RépubliqueAlgérienneDémocratiqueetPopulaire

MinistèredelʼEnseignementSupérieuretdelaRecherche Scientifique

UniversitéBlida 1

Faculté de technologie

Département d'électroniue

Mémoire de projet fe fin d'études

Présenté par :

Boudjerda Djahida & **Rehif Fedia**

Pour l'obtention du déplome de master en électroniue

Option : Electroniue des systèmes Embarués

Thème

Deep Learning pour la reconnaissance des symbols du langage des signes francais

Proposé par Dr . Bougherira H

Année Universitaire 2022- 2023

Remerciement

Louange à Dieu qui nous a donné la force, le courage, la patience, et l'espoir nécessaire pour accomplir ce travail et surmonter l'ensemble des difficultés.

On exprime notre gratitude, et nos remercîments à nos parents qui ont fait de leur mieux pour nous aider.

On tient à remercier vivement :

- *Nos plus remercimets s'adressent à* ***Mme bougherira hamida****, notre promotrice qui grace à ses efforts, son orientation ,sa patience et surtout sérieux a rendu possible la réalisation de ce projet .*
- *Je tiens aussi à remercier les membres du jury pour avoir accepté d'examiner et d'évaluer ce travail.*

- *L'ensemble des enseignants du département des électronique ayant consacré leurs vies sur la voie noble de l'enseignement et le partage*

On remercie aussi les personnes qui nous ont aidés et encouragés le long de ce travail.

Dédicace

Tout au début,je tiens à remercier le bon Dieu de m'avoir donné du courage et de patience afin de réaliser ce modeste travail que je dédieà:

Mes trè schers parents ***ABDELKADER****et* ***AICHA*** *pour leur amour,leur sen couragements,leur patience et leur saides continu elles le long Demon par cours d'études.*

Mes très chers sœurs **Chahira ,Fatha** *et* **Manel** *et chers frères* **Mouhammed** *et* **Ramzi** *à qui je souhaite une vie pleine de bonheur,de santé et de réussites.*

Toutes ma famille **Boudjerda et Reguige** *petits et grands,cousins et cousines, surtout oncle* ***Reguige Abd El Allah***

Ma cher binôme ***fedia,*** *et ma cher* ***Karima*** *je vous aime*

et mes amis ***: Malika, ,Siham, fozia , hadjira , aicha, fatima zahra ,Assma*** *et* ***Nesrine*** *et tout mes amies.*

Nos plus remercimets s'adressent à ***Mme bougherira hamida****, notre promotrice qui grace à ses efforts.*

Et à toute la promotion 2022 de système embarquer à qui je souhaite tout le bonheur et la réussite du monde.

Djahida

DÉDICACES

Je remercie Dieu tout puissant de m'avoir aidé pour achever ce travail que je dédie :

À mes chers parents, ma mère et mon père, qu'ils m'ont prodiguée avec tous les moyens et au prix de tous les sacrifices qu'ils ont consentis à mon égard, pour leur patience, leur amour et leurs encouragements. À ma chère binôme DJAHIDA.

Que ce travail leur apporte joie et fierté ; À mes chers frères et À mes chères sœurs et à toute la famille REHIF et famille DJAZAIRI ; À tous mes enseignants, particulièrement je remercierai mon encadreur Mm BOUGGHRIR.H, À mes très chers amis pour leurs aides et soutiens.

FEDIA

Résumé : Les gestes de la main jouent un rôle très important dans notre vie quotidienne comme l‟‟un des moyens de communication non-verbale les plus riches. Ainsi, dans le domaine d‟‟Interaction Homme-Machine (IHM). Dans notre mémoire on a utilisé l`apprentissage profond (Deep Learning) et les réseaux de neurones convolutifs (CNNs) pour classifier une base d`images en un ensemble de classes(d'images d'alphabets de la langue des signes américaine) , on a utilisé Matlab, nous allons présenter à travers ce mémoire nous avons présenté notre application en donnant quelques captures d'écran qui expliquent le déroulement et le fonctionnement de notre travail, ainsi que les resultants obtenus.

Mots clés : recherche d`images, classification, CNN : Convolutional Neural Networks, Deep Learning : apprentissage profond.

الملخص: تلعب إيماءات اليد دورًا مهمًا جدًا في حياتنا اليومية باعتبارها واحدة من أغنى وسائل الاتصال غير اللفظي. وهكذا ، في مجال التفاعل بين الإنسان والحاسوب (HMI). في أطروحتنا ، استخدمنا التعلم العميق والشبكات العصبية التلافيفية (CNNs) لتصنيف قاعدة بيانات الصور إلى مجموعة من الفئات (صور أبجدية لغة الإشارة الأمريكية) ، استخدمنا Matlab ، وسوف نقدمها من خلال هذه الذاكرة التي قدمناها التطبيق بإعطاء بعض اللقطات التي توضح سير العمل وعملنا وكذلك النتائج التي تم الحصول عليها.

الكلمات الرئيسية: البحث عن الصور ، التصنيف ، CNN: الشبكات العصبية التلافيفية ، التعلم العميق: التعلم العميق.

Abstract: Hand gestures play a very important role in our daily life as one of the richest means of non-verbal communication. Thus, in the field of Human-Computer Interaction (HMI). In our thesis, we used deep learning and convolutional neural networks (CNNs) to classify a database of images into a set of classes (images of American Sign Language alphabets), we used Matlab, we are going to present through this memory we presented our application by giving some screenshots which explain the progress and the functioning of our work, as well as the results obtained.

Keywords: image search, classification, CNN: Convolutional Neural Networks, Deep Learning: deep learning

Table des matières

Chapitre2: Deep learning (L'apprentissage profond).

Chapitre3 : Conception et implementation

Liste des figures

Chapitre 1

Chapitre2

Chapitre3

Liste des tableaux

Introduction générale

Nous vivons dans un monde numérique, où les informations sont stockées, traitées, indexés et recherchées par des systèmes informatiques, ce qui rend leur récupération une tâche rapide et pas cher. Au cours des dernières années, des progrès considérables ont été réalisés dans le domaine de classification d'images. Ce progrès est dû aux nombreux travaux dans ce domaine et à la disponibilité des bases d'images internationales qui ont permis aux chercheurs de signaler de manière crédible l'exécution de leurs approches dans ce domaine, avec la possibilité de les comparer à d'autres approches qu' ils utilisent sur les même bases.

L'homme fait recours toujours à la classification dans sa vie, et il essaie de répondre aux problèmes et questions sur les catégories des objets, c'est-à-dire réalisé l'affectation d'objets à leurs classes (formats, couleurs, tailles . . . etc.).

Généralement des bases d'observations caractérisent un domaine particulier (animaux, fruit, malades, gènes, . . . etc.), où on les regroupe en plusieurs classes.

La classification automatique d'images est une application de la reconnaissance de formes, qui consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification.

Les systèmes de recherche d'images cherchent à reconnaître les images similaires à une image requête, et cherche à obtenir de bonnes performances en moyenne sur toutes les images.

Parmi les problèmes rencontrés lors de la manipulation de grandes quantités d'images et la structuration, est la recherche. De ce fait l'utilisation de la classification peut contribuer à réduire la taille de ces problèmes.

Le deep learning est une technique d'apprentissage permettant à un programme, par exemple, de reconnaître le contenu d'une image ou de comprendre le langage parlé *« La technologie du deep learning apprend à* représenter le monde. *C'est-à-dire comment la machine va représenter la parole ou l'image par exemple ».*

Ce système d'apprentissage et de classification, basé sur des « réseaux de neurones artificiels » numériques est une technique courante en IA, permettant aux machines d'apprendre et reconnaitre des objets. Il approfondit sa compréhension de l'image avec des concepts de plus en plus précis.

Le but de notre projet est l'utilisation du deep learning pour la reconnaissance des symboles du langage des signes francais .

- Dans le premier chapitre on va présenter les Généralités sur des signe français.

- Le deuxième chapitre est consacré à la description des réseaux de neurones Convolutionnels ainsi que leur intérêt dans le domaine de la classification des images , et l'utilisation des réseaux de neurones dans la classification des images.

- Dans le troisième chapitre, on représente la conception de notre application, ainsi que la méthode d'implémentation de notre travail, en expliquant l'ensemble des choix techniques, (langage de programmation Matlab) utilisés pour la réalisation de cette application et on discute les différents résultats obtenus et à la fin on termine par une conclusion générale.

Chapitre 1

Généralité sur la reconnaissance du langage des signes français

1.1 Introduction

Le but de notre projet est de développer un algorithme basé sur l 'intelligence artificielle pour reconnaitre les signes du langage de signes français à partir d'images . Nous présentation donc, dans ce chapitre, le langage des méthodes classification de traitement d'images pour la reconnaissance d'objet et des méthodes basées sur l'intelligence artificielle.

1.2. Définition

La langue des signes française (LSF) est la langue de signes utilisée par les sourds francophones et leurs proches ainsi que certains malentendants pour communiquer avec les entendant. La LSF est une langue à part entière et un des piliers de l'identité de la culture sourde.[1]

La Langue des Signes Françaises (LSF) est une langue Visio-gestuelle. Même si chaque pays possède sa propre langue des signes, deux sourds se comprennent mieux, après un temps d'adaptation, que deux personnes entendantes ne parlant pas la même langue. Cette langue compte aujourd'hui entre 120000 et 150000 signeurs sourds ou non-sourds en France.

En LSF comme dans les autres langues des signes et quelques langues orales, il existe un ualphabet, appelé la dactylologie (voir figure 1). Celle-ci est représentée par différentes configurations de nos doigts et de nos mains. Elle permet entre autres d'épeler un mot, dont nous n'avons pas connaissance du signe ou qui n'en possède pas encore. Les signes doivent se réaliser avec la main dont nous nous servons pour écrire (la droite pour les droitiers et la gauche pour les gauchers).

Nous parlons de la LSF, pourtant nous pouvons remarquer des différences de signes au sein de cette même langue. En effet, la LSF que nous observons le plus est celle que nous voyons dans les médias : c'est le dialecte parisien. Mais il existe d'autres variétés régionales comme à Marseille, Aix-en-Provence ou encore Avignon.»

1.2.1 Alphabet dactylologique

L'alphabet dactylologique est la façon de signer l'alphabet latin. Il est utilisé pour épeler les noms propres ou les mots n'existant pas encore en LSF. La dactylologie de la LSF se fait d'une seule main, alors que les langues de la famille de la langue des signes britannique se pratiquent avec les deux mains.

Cet alphabet est constitué de signes dont les configurations gestuelles de la main représentent les lettres de l'alphabet [2].

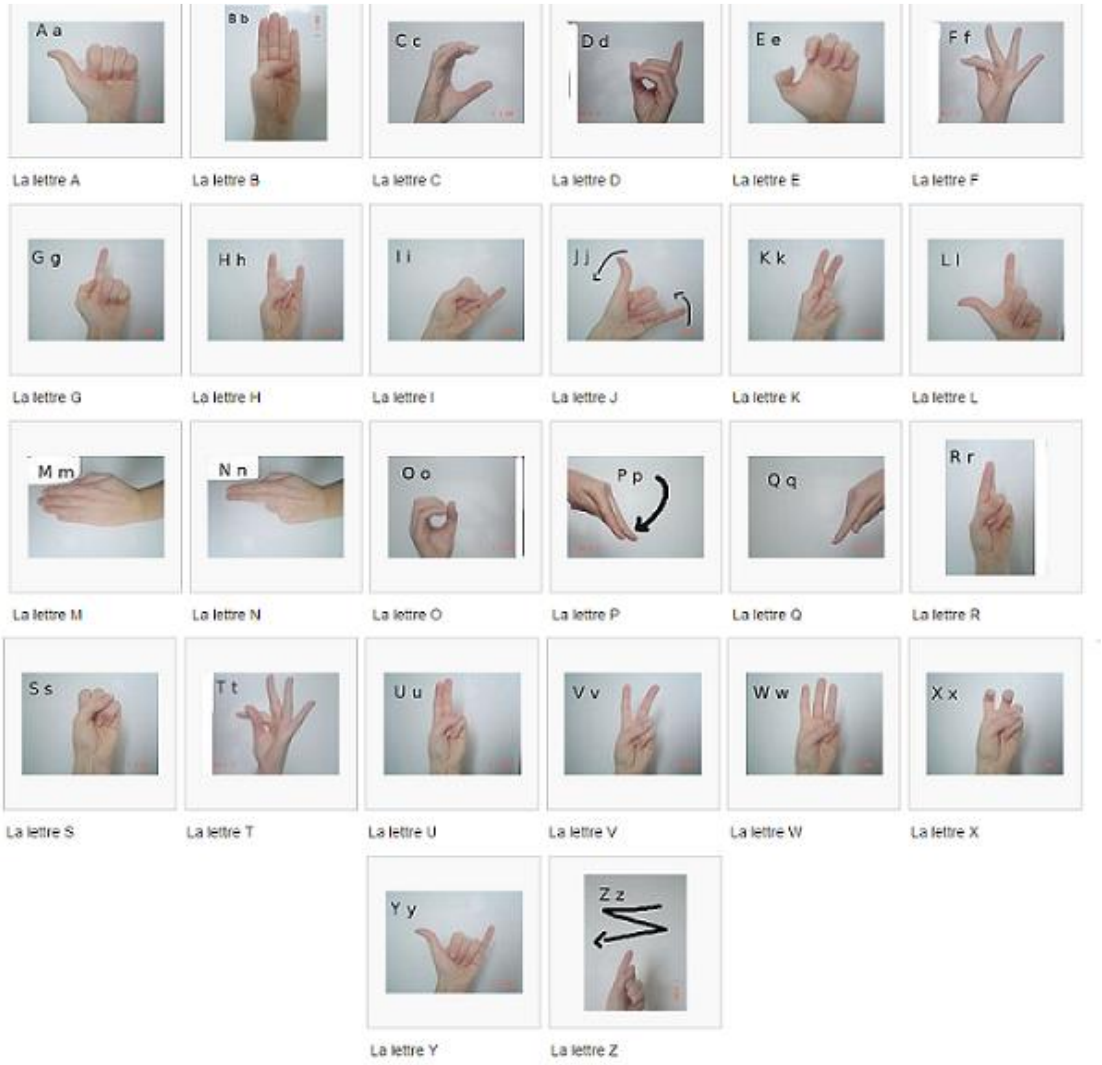


Figure 1.1: Alphabet dactylogique de la LSF

1.3. Structure de la langue des signes

Chaque geste d'une main peut être décomposé en cinq paramètres qui sont indépendants etpeuvent être aussi bien dynamiques qu'invariants durant l'émission du signe. Ces paramètressont définis comme suite [3]

- La configuration de la main
- L'orientation
- Le mouvement
- L'emplacement
- La direction du regard

- L’expression du visage
- Le mouvement du corps
- Le mouvement des lèvres

1.3.1.La configuration de la main

Une configuration est la forme prise par une main à un instant donné, et elle est définie parles postures des doigts et de la paume. En réalité, la forme d’une main varie d’une personne à une autre. Donc, elle n'est jamais parfaitement stable. Et correspond à la forme de la main définie par les doigts et la paume[4] La figure ci-dessous, montre un exemple :

Figure 1.2 : Le signe [**boire**]

1.3.2. L’orientation

L'orientation d'un signe peut avoir une valeur lexicale : un signe orienté vers différentes directions peut prendre des sens différents, comme pour le signe « ça va » (voir plus loin). Ce paramètre a aussi une valeur grammaticale, par exemple pour indiquer l'agent et le patient. Un signe comme « demander » orienté du locuteur vers l'interlocuteur signifie : je te demande ; orienté de l'interlocuteur vers le locuteur, il signifie : tu me demandes ; orienté du locuteur vers un emplacement sur le côté, il signifie : je lui demande, etc.

Figure1.3: L'orientation

Cette orientation est liée à l'espace situé devant le locuteur, où sont réalisés les signes. Il est appelé : espace de signation. Cet espace peut reproduire en miniature un espace réel, un peu comme une scène de théâtre — par exemple pour décrire un trajet, une rencontre, un lieu, un objet... Cet espace est aussi symbolique. Par exemple, placer une personne plus haut dans cet espace peut désigner une différence hiérarchique ou générationnelle. L'orientation d'un signe vers le haut ou le bas est ainsi signifiante. On place dans cet espace les différents personnages dont il est question, et les signes prennent sens par rapport à leur orientation vers ces différents emplacements.

