الجمـــهورية الـــــجزائرية الديمقـــــراطية الشعـــــــبية
République Algérienne Démocratique et Populaire
وزارة التعليـــــــــم العالـــــــــي والبحث العلمـــــــي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 20 Août 1955 – Skikda
Faculté des Sciences
Departement d'informatique

جـامعة 20 أوت 1955-سكيكدة
كـــــلية العلـــــوم
قسم الإعـــــلام الآلـــي

Mémoire de fin d'étude en vue de l'obtention du diplôme de Master en Informatique

Option : Systèmes Informatiques/Réseaux et Systèmes Distribués

Thème

Une approche pour la détection des émotions à partir de l'expression faciale en utilisant la technologie Yolo v4

Réalisé par les étudiantes :

- Amani Guetteche
- Chaima Bouchareb

Dirigé par :

Mr. Lazhar Benoudina

Année Universitaire 2021-2022

Remerciements

Nous remercions d'abord le bon Dieu qui nous a aidé et qui nous a donné le courage et la patience pour réaliser ce modeste travail.

Nous continuerons à remercier nos enseignants de l'université 20 Aout 1955 Skikda.

En particulier, nous remercions notre encadreur monsieur **Benoudina Lazhar** qui était toujours à notre disposition sans réservation pendant la préparation de ce mémoire de fin d'études.

Nos remerciements vont aussi aux membres de jury pour avoir accepté de juger ce modeste travail.

Nous remercions aussi Mr **Mazouzi Smaine** pour son aide.

On n'oublie pas nos parents pour leur contribution, leur soutien et leur patience. Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis. Qui nous ont toujours soutenue et encouragées au cours de la réalisation de ce travail.

Dédicaces

Pour chaque début une fin, et ce qui beau dans toute fin c'est la réussite et l'atteindre du but.

Tout d'abord, je rends grâce à ***Dieu*** *le tout puissant de m'avoir donnée le courage, la santé, et la volonté de mener à termine ce modeste travail.*

Je dédie ce mémoire à :

Mes chers parents :

Ma mère, *qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.*

Mon père : *qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie, Puisse Dieu faire en sorte que ce travail porte son fruit : merci pour les valeurs nobles, l'éducation et le soutient permanent venu de toi.*

Mes chères sœurs (Soumia, Meriem, Ibtissem, Hadjer) *pour leur soutien et leur présence dans ma vie.*

Mes chères neveu et nièce (Sief Elislem, Sidra) *qui j'aime beaucoup.*

Mes tantes et mes oncles *surtout mon oncle «* ***Sebti*** *» que je considère comme deuxième père pour moi.*

Mes amis (Roukaya, Iman, Chaima, Maroua) *qui nous marché ensemble en ouvrant ensemble la voie vers le succès et la créativité à qui nous nous joints main dans la main en cueillant une fleur, nous avons appris.*

Mes professeures*, qui m'ont appris des lettres d'or et des mots de perles et des phrases des expressions les plus hautes et les plus vénérées de la connaissance à ceux qui ont conçu pour moi des lettres de savoir et de leurs pensées comme un phare illuminant le chemin dès la connaissance pour nous et succès.*

Finalement, en espérant que cet humble travail trouvera acceptation et succès.

Guetteche Amani

Dédicaces

Je dédie ce travail

A l'homme, mon précieux offre du dieu qui doit ma vie, ma réussite et tout mon respect mon cher père.

A la femme qui souffert sans me laisser souffrir, qui n'a jamais dit non âmes exigences, qui m'a soutenu et encouragé durant ces années d'études. Qu'elle trouve ici le témoignage de ma profonde reconnaissance mon adorable mère.

A mon frère Mohamed et mes belles sœurs (Amani, Hanen, Ines) qui ont partagé avec moi tous le moment d'émotion lors la réalisation de ce travail. Ils m'ont chaleureusement supporté et encouragé tout au long de mon parcours.

A tous mes amis qui m'ont toujours encouragé, et à qui je souhaite plus de succès.

Sans oublier mon binôme pour son soutien moral, sa patience et sa compréhension tout au long de ce projet.

A tous ceux que j'aime

Bouchareb Chaima

ملخص

اليوم، يثبت التعرف التلقائي على المشاعر أنه أحد أكثر التطبيقات ذات الصلة في العديد من المجالات، وهي: التفاعل بين الإنسان والحاسوب، وعلم النفس، والطب، والتعليم، . . . إلخ بالإضافة إلى ذلك، كان نهج التعلم العميق والشبكات العصبية التلافيفية بشكل خاص (CNN) ناجحًا للغاية في مجال معالجة الصور والتعرف عليها. الغرض من هذا العمل هو البحث وتطبيق خوارزميات التعلم العميق للكشف عن المشاعر لمراقبة إنتاجية الموظف، ومرضى التوحد.

ركزنا في هذه الدراسة على خوارزمية YOLO للكشف عن الأشياء، وهي طريقة معترف بها ومعتمدة عمليًا.

لذلك، حاولنا إعادة تدريب نموذج YOLO على الفئات الفرعية لقاعدة البيانات. لقد حصلنا على نتائج مرضية للغاية بمعدل دقة 86٪، مع نظرة عامة تعكس التفوق الكبير لخوارزمية اكتشاف الأشياء في الوقت الحقيقي YOLO على العديد من الطرز الأخرى.

الكلمات الرئيسية: التعرف التلقائي على المشاعر، الشبكات العصبية التلافيفية (CNN)، التعلم العميق، YOLO.

Résumé

Aujourd'hui, la reconnaissance automatique des émotions s'avère être l'une des applications les plus pertinentes dans de nombreux domaines à savoir : Interaction homme-machine, psychologie, médecine, éducation, . . . etc. En plus, l'approche de deep Learning et plus particulièrement les réseaux de neurones convolutionels (CNN) ont connu un grand succès dans le domaine du traitement et de la reconnaissance d'images. Le but de ce travail est de rechercher et d'appliquer des algorithmes de détection des émotions d'apprentissage en profondeur pour suivi la productivité des employées, et les patients autistes.

Nous nous sommes concentrés dans cette étude sur l'algorithme de détection d'objet YOLO, qui est une méthode largement reconnue et approuvée.

Par conséquent, nous avons essayé de ré-entraîner le model YOLO sur des sous classes de la base de données. Nous avons obtenu des résultats très satisfaisants avec un taux de précision 86%, avec un aperçu qui reflète la grande supériorité de l'algorithme de détection d'objet YOLO en temps réel sur beaucoup d’autres modèles.

Mots clés : Reconnaissance automatique des émotions, Réseaux de neurones convolutif (CNN), Apprentissage en profondeur (Deep Learning), YOLO.

Abstract

Today, automatic emotion recognition is one of the most relevant applications in many fields such as: Human-Computer Interaction, psychology, medicine, education, etc. In addition, the deep learning approach and more particularly the convolutional neural networks (CNN) have been very successful in the field of image processing and recognition. The aim of this work is to research and apply deep learning object detection algorithms to monitor the productivity of employees, and autistic patients.

In this study, we focused on the YOLO object detection algorithm, which is a practically recognized and approved method.

Therefore, we tried to re-train the YOLO model on subclasses of the database. We obtained very satisfactory results with an accuracy rate of 86%, with an insight that reflects the great superiority of the YOLO real-time object detection algorithm over many other models.

Keywords: Automatic emotion recognition, Convolutional neural networks (CNN), Deep Learning, YOLO.

Table des matières

Table des figures

Liste des tableaux

Introduction générale

Introduction générale

Le visage étant la partie la plus expressive et communicative d'un être humain, parce qu'il peut lui montrer différentes expressions émotionnelles exprimant l'émotion intérieure de la personne. Les expressions faciales provoquent des changements physiologiques sur le visage, tels que la position de la bouche ouverts ou fermes, ou encore la manière de regarder. Il existe six expressions émotionnelles universelles : dégout, colère, bonheur, tristesse, surprise et peur. Ces expressions sont indéniables en observant les signes du visage.

L'analyse des expressions faciales, qui consiste à détecter et analyser les mouvements du visage afin d'en extraire l'émotion exprimée, représente l'une des approches les plus utilisées et les plus performantes. Néanmoins, il est difficile, autant pour l'homme que pour la machine, de différentier certaines émotions très similaires, telles que la peur et la surprise

En Informatique, la reconnaissance des émotions semble être la solution logique à une bonne interaction homme/machine. Dans un monde qui se penche de plus en plus vers les robots intelligents, il s'avère indispensable de doter ces derniers d'une capacité émotionnelle et affective afin d'assurer une interaction optimale avec les utilisateurs.

L'apprentissage automatique est un domaine de l'intelligence artificielle, qui fait référence à la capacité des systèmes informatiques au sein des machines à trouver indépendamment des solutions aux problèmes en percevant différents modèles de données. Parmi les algorithmes qui permettent à la machine d'apprendre par elle-même grâce à l'apprentissage profond en anglais deep learning, c'est la simulation des neurones du corps humain.