1.3.3.Le mouvement

Le mouvement a différentes fonctions :

- au niveau du vocabulaire, il fait partie de réalisafon, et donc du sens, d'un signe, comme « bonjour ». Il peut permettre aussi distinguer un nom d'un verbe. Par exemple lesigne « lapin » (désigné par ses oreilles) avec le mouvement des oreilles qui se replient signifie : se soumettre

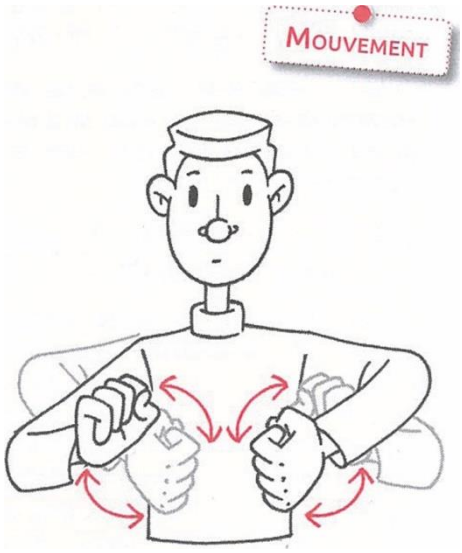


figure1.4: Mouvement

- au niveau de la syntaxe, comme pour l'orientafon, il permet de désigner l'agent, le patient ou le bénéficiaire
- au niveau de l'aspect, il peut indiquer différentes nuances. Par exemple, un mouvement répété plusieurs fois peut indiquer une action qui dure , un mouvement plus rapide ou plus lent est lié à la vitesse de l'action évoquée ; un mouvement réalisé avec une plus grande ampleur ou intensité apporte.

1.3.4.L'emplacement

CL’emplacement d'un signe dans l'espace de signassions a trois motivations :

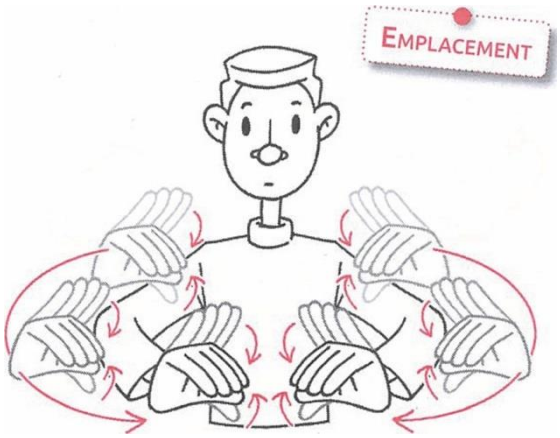


Figure1.5: Emplacement

- au niveau du discours, il peut reproduire un emplacement réel : la voiture était à gauche, le piéton à droite, le poteau devant...
- au niveau de la syntaxe, on peut attribuer un emplacement fictif aux protagonistes. Cela permet, par exemple pour les dialogues, de placer une personne à droite, l'autre à gauche, puis d'avoir simplement à tourner les épaules et le regard d'un côté ou de l'autre pour savoir qui parle.
- au niveau du vocabulaire, l'emplacement est souvent symbolique. Par exemple, beaucoup de signes dont le sens est lié à l'intellect sont réalisés au niveau de la tête (penser, réfléchir, rêver...).

Il est possible de « jouer » avec ces emplacements pour créer de nouvelles significations. Ainsi, « ouvrir la porte » se signe habituellement au niveau de l'abdomen. Le même signe réalisé au niveau du front prend le sens d'ouvrir son esprit ; ou encore, le signe « écouter » se signe au niveau de l'oreille. Le même signe réalisé au niveau des yeux signifie : écouter par son regard (c'est-à-dire la manière d'écouter des sourds). Ce procédé consistant à changer la signification d'un signe en changeant son emplacement est souvent utilisé pour la création de néologismes ou de jeux de signes.

1.3.5.Le regard

Le regard est très important en langue des signes. Il s'agit d'abord d'une langue où, pour se parler, il faut se regarder dans les yeux : tous les signes sont organisés pour que l'on n'ait pas à détacher son regard de l'interlocuteur — les signes éloignés du visage sont simples, afin d'être facilement perçus en vision périphérique.

Figure1.6: Regard

Le regard est au centre de l'organisation de la syntaxe : il désigne, dans l'espace de signation, l'emplacement dont il va être question. Il permet par exemple de viser un emplacement attribué à une personne, et de la pointer ainsi comme celle dont on va parler. Il est aussi au centre de l'organisation du discours : dans un discours rapporté, le regard du locuteur reproduit celui du personnage ou de l'entité dont il est question : lorsque le personnage, dans le récit, regarde à droite, le locuteur regarde à droite...

Lorsque le locuteur regarde de nouveau l'interlocuteur, il s'agit d'une interruption du récit, soit provisoire, par exemple pour insérer un commentaire, soit définitive — fin du récit.

Le regard a également un rôle pragmatique : par exemple, lorsqu'un locuteur ne veut pas être interrompu, il ne regarde pas son interlocuteur tout en continuant à signer— celui-ci ne peut pas alors « prendre la parole ».

Ces manières de gérer les échanges notamment par le regard nécessitent une certaine maîtrise, une grande finesse, sous peine de paraître grossier ou maladroit.

1.3.6.L'expression du visage

L'expression du visage est bien sûr liée à l'expression du sentiment du locuteur ou du personnage qu'il évoque. Accompagnant un verbe, elle peut refléter différentes valeurs modales : interrogatif, impératif, conditionnel, réprobatif, ironique, dubitatif, etc

Figure1.7: Expression du visage

Accompagnant un nom ou un adverbe, elle a souvent un rôle d'accentuation, comme en gonflant les joues, en fronçant les sourcils ou en plissant les yeux. Comme pour les autres paramètres, l'expression du visage peut avoir une certaine indépendance.

Ainsi, l'expression du visage qui accompagne un nom ou un verbe peut, dans certains contextes, être exprimée seule — l'interlocuteur comprenant l'ellipse ; ou encore l'expression du visage peut être en léger décalage par rapport au signe correspondant, afin d'accentuer un effet de surprise, d'interrogation, de doute...

1.3.7.Le mouvement du corps

Le mouvement des épaules est lié aux emplacernents attribués dans l'espace de signation : par exemple un mouvement vers la gauche focalisera l'attention sur la personne ou l'objet placé à gauche. Les épaules peuvent également être rentrées ou relevées en fonction de la taille ou de l'importance de ce qui est évoqué, ou en fonction de nuances apportées (timidité, peur, fierté, etc.).

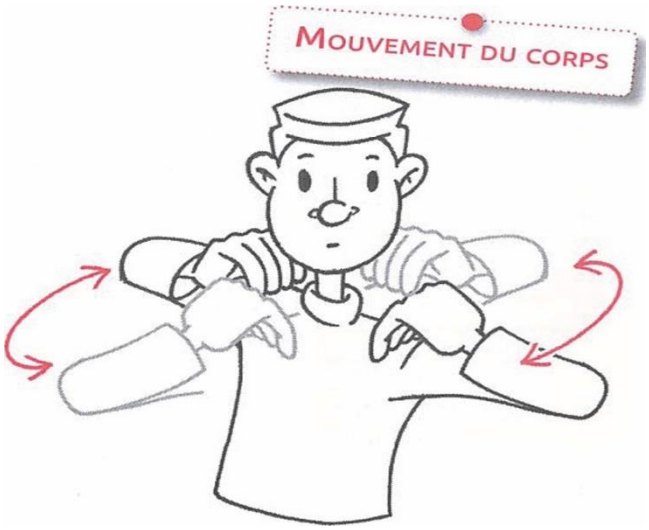


Figure1.8 : Mouvement du corps

Les nuances exprimées par les postures corporelles sont en fait très nombreuses et très fines. Il faut une certaine pratique de la langue des signes pour les maîtriser. Par exemple, un simple plissement du nez peut prendre le sens de « je suis d'accord ». Ainsi, . les épaules peuvent être rentrées ou relevées ; º les épaules peuvent avoir un mouvement vers l'avant ou vers l'arrière ; . la tête peut être tournée vers la droite ou la gauche, le haut ou le bas ; . les sourcils peuvent être relevés ou abaissés ; . les yeux peuvent être très ouverts ou fermés ; . les yeux peuvent cligner ; . les lèvres peuvent être pincées ou ouvertes ; . les joues peuvent être gonflées ou tendues ; . le nez peut être plissé...

1.3.8.Le mouvement des lèvres

Il existe deux types de mouvements :

- d'une part la labialisation de mots français, plus fréquente chez les personnes sourdes ayant suivi une éducation orale
- d'autre part différents mouvements comme les lèvres laissant passer un souffle, une série de « p », une forme d'explosion... Ces mouvements accompagnent des signes comme « fumer la pipe », « détonation », « ventilateur », etc.

figure 1.9: Mouvement des lévres

1.4. CLASSIFICATION

La classification est la dernière étape dans le processus de reconnaissance, l'essentiel de cette étape et d'identifier à quel classe appartient chaque geste on faisant une comparaison entre les caractéristiques extraites du geste actuelle et de celui enregistré dans la base de données, il existe 3 types de classification :

- Classification basée sur l'Apprentissage.
- Classification basée sur les Règles.
- Classification par les méthodes de l'ntelligence artificielle .

1.5.Classification basé Apprentissage

Quelques méthodes de classifications basées apprentissage sont discutés ci-dessous:

1.5.1. Les réseaux de neurones

Cette méthode est basée sur la modélisation du système nerveux humain élément appelé neurone et de son interaction avec les autres neurones pour transférer les informations. Chaque noeud comprend une fonction d'entrée qui calcule la somme pondérée et de la function d'activation afin de générer la réponse sur la base de la somme pondérée [5]. Il existe deux types de réseaux de neurones :« FeedForward » et « récurrent». Les méthodes basées sur cette approche risquent d'avoir un coût de calcul. éleve Pour les systèmes complexes, un tel modèle pourrait être très complexe. ajout de chaque nouveau geste / posture nécessite une reconversion complète du réseau de neurones [6]. Selon [7] cette méthode été appliquée pour

la classification des gestes sur des postures simple, mais cela prend du temps2 et lorsque le nombre de données d'apprentissage augmente, le temp nécessaire pour la classification est aussi augmenté.

1.5.2. Model de Markov caché

Le classificateur de Markov caché (MMC) appartient à la classe des classificateurs d'entrainement, il représente un modèle statistique, dans lequel la classe de geste de correspondance la plus probable est déterminé par un vecteur de caractéristiques donné, sur la base des données d'apprentissage. Cette méthode a été largement exploitée pour la reconnaissance des gestes temporelle [6]. Un MMC est constitué d'un ensemble d'états et des transitions d'état avec des probabilités d'observation. Pour regarder un geste, un distinct MMC est formé et la reconnaissance du geste est basée sur la génération de probabilité maximale par un MMC particulier[5].

1.5.3. K – plus proches voisins (KNN)

Cette méthode de classification utilise les fonctionnalités des vecteurs réunis dans la formation pour trouver les k voisins les plus proches dans un espace à n dimensions. La formation se compose principalement de l'extraction de possibles bonnes caractéristiques à partir d'images de formation, qui sont ensuite stockées pour une classification plus tard.

1.6.Classification basé règle

Dans les approches à base de règles, la règle est créée sur la base des vecteurs de caractéristiques, et ceux qui correspondent avec la règle peuvent être considérés comme le résultat final. C'est la méthode qu'on a suivis dans notre travail, cette méthode a des avantages comme elle a ces inconvénients aussi. Commençant par les avantages. L'avantage majeur et qui nous aide dans le cas de notre application qui doit être réalisé en temps réel c'est la vitesse. Au contraire des méthodes basé sur l'apprentissage qui prennent beaucoup de temps pour la reconnaissance, cette méthode donne des résultats avec une grande rapidité. L'inconvénient de cette approche est qu'elle dépond de la capacité de l'être humain à générer plusieurs règles qui assurent l'obtention d'une reconnaissance parfaite [8].

1.7 Traitement de l'image numérique

Le traitement d'images est une discipline de l'informatique qui étudie les images numériques et leurs transformations, dans le but d'améliorer leur qualité (Convolution et Filtrage) ou d'en extraire de l'information(Segmentation).

1.7.1. Convolution d'une image [9]

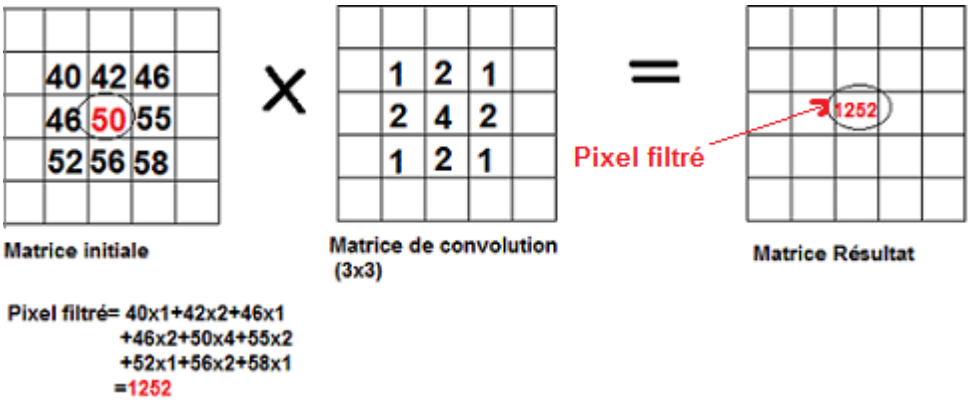
La convolution de l'image est un traitement d'une matrice par une autre appelée une matrice de convolution(ou noyau).

La plus part des filtres de traitement d'image utilisent des matrice de convolution.

Le filtre ou la matrice de convolution utilise une première matrice qui est l'image, c'est à dire, une collection de pixel en coordonnées rectangulaire 2D, et un noyau variable selon l'effet souhaité.

Les matrices de convolution les plus utilisées sont de la taille 3x3 ou 5x5 .

Le filtre étudie successivement chacun des pixels de l'image et pour chaque pixel que nous appelon "pixel initial", il multipli la valeur de ce pixel et de chacun des pixel qui l'entourent par la valeur correspondante dans le noyau. Il additionne l'ensemble des résulatats et le pixel initial prend alors la valeur du résultat final.



— A gauche, se trouve une partie de la matrice de l'image. Le pixel initial est encadré en rouge.

— Au centre, se trouve le noyau(la matrice de convolution) qui diffère selon le type de filtrage que l'on souhaite appliquer à l'image. Certain de ces filtres permettant la convolution portent le nom de ceux qui les ont créés comme le filtre de Sobel .

— A droite, se trouve le résultat de la convolution.

1.7.2. Filtrage d'une image [10]

Le filtrage est utilisé pour éliminer les bruits (parasites) sur l'image (lissage) pour des utilisations et applications ultérieurs.

Le principe du filtrage est de modifier la valeur des pixels d'une image, généralement dans le but d'améliorer son aspect. En pratique, il s'agit de créer une nouvelle image en se servant des valeurs des pixels de l'image d'origine et d'une matrice de convolution.

1.7.3.Segmentation de l'image [11]

La segmentation d'image est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères pré-définis. Les pixels sont ainsi regroupés en régions, qui constituent un pavage ou une partition de l'image. Il peut s'agir par exemple de séparer les objets du fond. Si le nombre de classes est égal à deux, elle est appelée aussi binarisation. La segmentation est une étape primordiale en traitement d'image. À ce jour, il existe de nombreuses méthodes de segmentation.

1.8.Types de segmentation

Il existe plusieurs méthodes de segmentation, Nous présentons dans ce qui suit quatre méthodes de segmentation: la première utilise le mouvement pour localiser la main, la seconde s'appuie sur la couleur particulière de la peau, la troisième utilise la notion de region et la quatrième utilise la notion de frontière.

1.8.1. Segmentation par différence image

La technique de différence d'image est bien connue en vision par ordinateur, elle consiste à soustraire une image par une autre, pixel à pixel, ce qui suppose que la caméra soit fixe afin qu'un pixel de l'image représente toujours le même lieu de l'espace au cours du temps. Cette approche se situe au niveau pixel, ce qui signifie qu'elle ne prend pas en compte les relations qui existent entre des pixels voisins [12].

1.8.2. Segmentation par couleur

La localisation par couleur est une approche classique dans le domaine de localisation de visages ou de main [13]. Elle s'appuie sur la segmentation de la couleur ou teinte-chair, qui permet de repérer des éléments dont la couleur est la plus proche de celle de la peau.

Pour détecter la main dans une vidéo nous utilisons l'analyse de la couleur de la peau, par ce qu'elle est très utilisée pour la détection du visage et des mains. C'est la première étape dans la plupart des systèmes informatiques pour l'interaction homme_ machine, de détection de mouvement, ou de la main, etc., et les performances de notre système est étroitement liés avec le résultat obtenu à cette étape.

En effet, Jones et Rehg[14] ont montré que la couleur de la peau présente une distribution caractéristique dans certains espaces colorimétriques, et que cette propriété peut être utilisée pour segmenter les régions de couleur de peau. Le résultat exporté est une image binaire contenant soit la valeur 1 pour les pixels de peau, soit la valeur 0 pour le reste des couleurs. Par mis les méthodes les plus célèbres utilisé dans ce type de segmentation, les méthodes base sur le choix d'un espace de couleur.

1.9.Classification par l'intelligence artificielle

1.9.1. L'intelligence artificielle (IA)

Est l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. Elle englobe donc un ensemble de concepts et de technologies, plus qu'une discipline autonome constituée. Des instances, telle la CNIL, notant le peu de précision de la définition de l'IA, l'ont présentée comme « le grand mythe de notre temps ».Souvent classée dans le groupe des mathématiques et des sciences cognitives, elle fait appel à la neurobiologie computationnelle (particulièrement aux réseaux neuronaux) et à la logique mathématique (partie des mathématiques et de la philosophie). Elle utilise des méthodes de résolution de problèmes à forte complexité logique ou algorithmique. Par extension, elle comprend, dans le langage courant, les dispositifs imitant ou remplaçant l'homme dans certaines mises en œuvre de ses fonctions cognitives. Ses finalités et enjeux ainsi que son développement suscitent, depuis l'apparition du concept, de nombreuses interprétations, fantasmes ou inquiétudes s'exprimant tant dans les récits ou films de science-fiction que dans les essais philosophiques5. La réalité semble encore tenir l'intelligence artificielle loin des performances du vivant ; ainsi, l'IA reste encore bien inférieure au chat dans toutes ses aptitudes naturelles[15]

1.9.2- **Les techniques de l'IA**

Il existe un débat scientifique et philosophique quant à l'intelligence[16], tels que :

- Comment le cerveau humain fonctionne-t-il ?

- Est-ce qu'on peut extraire l'intelligence humaine vers des machines ?

- Des machines peuvent-elles vraiment être intelligentes ?

Dans tous les secteurs d'activité, les techniques de l'IA tendent à élargir le champ d'actiondes machines, en leur donnant la possibilité de voir, d'entendre, de raisonner, de parler, d'agir,etc.

c'est-à-dire des systèmes qui possèdent des caractéristiques associées avec l'intelligence

dans le comportement humain ; le langage de compréhension, l'apprentissage, le raisonnement,les solutions des problèmes, etc.

On essaie seulement d'obtenir un comportement intelligent avec des ordinateurs, par le biais

des techniques : systèmes experts, logique floue, algorithme génétique, réseau de neurones, etc [17].

De ce fait, les différentes techniques d'IA nous permettent d'imiter le raisonnement humain et partagent un point commun qui est: l'aide à la décision

1.9.2- Domaines d'applications de l'Intelligence Artificielle

Comme l'Intelligence Artificielle, s'est développée en parallèle avec le calcul numérique et non en concurrence avec celui-ci. Elle s'attaque donc, à des secteurs d'application réputés difficiles ou impossibles à résoudre par le calcul numérique. Nous pouvons citer entre autres : - le calcul formel, - la vision : analyse de texte, d'images, - la robotique : génération de plans, - les machines autonomes : perception, interprétation, décision, action, - le langage naturel : traduction, compréhension, synthèse, - la démonstration de théorèmes, - les jeux, - la représentation des connaissances dans diverses disciplines (droit, médecine, automatique, etc.)[18].

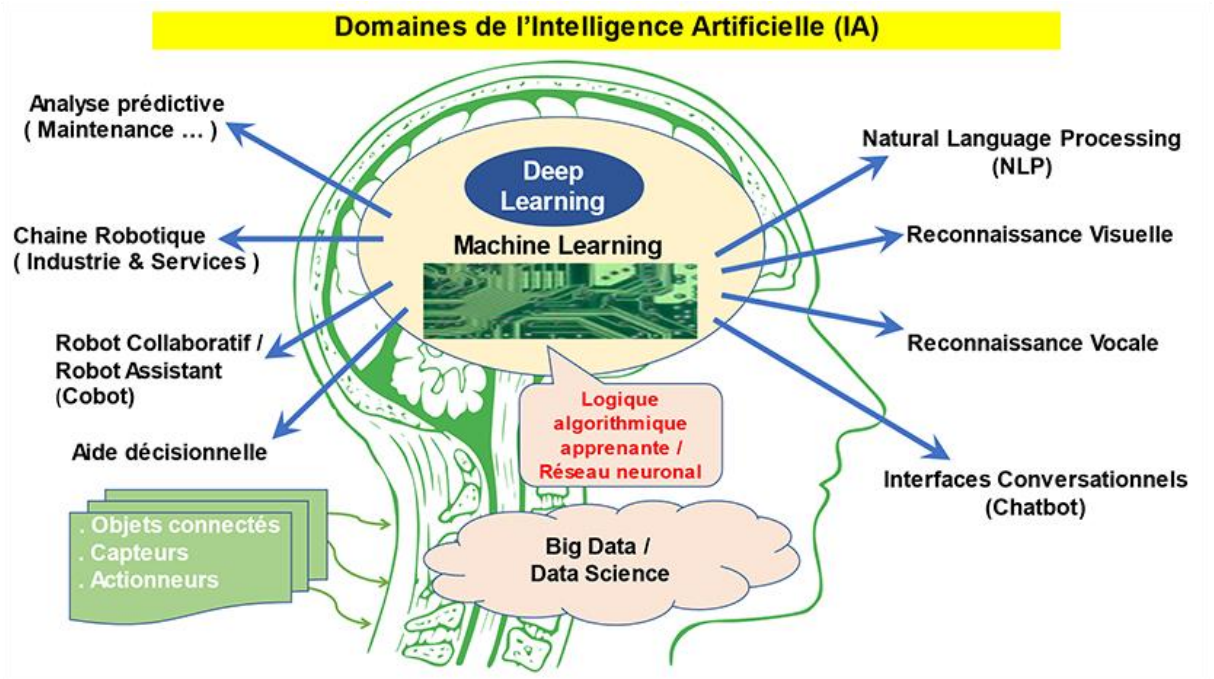


Figure 1-10 Domaines de l'intelligence artificielle (IA) .

1.10-Définition du Deep Learning

Le Deep Learning est un ensemble de méthodes d'apprentissage automatique (Machine Learning) tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires .Ces techniques ont permis des progrès importants et rapides dans les domaines de l'analyse du signal sonore ou visuel et notamment de la reconnaissance faciale, de la reconnaissance vocale, de la vision par ordinateur , du traitement automatisé du langage [16].

Le Deep Learning apprend à un modèle informatique comment réaliser des tâches de classification directement à partir d'images, de textes ou d'audio. Les modèles de Deep Learning peuvent atteindre un niveau de précision exceptionnel,. L'entraînement des modèles s'effectue via un vaste ensemble de données labellisées et d'architectures de réseaux de neurones qui contiennent de nombreuses couches. Le Deep Learning est une branche particulière du Machine Learning. Un processus de Machine Learning commence par l'extraction manuelle de caractéristiques pertinentes à partir d'images. En s'appuyant sur ces caractéristiques, un modèle qui catégorise les objets de l'image est ensuite créé. Dans un processus de Deep Learning, l'extraction de caractéristiques pertinentes à partir d'images est automatique[17]. En outre, le Deep Learning effectue un apprentissage « de bout en bout » : à partir de données brutes, un réseau se voit assigner des tâches à accomplir (une classification, par exemple) et apprend comment les automatiser. Un des avantages majeurs des réseaux de Deep Learning réside dans leur capacité à continuer à s'améliorer en même temps que le volume de vos données augmente .

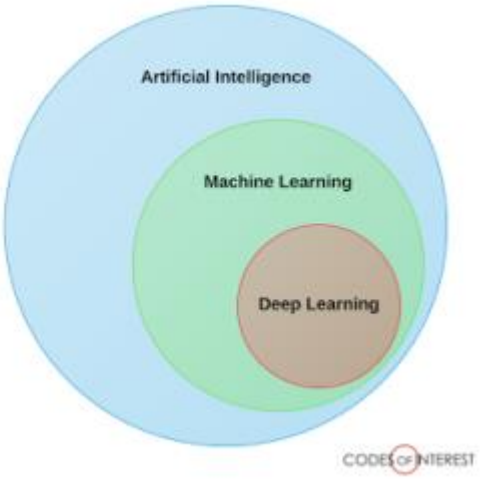


Figure1 .11-la relation entre l'intelligence artificielle, leML,et le deep Learning .

1.11. Conclusion

Dans ce chapitre, nous avons présenté la langue des signes français, ainsi que les trois types de signes utilisés dans l'alphabet de cette langue, la composition de signe (la Configuration, l'orientation, le mouvement, l'emplacement, mimique faciale, Dans le prochain chapitre, nous allons voir le principe de l`apprentissage profond (Deep learning) et détailler les réseaux de neurones et plus précisément l'utilisation des réseaux de neurones convolutionnels(CNNs) dans la classification des images afin d'appliuer le deeplearning à la reconnaissance des signes.

Chapitre II

Deep learning (L'apprentissage profond).

2 .1-INTRODUCTION

L'apprentissage profond (deeplearning) est une technique d'apprentissage automatique (machine learning) qui a considérablement amélioré les résultats dans de nombreux domaines tels que la vision par ordinateur, la reconnaissance de la parole et la traduction automatique. Les techniques d'apprentissage profond permettent, à l'aide de données, de résoudre de nombreux problèmes dans de nombreux domaines de l'économie tels que la santé, le transport, le commerce, la finance ainsi que l'énergie[19].

Le deep learning ou apprentissage profond passe par le déploiement d'un réseau de neurones artificiel préalablement entraîné. Il s'agit d'une pratique d'IA issue de l'apprentissage automatique ou machine learning[16].

L'apprentissage profond est une révolution en intelligence artificielle [20] .

Le deeplearning ou apprentissage profond, regroupe actuellement les méthodes les plus efficaces et les plus performantes appliquées dans la communauté de l'apprentissage automatique.

2.2-Fonctionnement du deep Learning

Le *deep Learning* s'appuie sur un réseau de neurones artificiels s'inspirant du cerveau humain. Ce réseau est composé de dizaines voire de centaines de « couches » de neurones, chacune recevant et interprétant les informations de la couche précédente. Le système apprendra par exemple à reconnaître les lettres avant de s'attaquer aux mots dans un texte, ou détermine s'il y a un visage sur une photo avant de découvrir de quelle personne il s'agit.

le Machine Learning et le deep Learning est basé sur l'idée des réseaux de neurones artificielles et il est taillé pour gérer de larges quantités de données en ajoutant des couches au réseau . Un modèle de deep Learning a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires[18] .

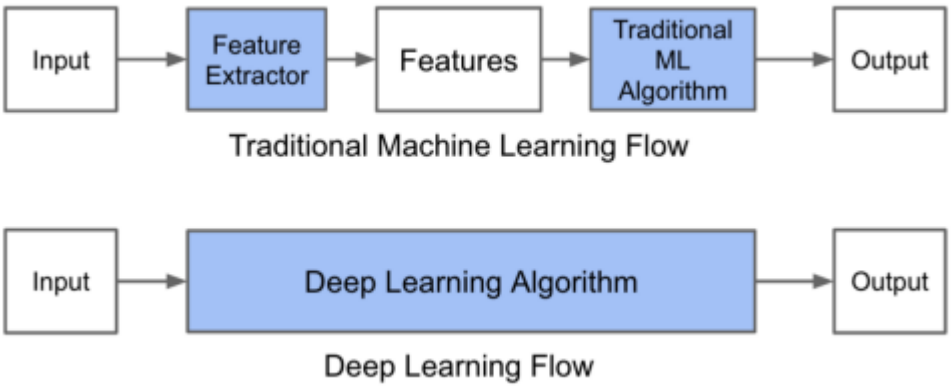


Figure2. 1 – Le procède du MLclassique comparé à celui du Deep Learning

2.3- Domaines d'application de l'apprentissage profonde

Ces techniques se développent dans le domaine de l'informatique appliquée aux NTIC (reconnaissance visuelle — par exemple d'un panneau de signalisation par un robot ou une voiture autonome — et vocale notamment) à la robotique, à la bio-informatique, la reconnaissance ou comparaison de formes, la sécurité, la santé, etc…, la pédagogie assistée par l'informatique, et plus généralement à l'intelligence artificielle[18]. L'apprentissage profond peut par exemple permettre à un ordinateur de mieux reconnaitre des objets hautement déformables et/ou analyser par exemple les émotions révélées par un visage photographié ou filmé, ou analyser les mouvements et position des doigts d'une main, ce qui peut être utile pour traduire le langage des signes, améliorer le positionnement automatique d'une caméra, etc… Elles sont utilisées pour certaines formes d'aide au diagnostic médical (ex. : reconnaissance automatique d'un cancer en imagerie médicale)[21], ou de prospective ou de prédiction (ex. : prédiction des propriétés d'un sol filmé par un robot)

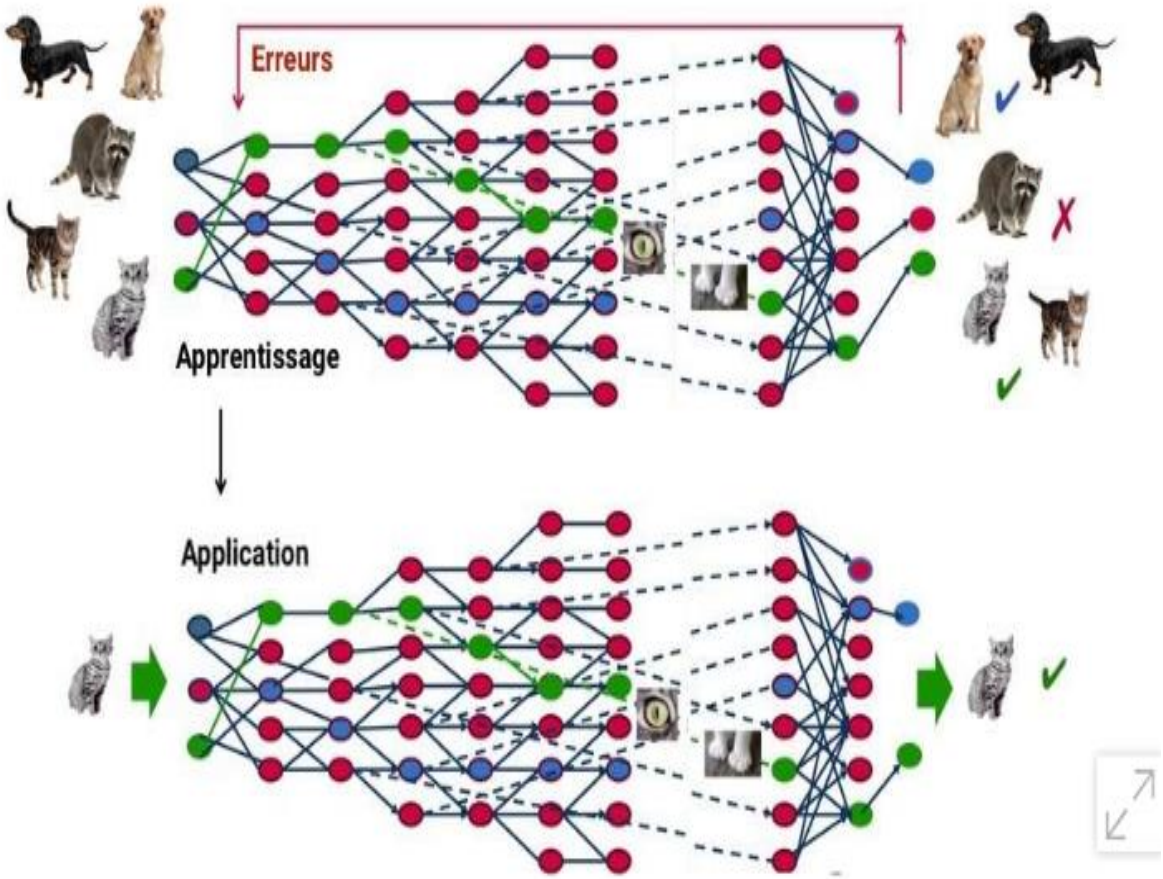


Figure2.2- réseau de neurones profond **.**

2.4- Architectures de réseaux de neurones profonds

Il existe un grand nombre de variables d'architectures profondes. La plupart d'entre elles sont dérivés de certaines architectures originales. Nous allons choisir les réseaux de neurones convolutifs (CNNs).

2.5- Quelques explications sur les CNNs

2.5.1- Principe d'architecture d'un CNN

Les réseaux de neurones convolutifs sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais Convolutional Neural Network, ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a deux dimensions pour une image aux niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu]. La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers d'une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions .certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. En fin, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN[23].[22]

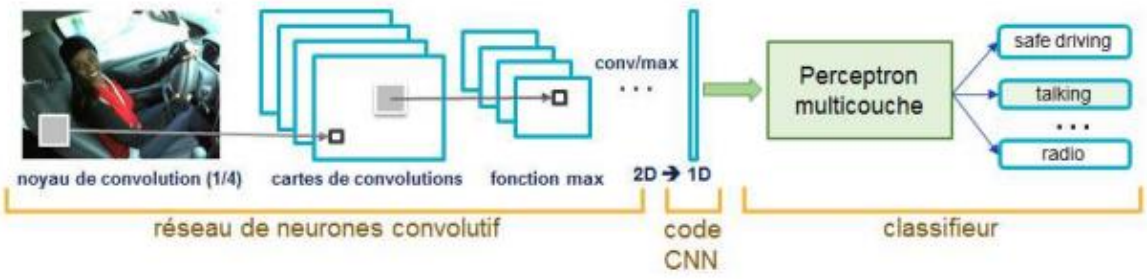


Figure2. 3-Les réseaux de neurones convolutifs.

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie

Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image. La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs

numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories .

2.6- Fonctionnement d'un CNN

Un CNN est composé de plusieurs couches d'opérations[24], chacune avec sa propre fonction. Le CNN prend une image en entrée et la transmet à la première couche. Les informations sur les entités sont propagées à travers les couches cachées. Pour chaque couche, les fonctions d'activation effectuent une activation élément par élément à la sortie produite par la couche précédente. La sortie de la couche finale est ensuite comparée à la sortie cible. Ce processus est appelé passage direct (forwardpass).