La plupart des recherches sur l'apprentissage en profondeur se concentrent sur la recherche de méthodes permettant d'obtenir un degré élevé d'abstraction en analysant un grand ensemble de données à l'aide de variables linéaires et non linéaires. D'où vienne la nécessité d'avoir des bases de données très riches et divers et surtout bien étiquetés.

L'objectif de ce projet de fin d'étude est de concevoir et de réaliser un système permet de reconnaitre les émotions à partir des expressions faciales d'un visage pour suivi la productivité des employées, patients autistes, patients en soins intensifs en utilisant la méthode d'apprentissage profond.

Ce mémoire se présente sous forme de trois chapitres :

Dans **le premier chapitre**, nous présentons quelques notions relatives aux expressions faciales et émotions telles que leurs historiques, leurs définitions, leurs types, et les principales difficultés,

Ensuite, **le chapitre 02** est consacré à l'étude des différentes architectures des méthodes de Deep learning pour la détection d'objets.

Le chapitre 03, nous présentons la conception de notre système de détection avec les résultats expérimentaux obtenus par le modèle YOLOv4 ainsi qu'une discussion avec interprétation des résultats.

Nous terminerons ce mémoire par une conclusion générale et des perspectives.

Chapitre 01
Expression faciale et la reconnaissance des émotions

1. Introduction

Au cours de la dernière décennie, la communauté de la recherche en vision par ordinateur a montré beaucoup d'intérêt pour l'analyse et la reconnaissance automatique des expressions faciales. Cette communauté de la vision par ordinateur et de la recherche scientifique envisageait de développer des systèmes capables de reconnaître les expressions faciales dans des vidéos ou de images, La plus plupart de ces systèmes d'analyse des expressions faciales tentent de classer les expressions en quelques grandes catégories émotionnelles, telles que la joie, la tristesse ; la colère, la surprise, la peur et le dégoût.

Dans ce premier chapitre, nous allons présenter quelques notions relatives aux expressions faciales et émotions.

2. Expression faciale

2.1. Définition

Le visage porte des informations sur l'identité d'une personne telle que la couleur des yeux, la forme de la bouche, la couleur des cheveux ainsi que la forme des oreilles, etc. Comme il peut exprimer aussi des expressions de communication et d'émotion qu'on peut reconnaitre en prêtant attention aux expressions faciales.

L'expression faciale est avant tout un ensemble des signes du visage, perceptible visuellement, dû à l'activation (volontaire ou non) a l'un ou plusieurs des 44 muscles qui composent le visage, qui traduisent un sentiment de changement dans le visage [1].

2.2. Historique :

La reconnaissance faciale est une technique biométrique relativement récente. Si l'empreinte digitale est la technique biométrique la plus ancienne inventée en 1903 pour rechercher les criminels, la reconnaissance des visages a été développé par "Benton et Van Allen" en 1968 pour évaluer la capacité d'identification des visages non familiers. Il ne s'agit pas d'un test de reconnaissance ménisque de visages familiers ou non familiers, mais d'une épreuve consistant à apparier des photographies de visages non familiers présentés sous

différents éclairages et selon des angles différents et nécessitant une bonne capacité d'intégration Visio-spatiale [2].

L'utilisation des techniques de reconnaissance faciale a connu un développement à grande échelle depuis le milieu des années 90, avec l'utilisation efficace de nouvelles technologies, notamment l'ordinateur et sa capacité de traitement d'images. L'utilisation de ces techniques existe depuis qu'une machine est capable de comprendre ce qu'elle « voit » lorsqu'on la connecte à une ou plusieurs caméras, c'est à dire que les premiers essais datent du début des années 70 (Benton et Van Allen en 1968), et sont basés sur des méthodes à bases d'heuristiques, basés sur des attributs faciaux mesurables comme l'écartement des yeux, des sourcils, des lèvres, la position du menton, la forme, etc. Ces méthodes sont très peu robustes, car elles font de nombreuses suppositions en se plaçant dans des cas très simples (visage de face, bonnes conditions d'illuminations, etc. L'une des premières tentatives de reconnaissance de visage est faite par Takeo Kanade en 1973 lors de sa thèse de doctorat à l'Université de Kyoto [3][4].

2.3. Qu'est-ce qu'une expression faciale émotionnelle

Les êtres humains peuvent exprimer leurs _émotions suivant différentes modalités telles

Que les expressions faciales, les expressions corporelles, la prosodie ou encore le langage.