2.7-ce que la convolution

La convolution est un outil mathématique simple qui est très largement utilisé pour le traitement d'image[25], ce qui explique que les réseaux de neurones à convolution soient particulièrement bien adaptés à la reconnaissance d'image.
La convolution agit comme un filtrage. On définit une taille de fenêtre qui va se balader à travers toute l'image (rappelez-vous qu'une image peut être vue comme étant un tableau). Au tout début de la **convolution**, la fenêtre sera positionnée tout en haut à gauche de l'image puis elle va se décaler d'un certain nombre de cases (c'est ce que l'on appelle le**pas**) vers la droite et lorsqu'elle arrivera au bout de l'image, elle se décalera d'un **pas** vers le bas ainsi de-suite jusqu'à ce que le **filtre** est parcourue la totalité de l'image

2.8-Les réseaux de neurones convolutifs

réseaux de neurones convolutifs (Convolution Neural Network (CNN)) sont un type de réseau de neurones spécialisés pour le traitement de données ayant une topologie semblable à une grille. Les exemples comprennent des données de type série temporelle, qui peuvent être considérées comme une grille 1D en prenant des échantillons à des intervalles de temps réguliers et des données de type image, qui peuvent être considérées comme une grille 2D de pixels. Les réseaux convolutifs ont connu un succès considérable dans les applications pratiques. Le nom « réseau de neurones convolutif » indique que le réseau emploie une opération mathématique appelée convolution. La convolution est une opération linéaire

spéciale. Les réseaux convolutifs sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication -matricielle dans au moins une de leurs couches[26].

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

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

La réponse d'un neurone individuel aux stimuli dans son champ réceptif peut être approchée mathématiquement par une opération de convolution.

2.9-Les couch de convolution

La couche de convolution est la composante clé des reseaux de neurones convolitifs et constitue toujours au moins leur premier couche

• La couche de convolution (CONV) qui traite les données d'un champ récepteur.

• La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de

l'image intermédiaire (souvent par sous-échantillonnage).

• La couche de correction (ReLU), souvent appelée par abus 'ReLU' en référence à la fonction

d'activation (Unité de rectification linéaire).

• La couche "entièrement connectée" (FC), qui est une couche de type perceptron dont le nombre correspond au nombre de classe.

• La couche de perte (LOSS)

Exemple d'une architecture CNN et sa sortie

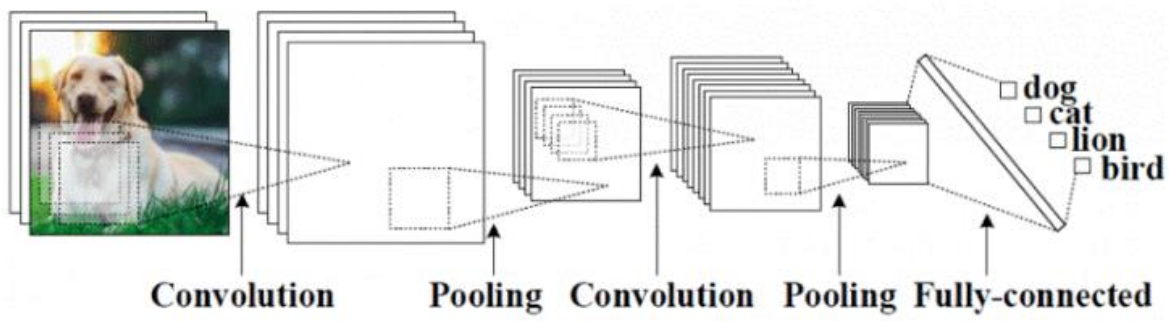


Figure2. 4-Exemples de modèles de CNN

Après la partie convolutive d'un CNN, vient la partie classification. Cette partie classification, commun à tous les modèles de réseaux de neurones, correspondant à un modèle perceptron multicouche (MLP).

Son objectif est d'attribuer à chaque échantillon de données une étiquette décrivant sa classe d'appartenance.

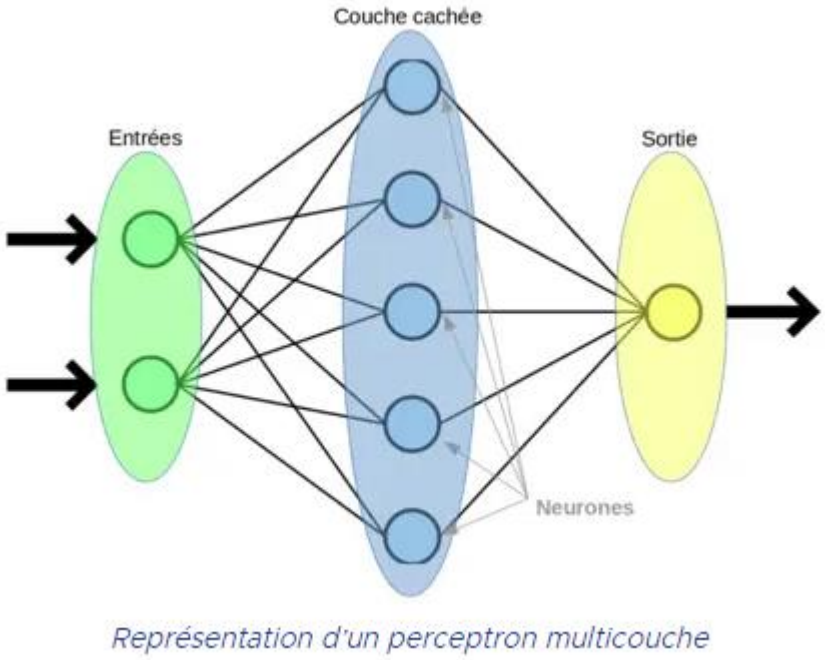


Figure2.5 :un perceptron multicouche

L'algorithme que les perceptrons utilisent pour mettre à jour leurs poids (ou coefficients de réseaux) s'appelle **la rétropropagation du gradient de l'erreur,** célèbre algorithme de descente de gradient que nous verrons plus en détail par la suite . Généralement, l'architecture d'un Convolutional Neural Network est sensiblement la même :

2.9.1Couche de convolution (CONV)

La couche de convolution est le bloc de construction de base d'un CNN. Trois paramètres permettent de dimensionner le volume de la couche de convolution la profondeur, le pas et la marge. 1. Profondeur de la couche : nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur). 2. Le pas : contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand. 3. La marge (à 0) ou zéro padding : parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce 'zéro-padding' est le troisième hyper paramètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée. Si le pas et la marge appliquée à l'image d'entrée permettent de contrôler le nombre de champs récepteurs à gérer (surface de traitement), la profondeur permet d'avoir une notion de volume de sortie, et de la même manière qu'une image peut avoir un volume, si on prend une profondeur de 3 pour les trois canaux RGB d'une image couleur, la couche de convolution va également

présenter en sortie une profondeur. C'est pour cela que l'on parle plutôt de "volume de sortie" et de "volume d'entrée", car l'entrée d'une couche de convolution peut être soit une image soit la sortie d'une autre couche de convolution. La taille spatiale du volume de sortie peut être calculée en fonction de la taille du volume d'entrée Wi la surface de traitement K (nombre de champs récepteurs), le pas S avec lequel ils sont appliqués, et la taille de la marge P.

Le nombre de neurones du volume de sortie est calculé selon la formule : $Wo = \frac{\mathrm{wi-k+2P}}{S} + 1$

Si W0 n'est pas entier, les neurones périphériques n'auront pas autant d'entrée que les autres. Il faudra donc augmenter la taille de la marge (pour recréer des entrées virtuelles)

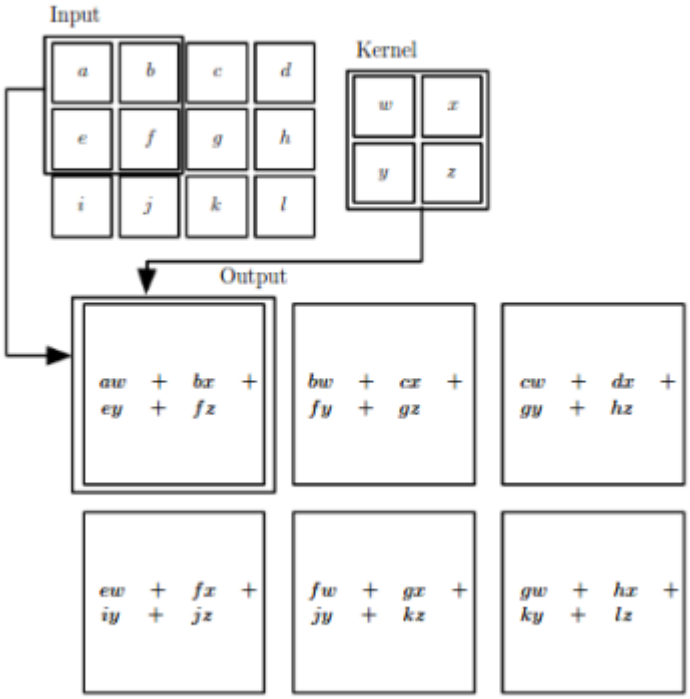


Figure2.6:Exemple d'une convolution 2D

dizaines ou des centaines de pixels. Cela signifie que nous pouvons stocker moins de paramètres, ce qui réduit les besoins en matière de mémoire du modèle et améliore son efficacité. Cela signifie également que le calcul des sortie nécessite moins d'opérations. Ces améliorations en matière d'efficacité sont généralement assez importantes.

• parameter sharing : (le partage des paramètres) Se réfère à l'utilisation du même paramètre pour plus d'une fonction dans un modèle. Dans un réseau de neurones convolutif, chaque élément du noyau est utilisé à chaque position de l'entrée (sauf peut-être quelques-uns des pixels des bords, selon le choix de conception concernant la frontière). Le paramètre sharing utilisé par l'opération de convolution signifie que, plutôt que d'apprendre un ensemble de paramètres distincts pour chaque emplacement, nous n'apprenons qu'un ensemble ce qui réduit encore davantage les exigences de stockage du modèle.

• equivariant représentations : Dans le cas de la convolution, la particularité du paramètre sharing fait que la couche possède une propriété appelée equivariant représentations. Pour dire qu'une fonction est équivariante signifie que si l'entrée change, la sortie change de la même manière.

Spécifiquement, une fonction f(x) est équivariante à la fonction g si f(g(x)) = g(f(x)). Dans le cas du traitement d'image, la convolution crée une carte 2-D de certaines caractéristiques

apparaissant dans l'entrée. Si nous déplaçons l'objet dans l'entrée, sa représentation va se déplacer de la même façon dans la sortie. Par exemple, dans le cas d'une image, Il est utile de détecter les arêtes dans la première couche d'un réseau convolutif. Les mêmes arêtes apparaissent plus ou moins partout dans l'image, donc il est pratique de partager des paramètres sur l'image entière.

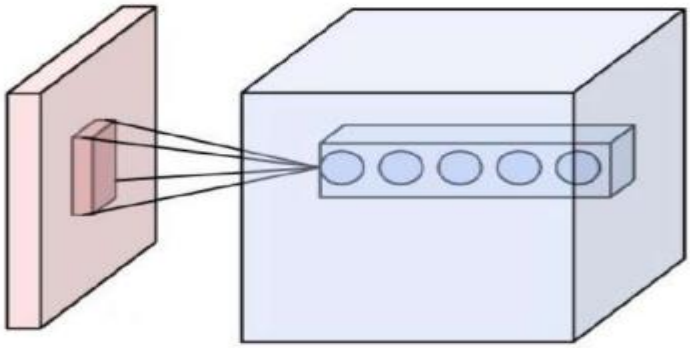


Figure 2.7-Ensemble de neurones (cercles) créant la profondeur d'une couche de convolution (bleu). Ils sont liés à un même champ récepteur (rouge).

2.9.2Couche de Pooling (POOL)

le pooling est Un autre concept important des CNN,ce qui est une forme de sous-échantillonnage de l'image. L'image d'entrée est découpée en une série de rectangles de n pixels de côté ne se chevauchant pas (pooling). Chaque rectangle peut être vu comme une tuile. Le signal en sortie de tuile est défini en fonction des valeurs prises par les différents pixels de la tuile. Le pooling réduit la taille spatiale d'une image intermédiaire, réduisant ainsi la quantité de paramètres et de calcul dans le réseau. Il est donc fréquent d'insérer périodiquement une couche de pooling entre deux couches convolutives successives d'une architecture CNN pour contrôler l'overfitting (sur-apprentissage). L'opération de pooling crée aussi une forme d'invariance par translation. La couche de pooling fonctionne indépendamment sur chaque tranche de profondeur de l'entrée et la redimensionne uniquement au niveau de la surface. La forme la plus courante est une couche de mise en commun avec des tuiles de taille 2x2 (largeur/hauteur) et comme valeur de sortie la valeur maximale en entrée. On parle dans ce cas de « MaxPool 2x2 ».

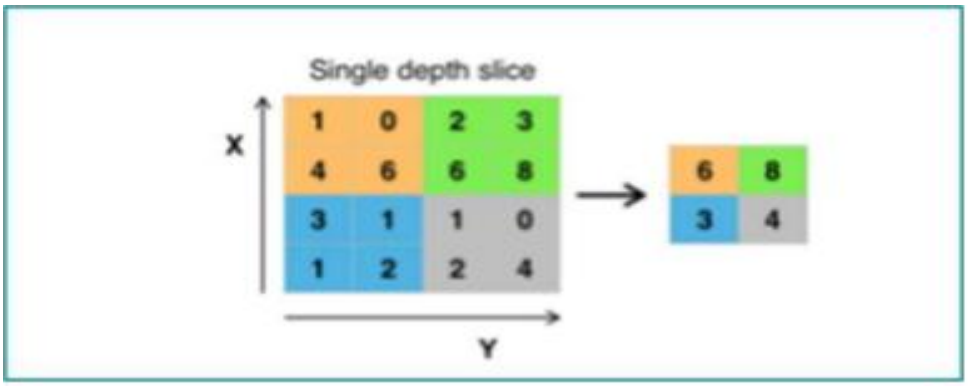


Figure 2.8 : Pooling avec un filtre 2x2 et un pas de 2

Il existe plusieurs types de pooling :

¬ Le « max pooling« qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image

¬ Le « meanpooling » (ou averagepooling), soit la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels

¬ Le « sumpooling « c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme) .

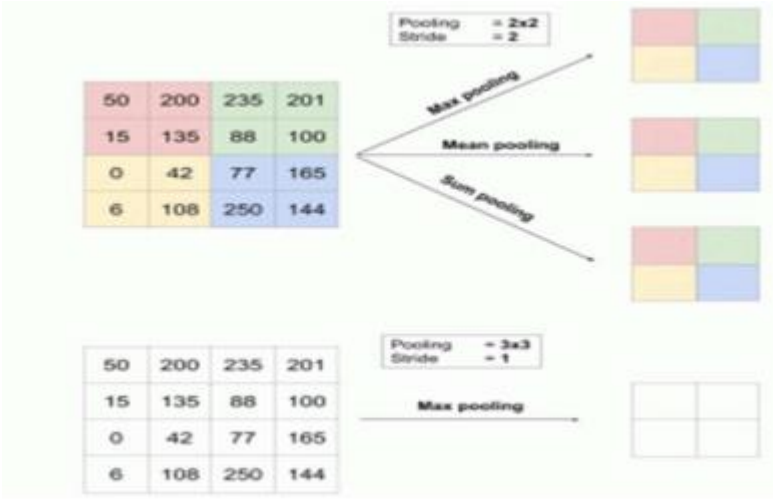


Figure2.9 : Calcul du pooling sur une image 4×4. Un pooling de 2×2 signifie que l'on sélectionne les pixels en carrés de 2×2.

2.9.3-La couched'activationRelu (Rectified Linear Units)

Il est possible d'améliorer l'efficacité du traitement en intercalant entre les couches de traitement une couche qui va opérer une fonction mathématique (fonction d'activation) sur les signaux de sortie. La fonction Relu (abréviation de Unités Rectifié linéaires) est définie comme suit : F(x)=max (0, x). Cette fonction force les neurones à retourner des valeurs positives

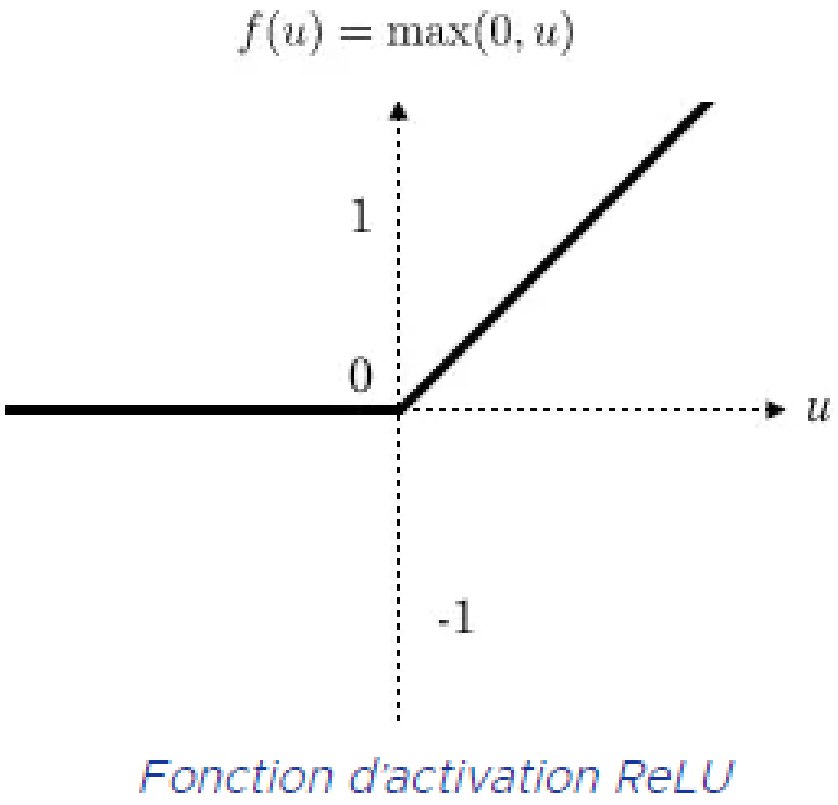


Figure2.10:fonction activation ReLU

2.9.4-Couche FullyConnected (FC)

- Ces couches sont placées en fin d'architecture de CNN et sont entièrement connectées à tous les neurones de sorties (d'où le terme fully-connected). Après avoir reçu un vecteur en entrée, la couche FC applique successivement une combinaison linéaire puis une fonction d'activation dans le but final

de classifier l'input image. Elle renvoie enfin en sortie un vecteur de taille d correspondant au nombre de classes dans lequel chaque composante représente la probabilité pour l'input image d'appartenir à une classe.

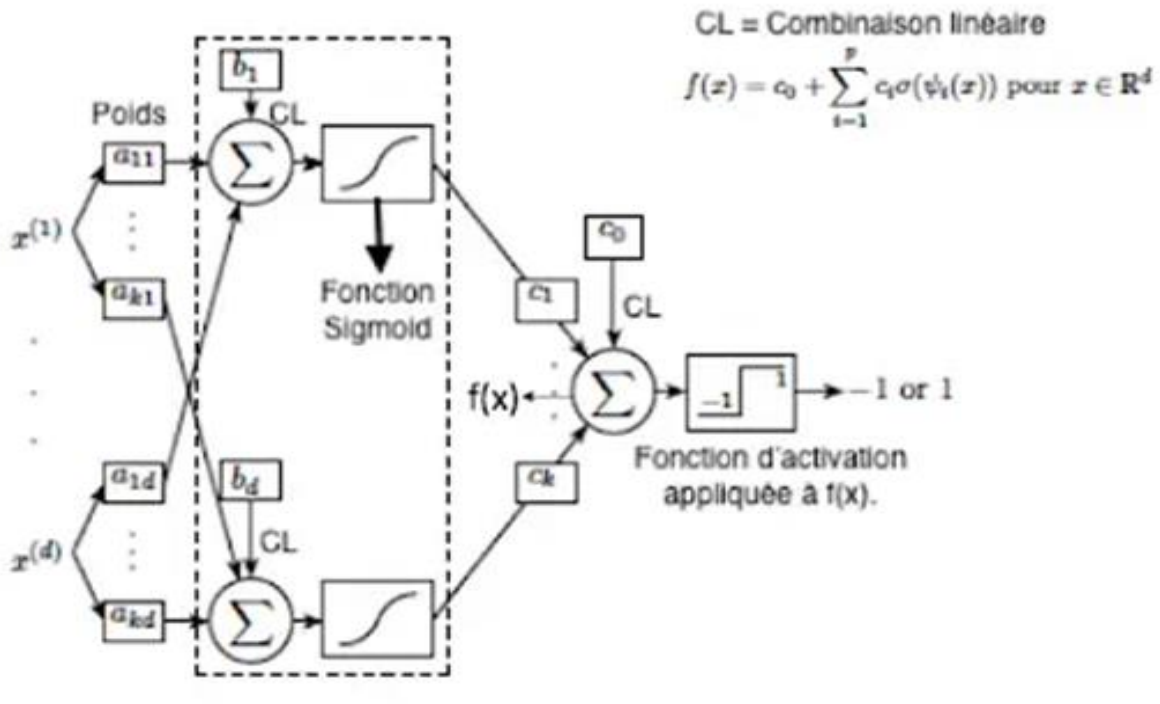


Figure2.11:fonctionnement d'un réseau neuronal a 2 couches cachees

2.10-Avantages de CNN

Un avantage majeur des réseaux convolutifs est l'utilisation d'un poids unique associé aux signaux entrant dans tous les neurones d'un même noyau de convolution. Cette méthode réduit l'empreinte mémoire, améliore les performances et permet une invariance du traitement par translation. C'est le principal avantage du CNN par rapport au MLP, qui lui considère chaque neurone indépendant et donc affecte un poids différent à chaque signal entrant. Lorsque le volume d'entrée varie dans le temps (vidéo ou son), il devient intéressant de rajouter un paramètre de temporisation (delay) dans le paramétrage des neurones. On parlera dans ce cas de réseau neuronal à retard temporel (TDNN).

2.11-Quelques réseaux convolutifs célèbres

• Le Net : Les premières applications réussies des réseaux convolutifs ont été développées dans les années 1990. Parmi ceux-ci, le plus connu est l'architecture Le Net utilisée pour lire les codes postaux, les chiffres, etc.

• Alex Net : Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était Alex Net, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton. Alex Net a été soumis au défi ImageNet ILSVRC en 2012 et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des 35 couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling).

• ZFnet : Le vainqueur de ILSVRC challenge 2013 était un réseau convolutif de Matthew Zeiler et Rob Fergus. Il est devenu ZFNet (abréviation de Zeiler et Fergus Net). C'était une amélioration de AlexNet en ajustant les hyper-paramètres de l'architecture, en particulier en élargissant la taille des couches convolutifs et en réduisant la taille du noyau sur la première couche.

• Google Net : Le vainqueur de ILSVRC challenge 2014 était un réseau convolutif de Szegedy et al. De Google. Sa principale contribution a été le développement d'un module inception qui a considérablement réduit le nombre de paramètres dans le réseau (4M, par rapport à AlexNet avec 60M). En outre, ce module utilise le global AVG pooling au lieu du PMC à la fin du réseau, ce qui élimine une grande quantité de paramètres. Il existe également plusieurs versions de GoogleNet, parmi elles, Inception-v4.

• RESNet : Résiduel network développé par Kai ming He et al. A été le vainqueur de ILSVRC 2015. Il présente des sauts de connexion et une forte utilisation de la batch normalisation. Il utilise aussi le global AVG pooling au lieu du PMC à la fin.

2.12- Classification des images et l'apprentissage machine

La classification automatique des images consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification. On retrouve ainsi la classification d'objets, de scènes, de textures, la reconnaissance de visages, d'empreintes digitale et de caractères. Il existe deux principaux types d'apprentissage : l'apprentissage supervisé et l'apprentissage

non-supervisé. Dans l'approche supervisée, chaque image est associée à une étiquette qui décrit sa classe d'appartenance. Dans l'approche nonsupervisée les données disponibles ne possèdent pas d'étiquettes. Dans notre travail on s'intéresse de l'approche supervisée[27].

L'objectif de la classification d'images est d'élaborer un système capable d'affecter une classe automatiquement à une image. Ainsi, ce système permet d'effectuer une tâche d'expertise qui peut s'avérer coûteuse à acquérir pour un être humain en raison notamment de contraintes physiques comme la concentration, la fatigue ou le temps nécessité par un volume important de données images. Les applications de la classification automatique d'images sont nombreuses et vont de l'analyse de documents à la médecine en passant par le domaine militaire. Ainsi on retrouve des applications dans le domaine médical comme la reconnaissance de cellules et de tumeurs, la reconnaissance d'écriture manuscrite pour les chèques les codes postaux. Dans le domaine urbain comme la reconnaissance de panneaux de signalisation la reconnaissance de piétons la détection de véhicules la reconnaissance de bâtiments pour aider à la localisation. Dans le domaine de la biométrie comme la reconnaissance de visage, d'empreintes, d'iris [21].

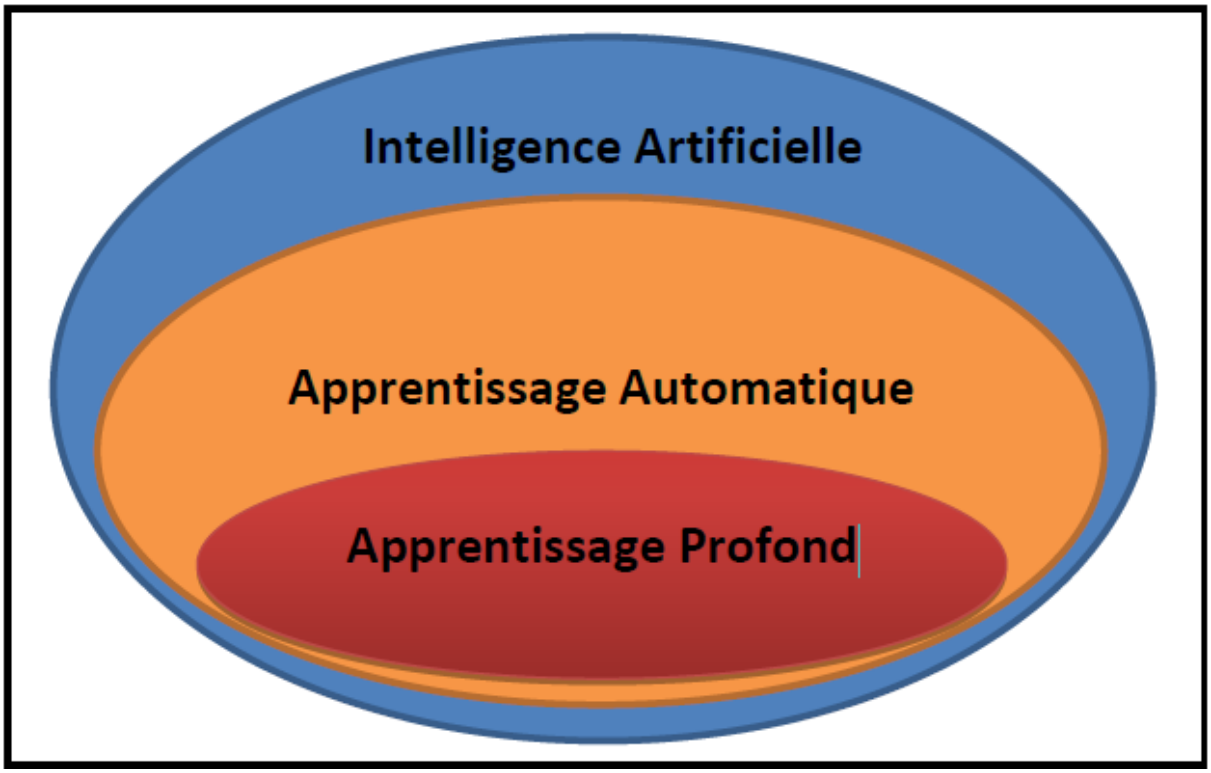


Figure2.12:Intelligence artificielle est ses sous-domaines

2.13- Méthodes de classification

De nombreuses méthodes classiques ont été consacrés, elles peuvent être séparées en deux grandes catégories : les méthodes de classification supervisée et les méthodes de classification non supervisée[28].

2.13.1 Méthodes supervisées

L'objectif de la classification supervisée est principalement de définir des règles permettant de classer des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets. On dispose au départ d'un échantillon dit d'apprentissage dont le classement est connu. Cet échantillon est utilisé pour l'apprentissage des règles de classement. Il est nécessaire d'étudier la fiabilité de ces règles pour les comparer et les appliquer, évaluer les cas de sous apprentissage ou de sur apprentissage (complexité du modèle). On utilise souvent un deuxième échantillon indépendant, dit de validation ou de test.

2.13.2 Méthodes non supervisées

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

.

2.14-Premiers pas avec l'apprentissage en profondeur

Si vous débutez dans l'apprentissage en profondeur, un moyen simple et rapide de commencer est d'utiliser un réseau existant, tel que Google Net, un CNN formé sur plus d'un million d'images. GoogLeNet est le plus couramment utilisé pour le classement des images. Il peut classer les images en 1000 images différentes catégories, y compris les claviers, les souris d'ordinateur, les crayons et autres matériel de bureau, ainsi que diverses races de chiens, chats, chevaux, et d'autres animaux.

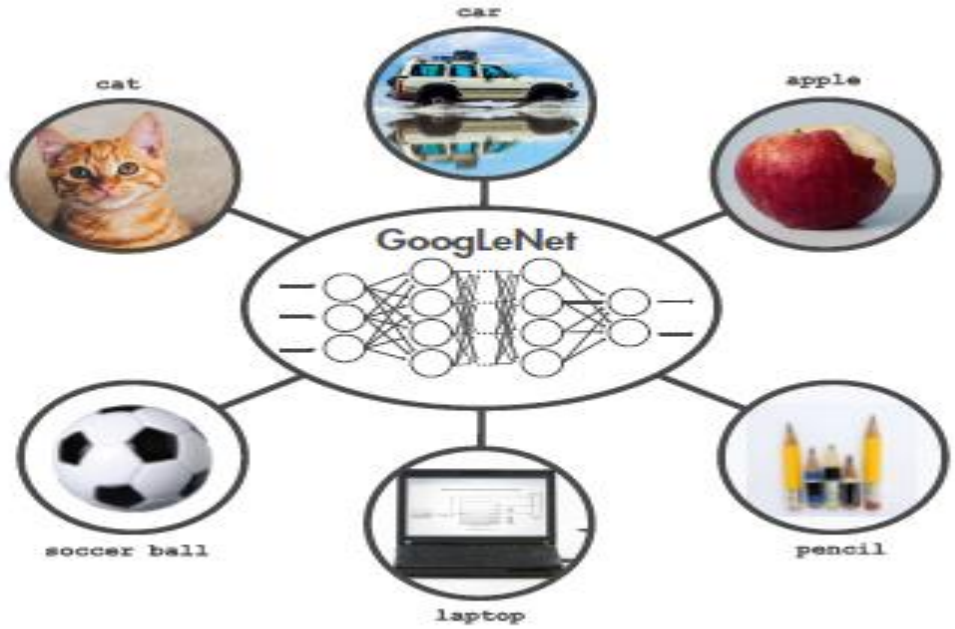


Figure2.13:GoogLeNet

2.15-Comprendre le modèle GoogLeNet – Architecture CNN

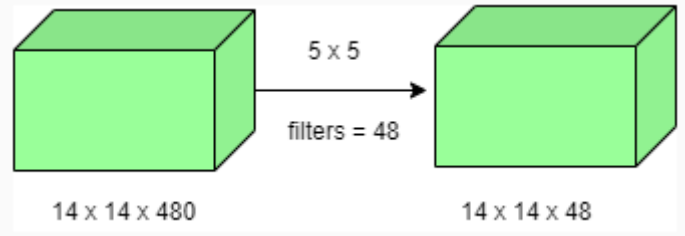
Google Net (ou Inception V1) a été proposé par la recherche de Google (avec la collaboration de diverses universités) en 2014 dans le document de recherche intitulé « Going Deeper with Convolutions ». Cette architecture a remporté le concours de classification d'images ILSVRC 2014. Il a fourni une diminution significative du taux d'erreur par rapport aux gagnants précédents AlexNet (gagnant de l'ILSVRC 2012) et ZF-Net (gagnant de l'ILSVRC 2013) et un taux d'erreur nettement inférieur à celui de VGG (finaliste de 2014). Cette architecture utilise des techniques telles que les convolutions 1×1 au milieu de l'architecture et la mise en commun de la moyenne globale.

2.16 -Fonctionnalités de Google Net

L'architecture GoogLeNet est très différente des architectures de pointe précédentes telles qu'AlexNet et ZF-Net. Il utilise de nombreux types de méthodes telles que la convolution *1 × 1* et la mise en commun des moyennes globales qui lui permettent de créer une architecture plus profonde. Dans l'architecture, nous aborderons certaines de ces méthodes :

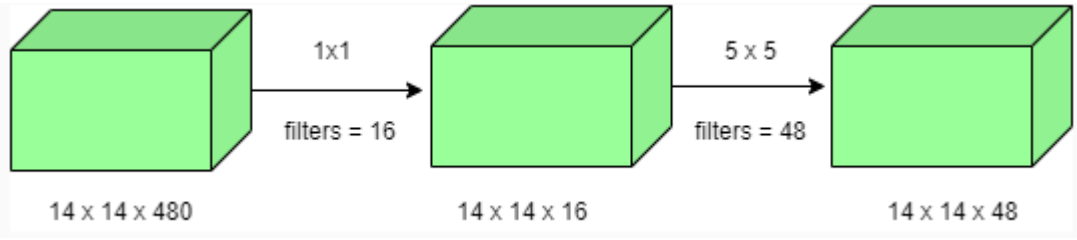
- ***Convolution 1 × 1 :*** L'architecture de démarrage utilise la *convolution 1 × 1* dans son architecture. Ces convolutions permettent de diminuer le nombre de paramètres (poids et biais) de l'architecture. En réduisant les paramètres, nous augmentons également la profondeur de l'architecture. Regardons un exemple deconvolution *1×1 ci-dessous :*

  Par exemple, si nous voulons effectuer une convolution *5 × 5* avec 48 filtres sans utiliser de convolution *1 × 1* comme intermédiaire :



Nombre total d'opérations : *(14 x 14 x 48) x (5 x 5 x 480) = 112,9 M*

Avec convolution 1×1 :



(14 x 14 x 16) x (1 x 1 x 480) + (14 x 14 x 48) x (5 x 5 x 16) = 1,5 M + 3,8 M = 5,3 M, ce qui est beaucoup plus petit que 112,9 M.