Ces différents modes d'expressions des _émotions ne s'opposent pas et peuvent très bien

Être utilisés simultanément. Nous pouvons par exemple exprimer de la peur avec une voix

Effrayée, un visage effrayé et une expression corporelle de peur. L'incongruité entre ces

Modalités peut entrainer une perturbation de la reconnaissance de l'expression faciale, qui

Se trouve alors biaisée par l'expression émotionnelle corporelle (Meeren et coll., 2005) [5].

2.4. Principales difficultés de la reconnaissance faciale

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représentent un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations

associées aux images de visages : inter et intra sujet [6]. La variation inter-sujette est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra-sujette est plus vaste. Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous.

- **Changement d'illumination**

Les variations d'éclairage rendent la tache de reconnaissance de visages très difficile. En effet, le changement d'apparence d'un visage du a l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entrainer une mauvaise classification des images d'entrée [6].

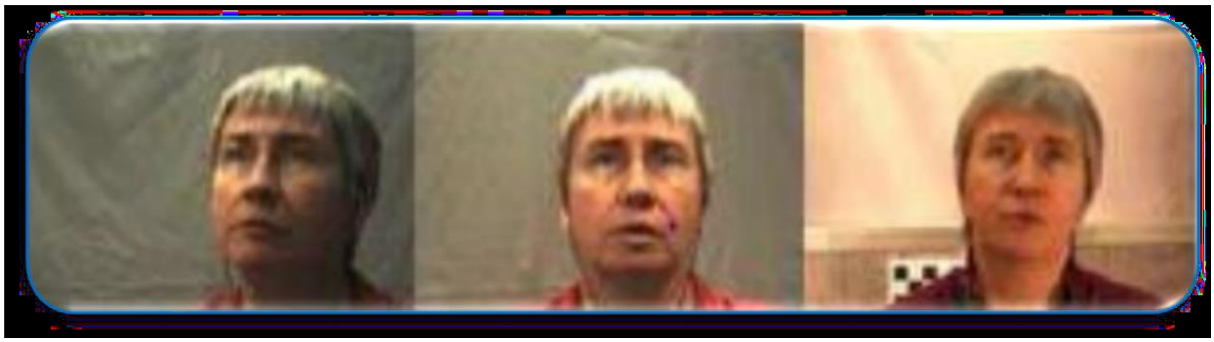


Figure 1.1: une Exemple de variation d'éclairage [6].

- **Variation de pose**

Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation < 30°), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30°, la normalisation géométrique n'est plus possible [6].

Figure 1.2 : Exemple de variation de pose [6].

- **Expression faciale**

La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l’expression faciale modifie l'aspect du visage, elle entraine forcement une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d’actualité et qui reste non résolu. [6]

Figure 1.3 : Exemples de variation d’expressions [6].

2.5. Systèmes de reconnaissance faciale

La reconnaissance faciale est un système qui repose sur la manipulation de l’image ou la vidéo pour extraire automatiquement les traits du visage, afin de les conserver en tant que marqueurs numériques dans une base de données où chaque personne dispose d’une signature numérique unique. Ces systèmes fonctionnent sur deux modes soit l’identification ou l’authentification des personnes sans aucune autre information ou aide [7].

2.6. Processus de la reconnaissance faciale

Les systèmes de reconnaissance faciale passent par plusieurs étapes de traitement. Les sections suivantes présentent l'essentiel de ces traitements qui sont montrées dans la figure 1.4.