2.17-Global Average Pooling

Dans l'architecture précédente comme AlexNet, les couches entièrement connectées sont utilisées à la fin du réseau. Ces couches entièrement connectées contiennent la majorité des paramètres de nombreuses architectures, ce qui entraîne une augmentation du coût de calcul. Dans l'architecture GoogLeNet, il existe une méthode appelée mise en commun des moyennes globales qui est utilisée à la fin du réseau. Cette couche prend une carte d'entités de 7×7 et la moyenne à 1×1 . Cela réduit également le nombre de paramètres pouvant être entraînés à 0 et améliore la précision du top 1 de 0,6 %.

de démarrage : Le module de démarrage est différent des architectures précédentes telles qu'AlexNet, ZF-Net. Dans cette architecture, il existe une taille de convolution fixe pour chaque couche.
Dans le module Inception , la convolution $1 \times 1, 3 \times 3, 5 \times 5$ *et la mise en commun maximale* 3×3 effectuées de manière parallèle à l'entrée et à la sortie de celles-ci sont empilées pour générer la sortie finale. L'idée derrière ces filtres de convolution de différentes tailles gérera mieux les objets à plusieurs échelles.

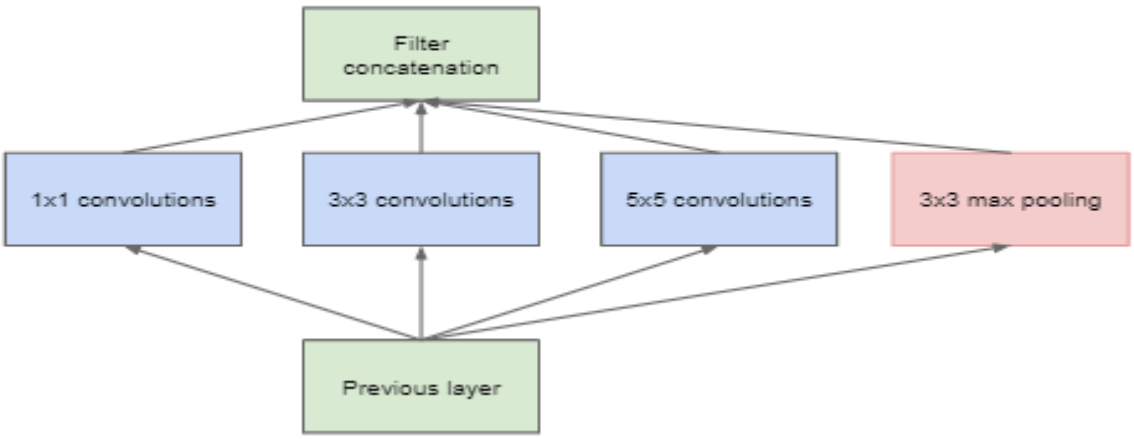


Figure2.14 :inception module,naive version

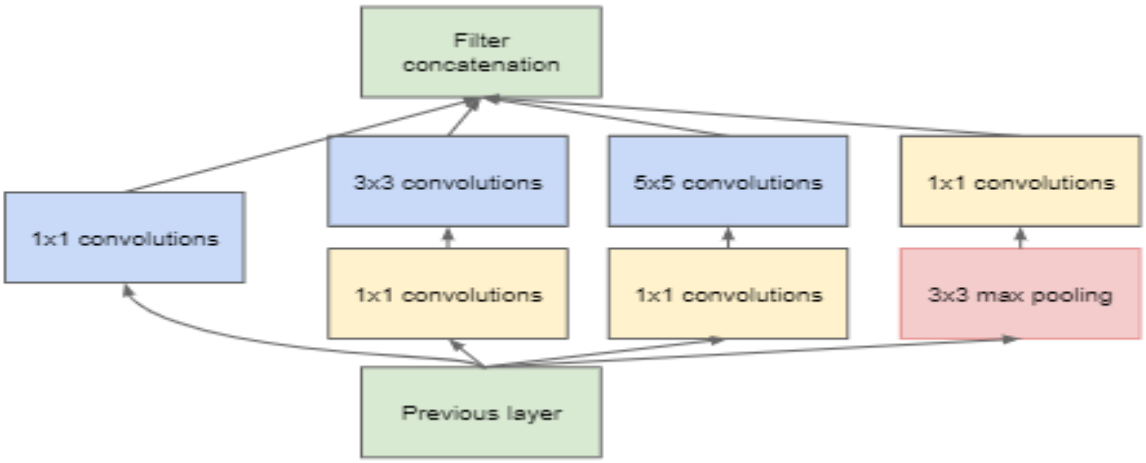


Figure2.15 :Inception module dimension reductions

2.18-Classificateur auxiliaire pour la formation

Architecture de démarrage a utilisé des branches de classificateur intermédiaires au milieu de l'architecture, ces branches sont utilisées uniquement pendant la formation. Ces branches consistent en une couche de regroupement moyenne 5 × 5 avec une foulée de 3, une convolution *1 × 1 avec 128* filtres, deux couches entièrement connectées de 1024 sorties et 1000 sorties et une couche de classification softmax. La perte générée de ces couches s'ajoute à la perte totale avec un poids de 0,3. Ces couches aident à lutter contre le problème de disparition du gradient et assurent également la régularisation.

2.19-Modèle d'architecture

les détails architecturaux couche par couche de GoogLeNet. L'architecture globale est profonde de 22 couches. L'architecture a été conçue pour garder à l'esprit l'efficacité des calculs. L'idée sous-jacente que l'architecture peut être exécutée sur des appareils individuels même avec de faibles ressources de calcul. L'architecture contient également deux couches de classificateurs auxiliaires connectées à la sortie des couches Inception (4a) et Inception (4d).

Les détails architecturaux des classificateurs auxiliaires sont les suivants :

- Une couche de regroupement moyenne de taille de filtre 5 × 5 et de foulée 3.
- Une convolution 1 × 1 avec 128 filtres pour la réduction de dimension et l'activation ReLU.

- Une couche entièrement connectée avec 1025 sorties et activation ReLU
- Régularisation des abandons avec taux d'abandon = 0,7
- Un classificateur softmax avec une sortie de 1000 classes similaire au classificateur softmax principal.

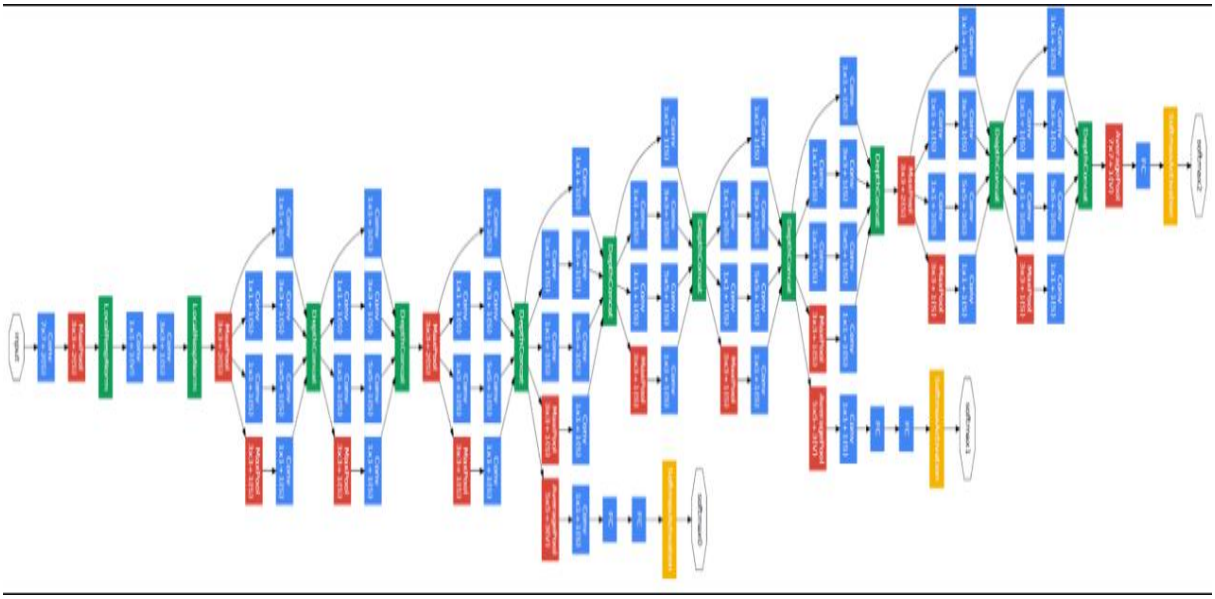


Figure2.16 :GoogleNet –CNN Architecture

Cette architecture prend une image de taille *224 x 224* avec des canaux de couleur RVB. Toutes les convolutions à l'intérieur de cette architecture utilisent des unités linéaires rectifiées (ReLU) comme fonctions d'activation.

	GoogleNet	GoogleNet++
Conv1	3 224 224	3 224 224
Pool1	64 212 112	64 112 112
Conv2	64 56 56	64 56 56
Pool2	192 56 56	192 56 56
Inception3a	192 28 28	192 28 28
Inception3b	256 28 28	256 28 28
Pool3	480 28 28	480 28 28
Inception4a	480 14 14	480 14 14
Inception4b	512 14 14	512 14 14
Inception4C	512 14 14	512 14 14
Inception4d	512 14 14	512 14 14
Inception4e	528 14 14	528 14 14
Pool4	832 14 14	832 14 14
Inception5a	832 7 7	832 7 7
Inception5b	832 7 7	832 7 7
Pool5	1024 7 7	1856 7 7
Fc	1024 7 7	1856 1 1
Prob	1000	1000

Tableau2.1:Taille d'entrée des couches de GoogLeNet et GoogLeNet++

2.20-Conclusion :

Dans ce chapitre on a présenté les notions importantes qui sont en relation avec l'apprentissage profond (définition, Architectures…etc). Aussi qu'une vision générale sur l'apprentissage profond, toute on donnant en détail la méthode choisie dans notre travail de recherche qui est le CNN,PUISnous avons présenté les réseaux de neurones convolutionnels. Ces réseaux sont capables d'extraire des caractéristiques d'images présentées en entrée et de classifier ces caractéristiques, Ils sont fondés sur la notion de « champs récepteurs » (receptivefields), ils implémentent aussi l'idée de partage des poids qui permettant de réduire beaucoup de nombre de paramètres libres de l'architecture. Ce partage des poids permet en outre de réduire les temps de calcul, l'espace mémoire nécessaire, et également d'améliorer les capacités de généralisation du réseau[17].

Le prochain chapitre, traite les détails de la conception, ainsi que la méthode et les outils utilisés pour la réalisation de notre application.

Chapitre III:

Implementation
Et résultats

3.1- Introduction :

Dans ce chapitre, nous allons présenter les résultats de notre approche basée sur l'utilisation des réseaux convolutionnels avec l'architecture proposée par google net,et nous allons faire la conception de notre application Nous allons presenter aussi la mise en oeuvre de notre application en utilisant le langage Matlab. On commençant tout d'abord par une présentation du langage de programmation choisi. Ensuite nous présentons des captures d'écran de l'exécution (résultat obtenir) de notre application.

3.2-Présentation des outils de développement :

3.2.1- Matériel :

La configuration du matériel utilisé dans notre implémentation est :

- Un PC portable intel R core TM2 Duo P7350 CPU 2.00 GHZ
- Nom de lappareil DESKTOP NL2G8MP
- RAM de taille 3 GO
- Disque dur de taille 117 GO
- Système d'exploitation 64 bits Processeur 64

3.2.2- Matlab :

MATLAB_ (Matrix LABoratory) est un logiciel interactif basé sur le calcul matriciel. Il est 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 ce 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". [30]

Son objectif, par rapport aux autres langages, est de simplifier au maximum la transcription en langage informatique d'un problème mathématique, en utilisant une écriture la plus proche possible du langage naturel scientifique.

Les ingénieurs et les scientifiques utilisent MATLAB dans de nombreux domaines, tels que le traitement des images et des signaux, les communications, les systèmes de contrôle du secteur industriel, la conception de réseaux intelligents, la robotique et la finance computationnelle.

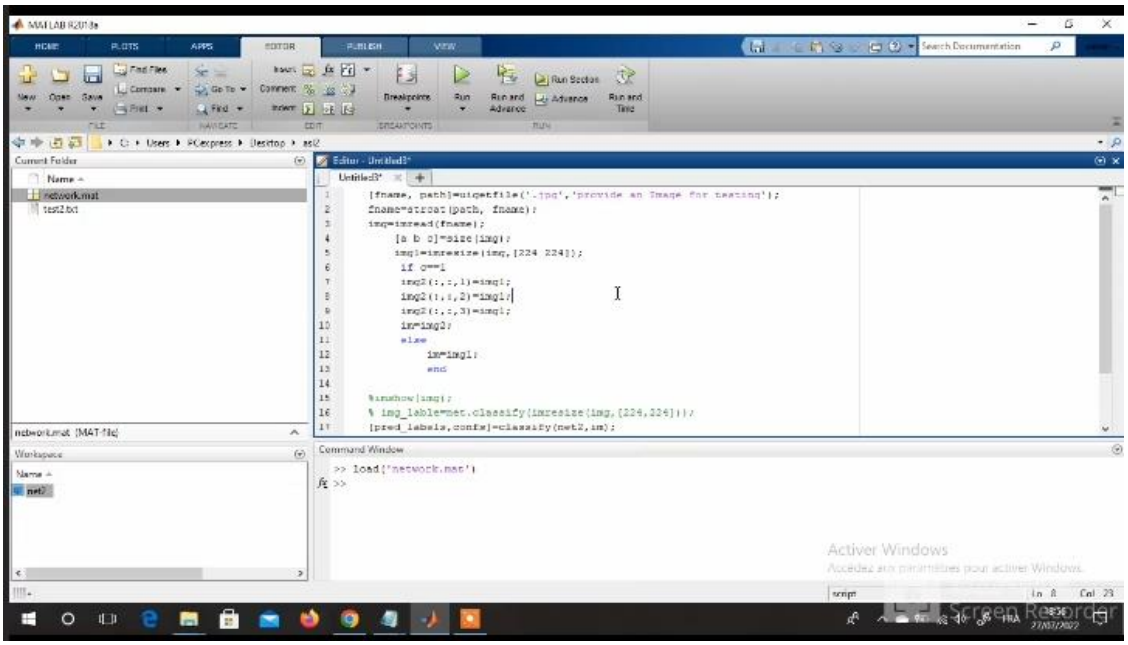


Figure 3.1-L` environnement.Matlab.

3.2.3: Outils Toolbox

Deep Learning Toolbox fournit un cadre pour la conception et la mise en œuvre de réseaux de neurones profonds avec des algorithmes, des modèles pré-entraînés et des applications. Vous pouvez utiliser des réseaux de neurones convolutifs (ConvNets, CNN) et des réseaux de mémoire longue à court terme (LSTM) pour effectuer une classification et une régression sur des données d'image, de séries chronologiques et de texte. Vous pouvez créer des architectures de réseau telles que des réseaux antagonistes génératifs (GAN) et des réseaux siamois à l'aide de la différenciation automatique, de boucles d'entraînement personnalisées et de pondérations partagées. Avec l'application Deep Network Designer, vous pouvez concevoir, analyser et former des réseaux graphiquement. L'application Experiment Manager vous aide à gérer plusieurs expériences d'apprentissage en profondeur, à suivre les paramètres d'entraînement, à analyser les résultats et à comparer le code de différentes expériences. Vous pouvez visualiser les activations de couches et surveiller graphiquement la progression de la formation.

3.3-Dataset:

Référentiel GitHub pour la langue des signes à la parole : non exprimé [https://github.com/grassknoted/Unvoiced], [31]

Ils sont disponibles sur lien suivant:[32]

https://www.kaggle.com/datasets/grassknoted/asl-alphabet?select=asl_alphabet_train

3.4-la base données

L'ensemble de données est une collection d'images d'alphabets de la langue des signes américaine, séparés en 29 dossiers qui représentent les différentes classes.

L'ensemble de données d'apprentissage contient 87 000 images de 200 x 200 pixels. Il existe 29 classes, dont 26 pour les lettres A-Z et 3 classes pour SPACE, DELETE et NOTHING.[33]

Ces 3 classes sont très utiles dans les applications en temps réel et la classification.

figure3.2:Base d'image asl_alphabet_test

Les données sont disponible sur le site Github pour la langue des signes et la parle et Kaggle[31] ,[32]

3.5- Présentation de l'application :

Nous allons maintenant présenter les résultats obtenus grâce à des expériences réalisées par l'application

On utilise deep learning et la méthode de classification CNN.

Le schéma de notre système est le suivant :

1. Nous avons utilisé Google Note un réseau de neurones pré-formé sur La base de données contenant des images en plusieurs classes comme exemple (class animaux, class fruit, class informatique, transport,….etc)

2. nous avons pris ce réseau de neurones et nous avons utilisé Transfer learning .

pour le Transfer learning nous avons supprimé les trois dernières couches du réseau de neurones:

- **Connected layer fully** : Au sommet du réseau neuronal se trouve la couche entièrement connecrée.
- **softmax layer**: Il aide à prendre les informations de l'entrée et à les organiser d'une manière facile à comprendre pour un humain.

- **classification layer**: Cette couche utilise un ensemble de règles prédéfinies et filtre les données en fonction de ces règles, classant les données en groupes. Le but de cette couche est de fournir un contexte et une structure supplémentaires aux données non structurées, ce qui peut être difficile à comprendre pour une machine seule.

3. Nous les avons remplacées avec d'autres couches afin de pouvoir les entraîner sur un autre base de donnée, la base de données des signes alphabets.

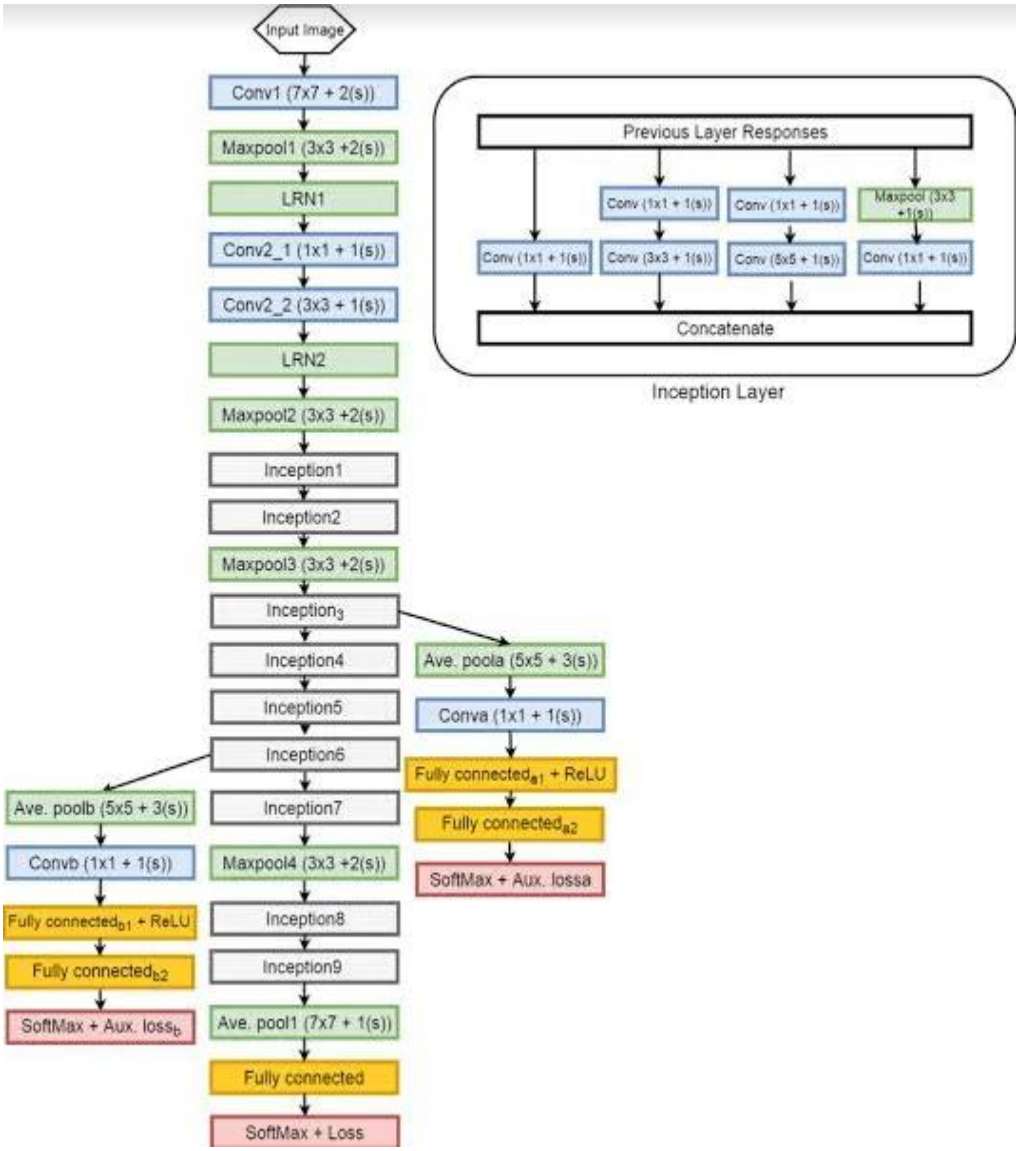


Figure3.3 : le modele googlenet architecture cnn

3.6- les étapes de creation du code principal

- appeler le réseau de neurones

```
net=googlenet;
```

- Créer un magasin de données (images d'entraînement)

```
images=imageDatastore('E:\archive2\asl_alphabet_train\asl_alphabet_train','IncludeSubfolders',true,'LabelSource','foldernames');
```

- Redimensionner les images de 300*300 à 224*224*3

```
for i=1:length(images.Labels)
    img=readimage(images,i);
    [a b c]=size(img);
```

```
    img1=imresize(img,[224 224]);
    if c==1
    img2(:,:,1)=img1;
    img2(:,:,2)=img1;
    img2(:,:,3)=img1;
    im=img2;
    else
       im=img1;

    end

    imwrite(im,cell2mat(images.Files(i)))
end
```

- Extraction de la structure du réseau

```
graph=layerGraph(net);
```

- Calculer le nombre de catégories existantes (28 caractères)

```
numClasses=numel(categories(images.Labels));
```

- Supprimez les trois dernières couches (Connected layer fully ,softmax layer ,classification layer)

```
graph=removeLayers(graph,{'loss3-classifier','prob','output'});
```

- Créer trois nouveaux couches

```
newLayers=[
fullyConnectedLayer(numClasses,'name','fullyConnected','WeightLearnRateFactor',16,'BiasLearnRateFactor',16),
softmaxLayer('Name','softmax'),
classificationLayer('Name','output')
];
```

- Ajout de nouvelles couches à la structure

```
graph=addLayers(graph,newLayers);
```

- Connecter de nouvelles couches avec des couches existantes dans la structure

```
graph=connectLayers(graph,'pool5-drop_7x7_s1','fullyConnected');
```

- Définir les paramètres d'entraînement

```
options=trainingOptions('sgdm',...
'MaxEpochs',100,...
'MiniBatchSize',10,...
'plots','training-progress',...
'InitialLearnRate',0.001);
```

- le processus de tri commence sur la banque de donnée d'images contenant les images

```
net2=trainNetwork(images,graph,options);
```

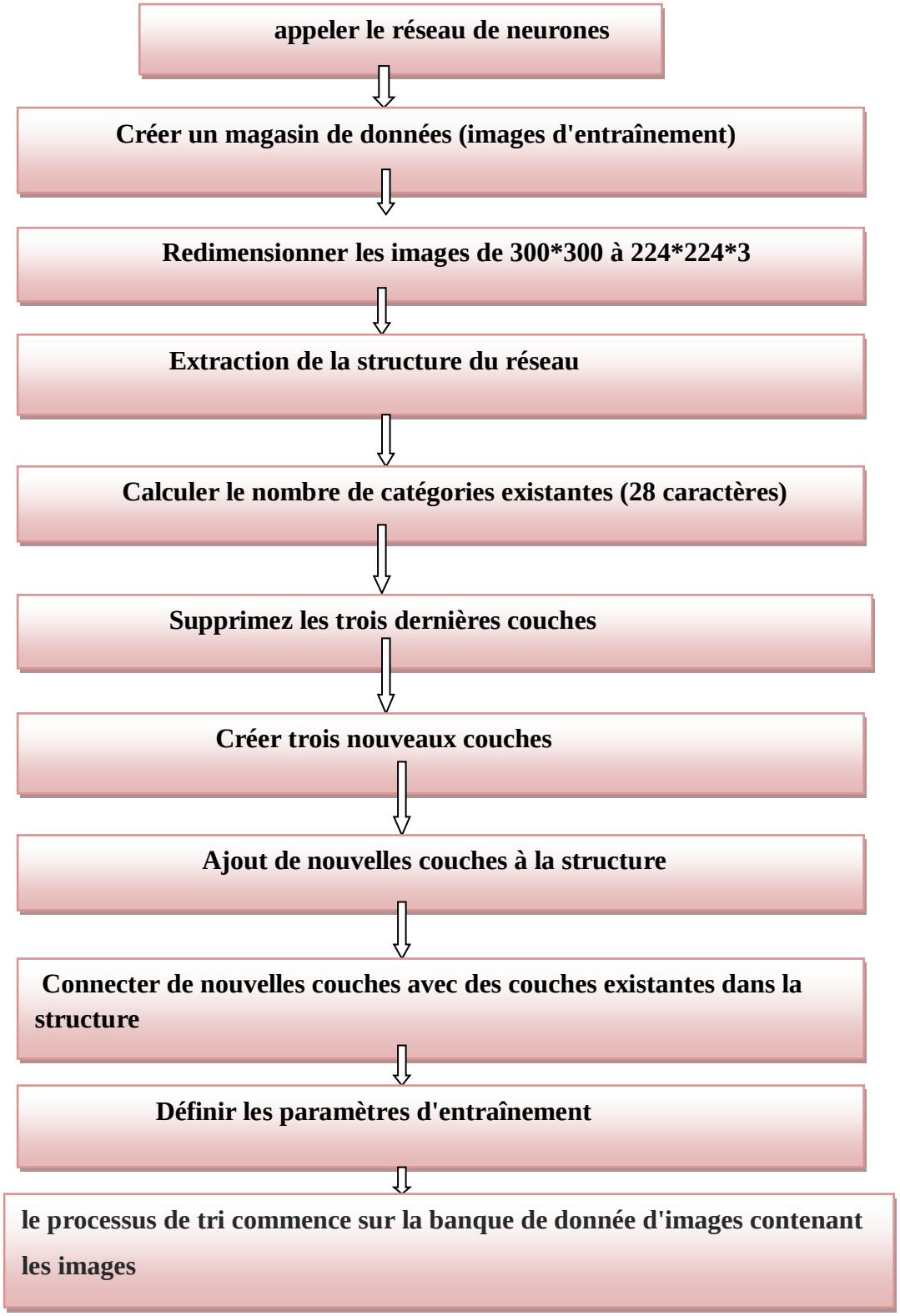


Figure3.4 : Organigramme des étapes

3.7- CNN - une couche de convolution et une couche entièrement connectée

3.7.1 - Étape 1 : Préparation de l'image étiquetée

Pour appliquer un réseau de neurones pour la classification contenant les symboles de la main pour les lettres des alphabets (un type d'application de classification d'images), vous devez d'abord préparer un ensemble d'images étiquetées. Cela peut être un bon exemple de diverses préparations de données que j'ai mentionnées dans la page de ce qu'ils font. Il peut y avoir différentes façons de préparer les données étiquetées pour la formation, mais la façon dont nous prendons est la suivante.

i) Tout d'abord, nous avons créé 28 dossiers avec le nom du dossier correspondant à l'étiquette des images stockées dans le dossier. c'est-à-dire que nous avons créé 28 dossiers avec le nom A, B, C, D, E, F,G, H, I, G, K ,L, M, N , nothing- test,O ,P, Q, R, S, space- test, T, U ,V, W, X ,Y, Z, comme indiqué ci-dessous.

ii) Ensuite, nous avons mes les images avec des alphapets écrits (figure3.5).

NOTE 1 : Pour plus de simplicité, tous les fichiers crées ont des image avec la même taille. En situation réelle, vous pouvez obtenir toutes ces images avec des tailles différentes.

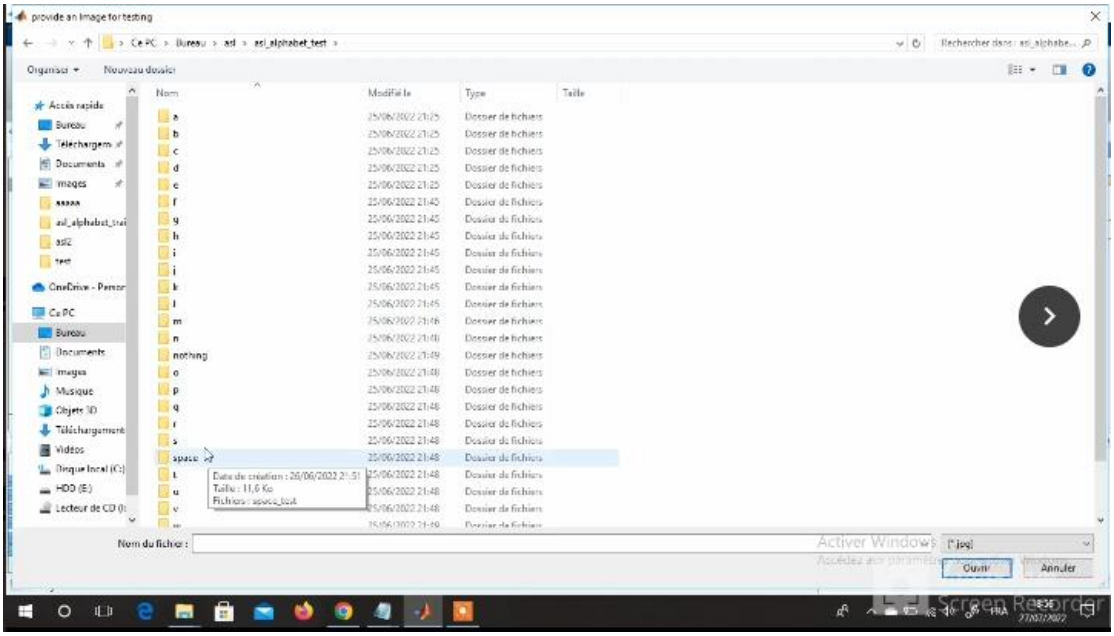


Figure3.5:les dossiers avec le nom

Nous cliquons sur le dossier que nous voulons choisir et l'ouvrons puis l'image dans le dossier s'affichera comme suit :

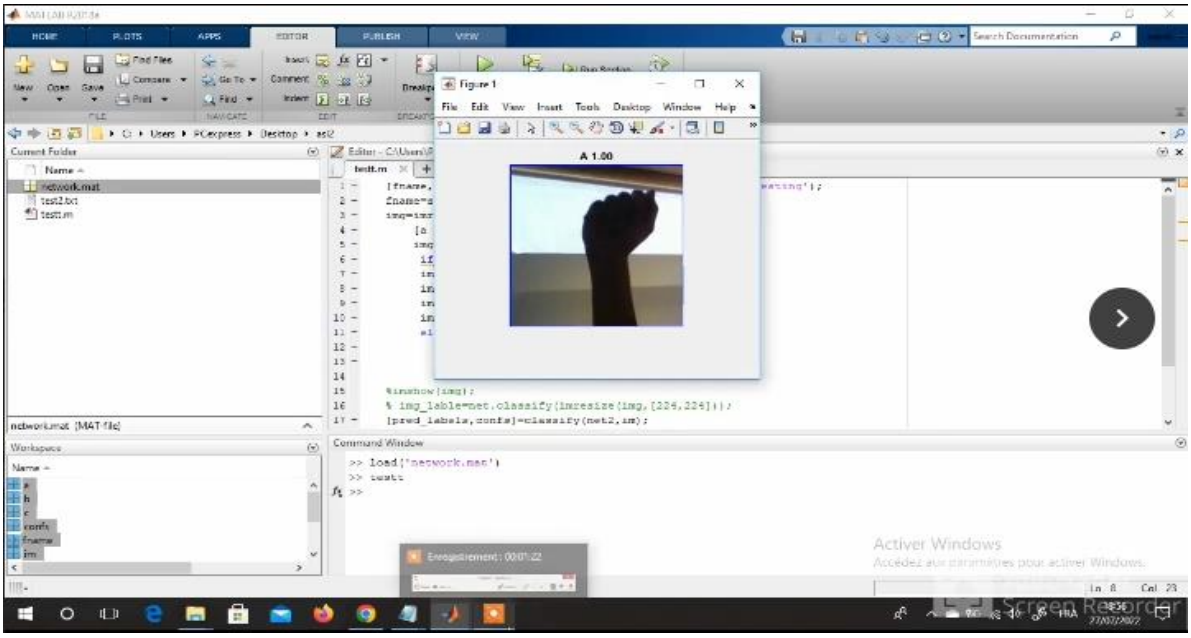


Figure3.6:le résultat du traitement de limage contenant le symbole de la main pour la lettre" A"