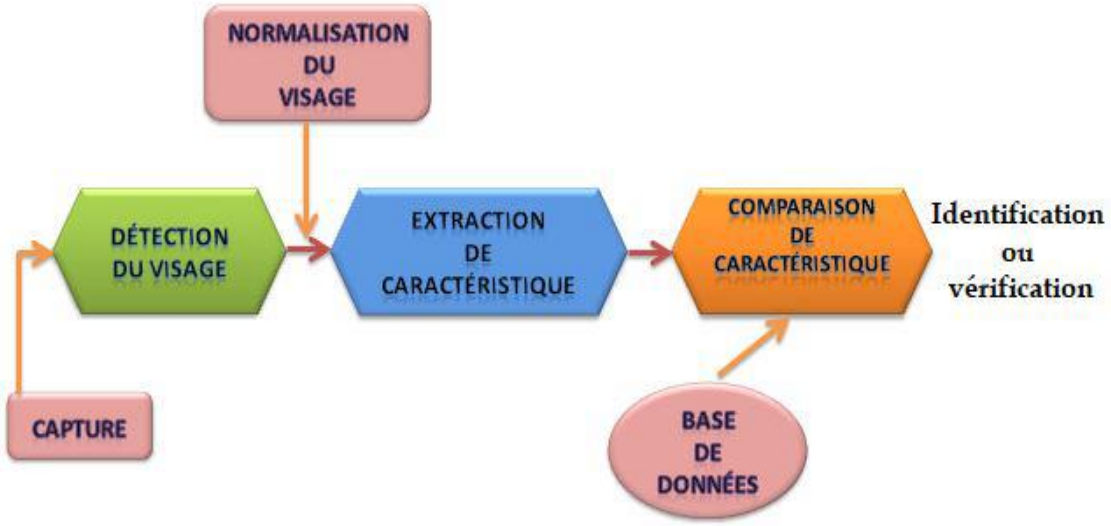


Figure 1.4 : Schéma général d'un système de reconnaissance des visages

2.6.1. Capture

C'est la phase préliminaire qui consiste à capturer l'image naturelle et la transformer vers une matrice dont la valeur de chaque élément représente une intensité discrète de la lumière en cas de photo noire et blanc ou couleur cette matrice est appelée image digitale. En effet, l'image de bonne qualité donne de meilleurs résultats dans la phase de la reconnaissance. Donc il faut capturer l'information pertinente sans bruit. On trouve plusieurs types de capteurs d'image tels que : les captures classique 2D, les captures 3D, Les captures en proche infrarouge. Chaque type de capteur présence des avantages et des inconvénients. Dans la reconnaissance des visages, on peut utiliser les capteurs 3D par exemple pour s'affranchir des problèmes de pose ; mais leur prix excessif ne permet pas une utilisation à grande échelle. Les captures en proche infrarouge sont utilisées pour éliminer les problèmes d'illumination [7]

2.6.2. Détection du visage

La détection de visage dans une image est la deuxième phase de traitement avant la reconnaissance. Le traitement repose sur la détection de la région du visage qui contient les yeux, la bouche et le nez. Il existe des méthodes de détection du visage on peut le catégorisées en 4 méthodes.

• Méthodes basées sur les connaissances

Ces méthodes sont basées sur la connaissance des principales caractéristiques du visage et les relations qui existent entre eux. Par exemple, un visage apparaît dans une image avec deux yeux dans des positions relatives entre un nez, et une bouche. [8]

• Méthodes basées sur les caractéristiques invariantes

Ces approches se basent sur des caractéristiques structurelles (traits faciaux, texture, couleur de la peau) qui existent même quand la pose, le point de vue, ou les conditions d'illumination varient. [9]

• Méthodes basées sur la mise en correspondance de modèles

Ces méthodes consistent à l'utilisation des modèles prédéfinis des visages ou une partie de visage (bouche, œil, nez). Une comparaison s'effectue entre chaque modèle de l'ensemble existent et l'image entrée pour identifier la présence de visage ou non dans l'image. [10]

• Méthodes basées sur l'apprentissage

Ces méthodes se basent sur le même principe des modèles prédéfinis des visages mais les modèles sont ici des modèles appris à partir d'un ensemble d'images d'apprentissage qui doivent permettre de caractériser la variabilité de l'apparence d'un visage. Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps d'entraînement. [8]

Ces méthodes de détection du visage sont liées aux qualités de l'image de visage extraite, donc on peut effectuer des améliorations sur l'image avant de passer aux étapes suivantes, voici quelque amélioration :

• Normalisation géométrique

La normalisation géométrique consiste à extraire la zone du visage de l'image originale, ensuite une rotation du visage est effectuée afin d'aligner l'axe des yeux avec l'axe horizontal. Enfin, une réduction proportionnelle à la distance entre les centres des deux yeux est appliquée. On obtient alors une image de visage dont la distance entre les centres des yeux est fixe. Les dimensions

De l'image du visage sont calculées à partir de la distance à obtenir entre les centres des deux yeux [11]. La figure 2.2 illustre la méthode de la normalisation géométrique.

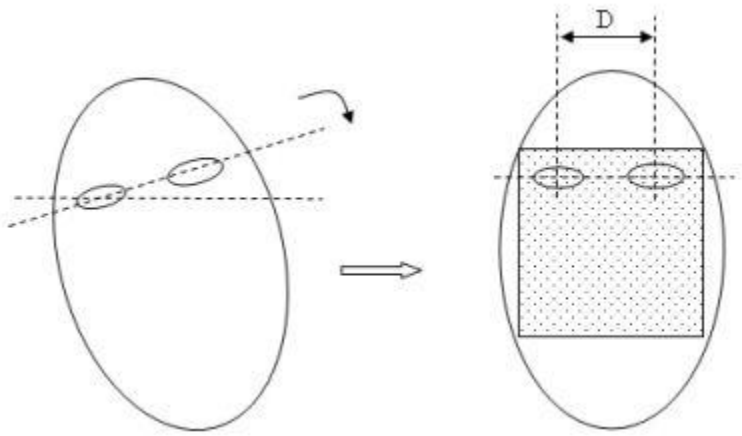


Figure 1.5 : Normalisation géométrique. [7]

• Normalisation photométrique

Ce prétraitement est nécessaire pour régler l'éclairage dans une image et minimiser l'influence de l'illumination. Cela on peut être effectué soit par des méthodes simples tell que l'égalisation d'histogrammes comme elle présente dans la figure 6 et correction de gamme, ou par des méthodes complexes telles que le lissage anisotropie [7].

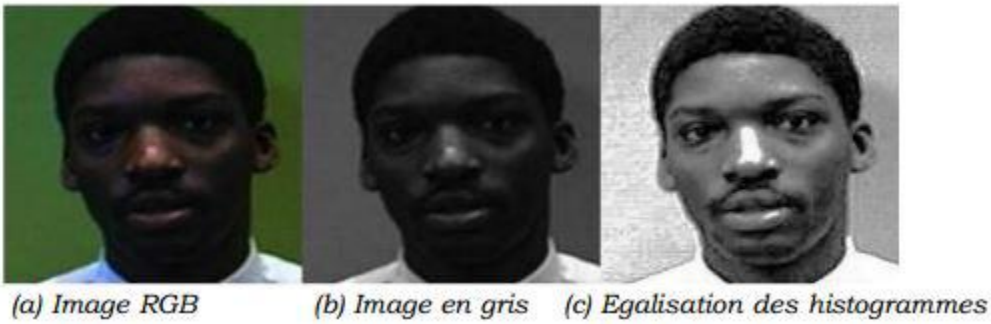


Figure 1.6 : Normalisation photométrique par égalisation d'histogramme. [11]

2.6.3. Extraction de caractéristiques

Après la détection de la zone de visage, le système va extraire Les caractéristiques du visage à l'aide de deux pratiques différentes. La première pratique repose sur l'extraction des régions entières du visage, elle est souvent implémentée avec une approche globale de reconnaissance de visage. La deuxième pratique extrait des points particuliers des différentes régions caractéristiques du visage, tels que les coins des yeux, de la bouche et du nez. Elle est présentée comme une approche locale de reconnaissance [12].

2.6.4. Comparaison des caractéristiques

Selon les caractéristiques extraites à l'étape précédent, les algorithmes de comparaison sont différés. On trouve dans la littérature plusieurs approches tel que le calcul de distance, calcul de similarité. D'autres méthodes se basent sur la classification des caractéristiques par un classifier (machine à vecteurs de support, classifier bayésien, etc.) [7].

3. Emotion

3.1. Définition

L'émotion est l'état mental d'une personne qui dépend de son humeur, elle peut changer à cause d'un événement qui se produit ou d'une situation particulière.

Une émotion peut rester intérieure à une personne et rester limité à son expérience intime, mais même dans ce cas l'envie de passer à l'action est toujours présent, que ce soit par un ressenti ou à travers l'imagination [13]

3.2. Types d'émotions

Les émotions sont des séquences courantes qui interviennent tout au long de nos journées, on les reconnaît grâce à des expressions faciales, à la tonalité de la voix ou même à des gestes corporels, il existe trois types d'émotions : émotion primaire, émotion secondaire et émotion sociale.

3.2.1. Émotion primaire

Les émotions primaires (joie, tristesse, +colère, peur, dégoût, surprise) sont engendrées par des évènements quotidiens et elles sont à la base des réactions humaines qui cause des comportements spécifiques [14].

Figure 1.7 : Les émotions faciales de gauche à droite 1 : dégoût ; 2 : peur ; 3 : joie ; 4 : Surprise ; 5 : tristesse ; 6 : colère. [15]

Les deux types d'émotions suivants sont appris et constitué à partir des émotions primaires.

3.2.2. Émotion secondaire

Les émotions secondaires sont les plus difficiles à identifier, car elles sont souvent le résultat d'une combinaison de deux émotions primaires, elles sont à l'origine d'un souvenir évoqué et arrivent à maturation à l'âge adulte [14].

3.2.3. Émotion social

Les émotions sociales (culpabilité, honte, jalousie, timidité, humiliation, etc.) sont acquises en fonction de l'éducation et la culture d'une personne, elles nous permettent de nous adapter aux autres afin de pouvoir vivre en société [14].