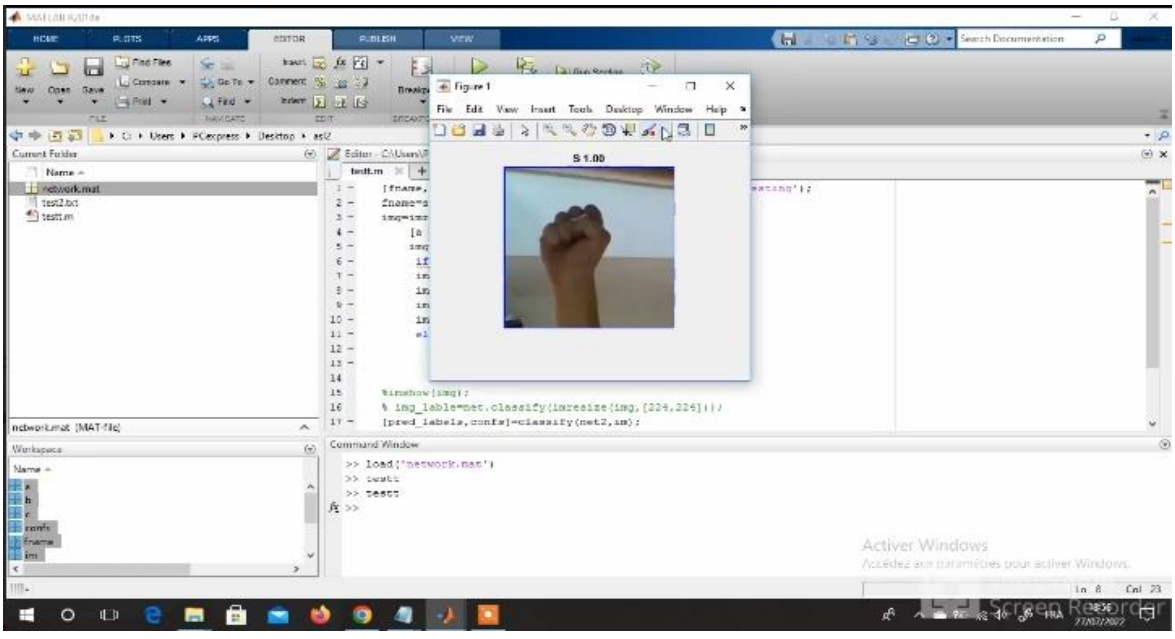


Figure3.7:le résultat du traitement de l'image contenant le symbole de la main pour la lettre" S «

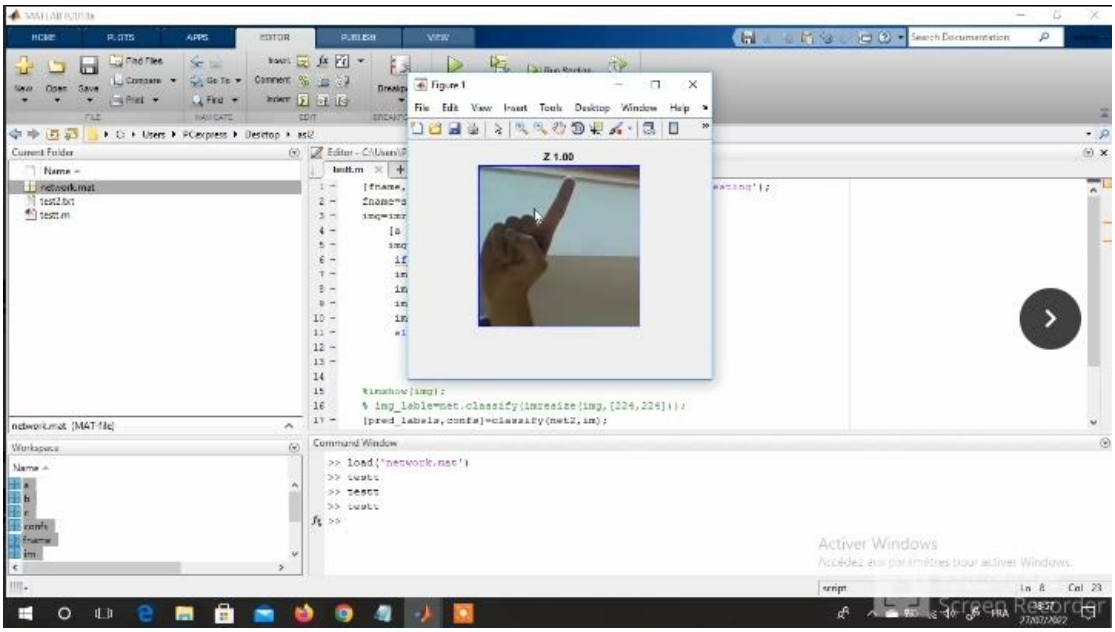


Figure3.8:le résultat du traitement de limage contenant le symbole de la main pour la lettre" Z"

Donc on continue de la même méthode pour toutes les lettres.

3.7.2- Étape 2 : Lire l'image étiquetée et générer les données d'entraînement

Dans cette étape, Nous utilisons une fonction spéciale appelée fonction imageDastor() qui lit essentiellement tous les fichiers image dans tous les dossiers et les enregistre dans un variable sans utiliser de boucle for ou while (bien sûr, en interne Dans cette fonction, il y aura des boucles en cours d'exécution pour prédire les images un par un comme suit:

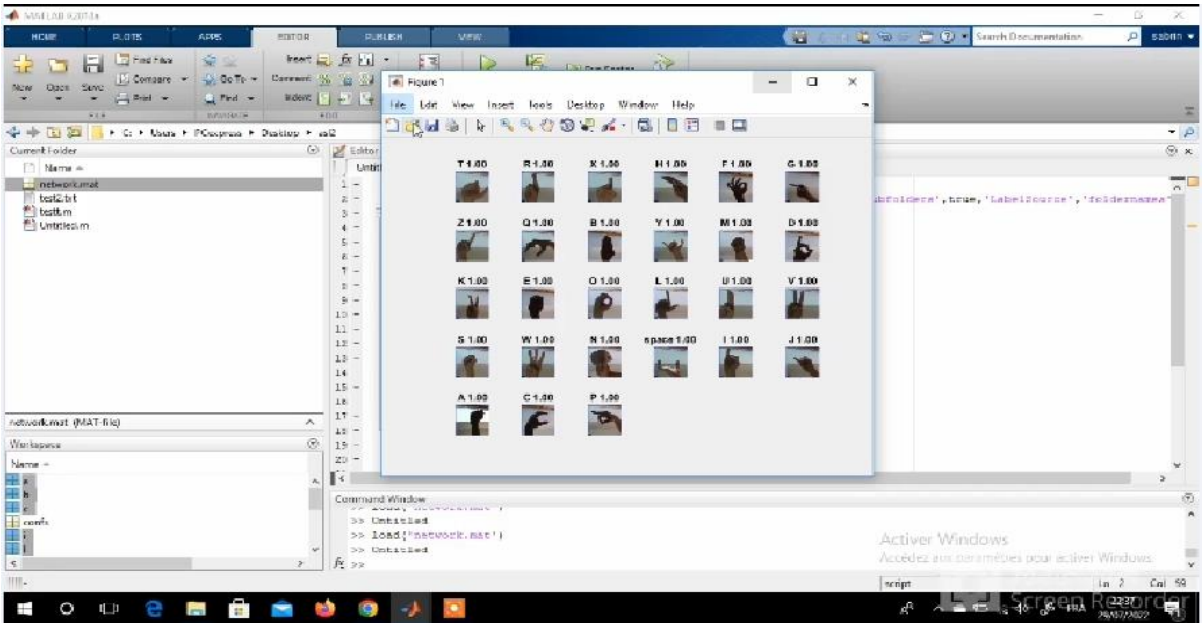


Figure3.9:les résultat affichée pour toue les image

3.7.3 :Code pour l'utilisation de image

```
net=load('network.mat');
images=imageDatastore('D:\archiv\asl_alphabet_test\asl_alphabet_test','IncludeSubfolders',tru
e,'LabelSource','foldernames');
 for i=1:length(images.Labels)
    img=readimage(images,i);
    [a b c]=size(img);
    img1=imresize(img,[224 224]);
    if c==1
    img2(:,:,1)=img1;
    img2(:,:,2)=img1;
    img2(:,:,3)=img1;
    im=img2;
    else
       im=img1;

    end

    imwrite(im,cell2mat(images.Files(i)))
```

```
 end
n=numel(images.Labels);
id=randperm(n,27);
figure;
for i=1 : 27


    subplot(5,6,i);
    I=readimage(images,id(i));
    [pred_labels,confs]=classify(net2,I);
    imshow(I);
    title(sprintf('%s %.2f',char(pred_labels),max(confs)))
end
```

Figure3.10: les images obtenues avec une prédiction plus précise

3.8- Discussion des résultats

Une fois que vous avez remarqué les résultats obtenus en appliquant l'algorithme CNN via le programme Matlab, Loss et précision et discuter de lettres ont été reconnues à travers des images qui contiennent un symbole de main qui symbolise à chaque fois une lettre spécifique, que ce soit en expérimentant test avec une image ou plusieurs images avec une prediction précise.

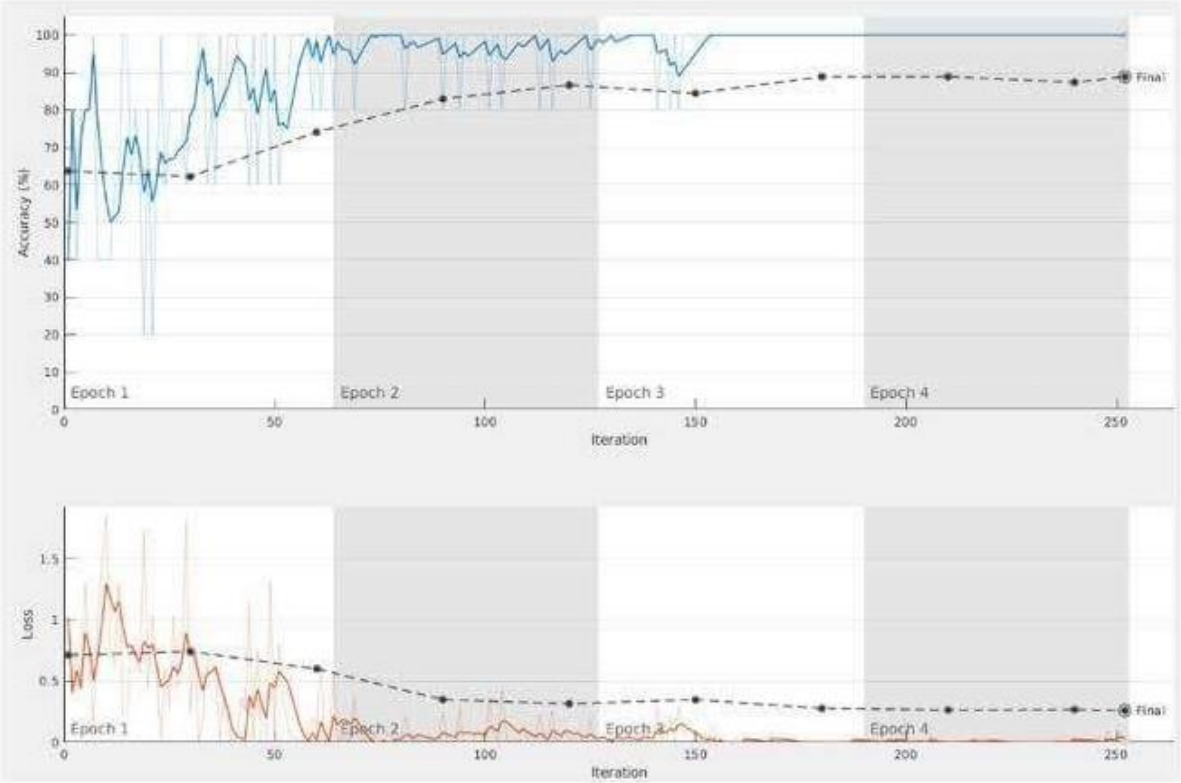


Figure3.11 : Graphique de validation et de perte de la formation CNN

Comment mesurer l'exactitude et la précision

- Si vous souhaitez calculer la perte pour chaque époque, divisez le running_loss par le nombre de lots et ajoutez-le à train_losses à chaque époque. La précision est le nombre de classifications correctes / le nombre total de classifications.
- Valeur moyenne = somme des données / nombre de mesures.
- Écart absolu = valeur mesurée - valeur moyenne.
- Ecart moyen = somme des écarts absolus / nombre de mesures.

- Erreur absolue = valeur mesurée - valeur réelle.
- Erreur relative = erreur absolue / valeur mesurée.

3.9-Conclusion

Après avoir achevé notre conception nous avons donné les outils nécessaires pour la réalisation de notre travail. Nous avons présenté aussi l'environnement de développement. A la fin nous avons présenté notre application en donnant quelques captures d'écran qui expliquent le déroulement et le fonctionnement de notre travail, ainsi que les resultants obtenus.

On a donné un exemple de classification par notre application d'un ensemble d'images qui sont déjà classifiée au paravent.

Nous avons montré l`impact de la méthode choisi (CNNs) sur les résultats de la classification des signes de la langue des et la la reconnaissances.

Conclusion général

La classification d'images est une tâche importante dans le domaine de la vision par ordinateur, la reconnaissance d'objets et l'apprentissage automatique. Bien que les capacités des activités réalisées 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 des 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 en un ensemble de classes (les gestes de la main).

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 CNNs 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 et/ou d'hybridation avec d'autres méthodes de classification.

Références

[1] https ://fr.wikipedia.org/wiki/Langue-des-signes-francaise.

[2] : Aliyah Morgenstern «Introduction à la langue des signes française, La place du Sourd et de sa langue en France »

[3] : Bossard Bruno « Problèmes posés par la reconnaissance de gestes en Langue des Signes» Université Paris XI, juin 2002.

[4] : Cuxac C «Autour de la langue des signes » vol.10 Université de Paris1983.

[5] YannB,NicolasP.Rougier, Apprentissage et Memoires,Introduction aux reseaux de neurones artificiels,Janvier,Master 1-Sciences Cognitives Université De Lorraine,janvier2012

[6] Gag, Aired.CONVERTING AMERICAN SIGN LANGUAGE TO VOICE USING RBFNN. s.l. ,A Thesis Presented to the Faculty of San Diego State University in Partial Fulfillment of the Requirements for the Dergree :Master of Science in Computer Science,Summer 2012.

[7] Rafiqul Z. Khan, Noor A. Ibraheem. 4,Hand Gesture Recognition : A Literature Review, International Journal of Artificial Intelligence & Applications And Information Science,(IJAIA), Vol. 3, 2012.

[8] G. R. S. Murthy, R. S. Jadon ,A Review of Vision Based Hand Gestures Recognition, International Journal of Information Technology and Knowledge Management, Vol. 2(2), pp. 405-410, 2009.

[9] https ://docs.gimp.org/fr/plug-in-convmatrix.html.

[10] http ://xphilipp.developpez.com/articles/filtres 2012.

[11] https ://fr.wikipedia.org/wiki/Segmentation-d-image.

[12] Simon conseil , suivi tridimensionnel l de la main et reconnaissance de gestes pour les interfaces homme –machine, thèse pour obtenir le grade de docteur de l'université Paul Cézanne faculté des sciences et techniques, 13 mars 2008.

[13] M BICHSEL, éditeur. Proceedings of the international Worksop on Automatic Face and Gesture Recognition(IWAFGR'95),Zurich, Switzerland, juin 1995.

[14] Dr. John, C. Russ , The Image Processing and measurement Cookbook, 2000.

[15] http://www.techno-science.net

[16] https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501333-deep-learning-definition-et-principes-de-l-apprentissage-profond/

[17] L. Deng, D. Yu, et al., "Deep learning : methods and applications," Foundations and Trends R in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014. 73

[18] Y. Bengio et al., "Learning deep architectures for ai," Foundations and trends R in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.

[19] L. Deng, D. Yu, et al., "Deep learning : methods and applications," Foundations and Trends R in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.

[20] https://fr.wikipedia.org/wiki/Apprentissage_supervis%C3%A9

[[21] J. Schmidhuber, "Deep learning in neural networks : An overview," Neural networks, vol. 61, pp. 85– 117, 2015.

[22] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in Advances in neural information processing systems, pp. 2643–2651, 2013.

[23] R. Collobert and J. Weston, "A unified architecture for natural language processing : Deep neural networks with multitask learning," in Proceedings of the 25th international conference on Machine learning, pp. 160–167, ACM, 2008.

[24]http://geekflare.com

[25]https://actualiteinformatique.fr

[26]wikipedia.org]

[27] https://fr.acervolima.com › comprendre-le-modele-goo...

[28]http://www.reseurchgate.net

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

[30]. A.Seghir. méthode des éléments finis. s.l. : Département de génie civil, mémoire magister, Université A.Mira de Bejaia.

[31] [https://github.com/grassknoted/Unvoiced],

[32] https://www.kaggle.com/datasets/grassknoted/asl-alphabet?select=asl_alphabet_train