3.3. Analyse des émotions faciales

Les émotions jouent un rôle essentiel dans notre vie puisqu'elles permettent d'améliorer la conversation humaine. L'analyse des émotions faciales est un sujet de recherche actif pour les spécialistes du comportement depuis 1872 par les travaux de Charles Darwin [16]. Et ceux du psychologue Paul Ekman et ses collègues depuis les années 1970, où cette équipe définit en

1978 l'ensemble des émotions de base et établit qu'elles sont universelles (se présentent dans tous les cultures de la même façon) [17].

La reconnaissance des émotions faciales(REF) à partir d'images est un sujet de recherche intéressant où Suwa et Motoi ont présenté une première tentative d'analyse automatique des expressions faciales en traçant le mouvement de 20 points identifiés sur une séquence d'images en 1978 [18].

Bien que la reconnaissance des émotions faciales (REF) est un sujet de recherche depuis de nombreuses années dans le domaine de la vision par ordinateur les systèmes de reconnaissance des émotions trouvés son utilité dans divers domaines tels que la communication, les sciences du comportement, les jeux vidéo, l'animation, la psychiatrie, etc. Ce domaine présente encore de nombreux défis liés à la complexité des émotions, les changements dans la pose du visage, les conditions d'éclairage et les variations entre les individus en termes d'attributs tels que l'âge, le sexe [19].

Dans cette partie, nous présentons les travaux de recherches sur la REF automatiques que nous divisons en deux groupes selon que les caractéristiques sont créées manuellement ou générées via un réseau de neurones profonds.

La figure 1.8 présente la structure de base des systèmes de reconnaissance des émotions faciales qui partagent les mêmes premières étapes des systèmes de reconnaissance faciale. Le système repose sur la détection et la localisation de la zone du visage pour passer à l'étape de l'extraction de caractéristiques du visage selon deux approches : les approches basées sur les caractéristiques géométriques ou les approches basées sur les caractéristiques d'apparence ou une approche hybride. Enfin, l'étape de classification qui utilise les méthodes d'apprentissage traditionnelles (tel que SVM, RNA) ou d'apprentissage profond tel que les réseaux convolutionnells

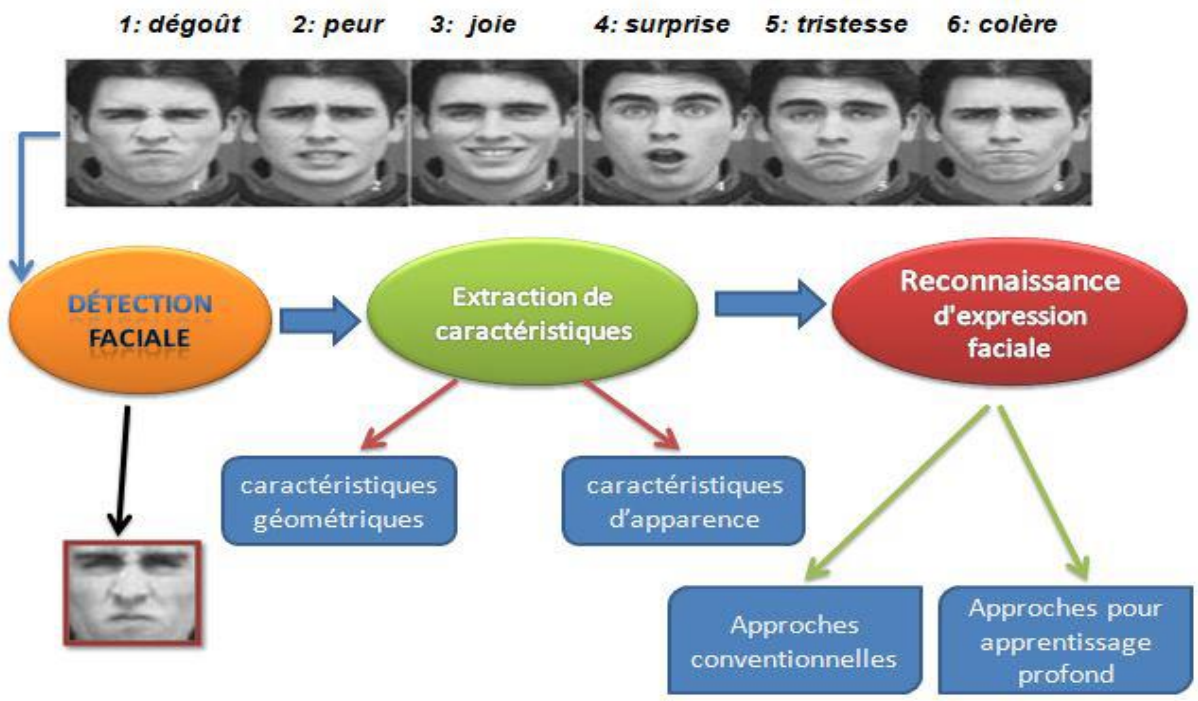


Figure 1.8 : Structure de base des systèmes de reconnaissance des émotions faciales.

3.4. Pourquoi s'intéresser aux émotions

L'émotion a un impact sur notre jugement [Clore 1994] et notre raisonnement [Damasio 1994]. Elle influe également sur l'attention, la motivation, la mémoire, la résolution de problèmes ou laprise de décision.

Lors de la réunion plénière Humaine de 2007, Paul Ekman a décrit un outil qu'il a mis en placeafin de déterminer l'émotion d'une personne (peur, colère, …) par l'observation des « unités d'action » du visage. Il rapportait que les gens amélioraient leur reconnaissance des émotions, que cette amélioration perdurait plusieurs mois après l'apprentissage et que certaines personnes avaient constaté une amélioration de leurs relations avec les autres, suite à cet apprentissage.

R. Picard a fait des observations similaires, particulièrement lors d'études avec des individus ayant du mal à reconnaître les émotions comme les autistes. Des expériences sur des simulateurs de voiture [Nass et al. 2005] ont mis en évidence le fait qu'une personne de bonne humeur réagissait mieux à une voix de bonne humeur (moins d'accidents). Par contre, une personne stressée allait mieux réagir à une voix plus sobre et une voie joyeuse allait au contraire l'irriter encore plus et augmenter le nombre d'accidents. Dans toutes les études où on compare deux

versions d'un outil l'une avec un module de traitement sur les états affectifs, aussi basique soit-il, et un autre sans ce module, l'utilisateur va systématiquement préférer la version affective et cela va souvent se répercuter sur ses performances. De même, dans le domaine de l'éducation, un tuteur virtuel qui adapterait sa stratégie à l'état émotif d'un élève pourrait lui permettre de progresser plus rapidement et avec plus de plaisir. Au MIT également, des études en cours proposent l'intégration des bio senseurs à des produits du type iPod ou téléphone portable afin par exemple d'adapter la musique de l'iPod à l'humeur du sujet ou de prévenir les rechutes d'anciens toxicomanes en détectant les signaux physiologiques de manque. [20]

3.5. Différents modèles du processus émotionnel

Il existe différents modèles du processus émotionnel, on peut citer les modèles suivants :

3.5.1. Modèle de James

Un stimulus externe, est perçu avec les aires sensorielles du cortex cérébral. Les réponses sont contrôlées par le cortex moteur. Elles produisent des sensations qui retournent au cortex cérébral où elles sont perçues (figure 1.9). Cette perception, associée à la réponse émotionnelle est ce qui donne à l'émotion ses qualités propres dans la théorie de James. Pour James les réponses émotionnelles précèdent donc et déterminent les expériences conscientes [21].

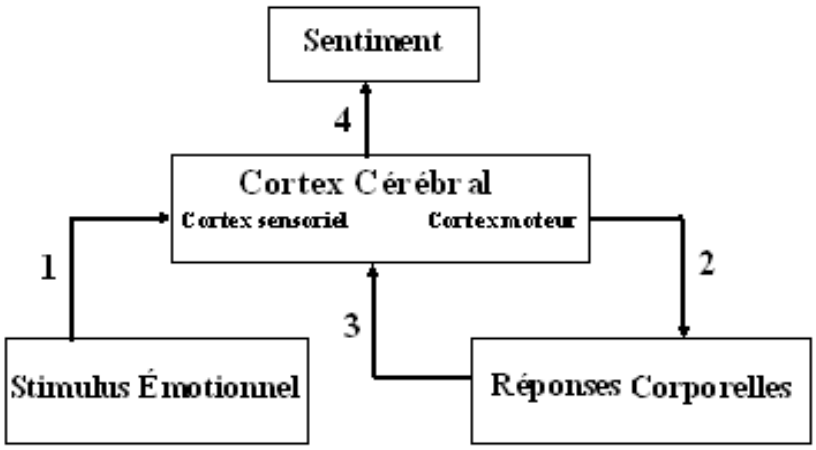


Figure 1.9 : Le modèle de James.

3.5.2. Modèle de Cannon

Cannon pensait que les stimuli externes traités par le thalamus étaient dirigés vers le cortex cérébral et vers l'hypothalamus (figure 1.10). Ce dernier envoie des messages aux muscles et aux organes dans le corps ainsi qu'au cortex [21].

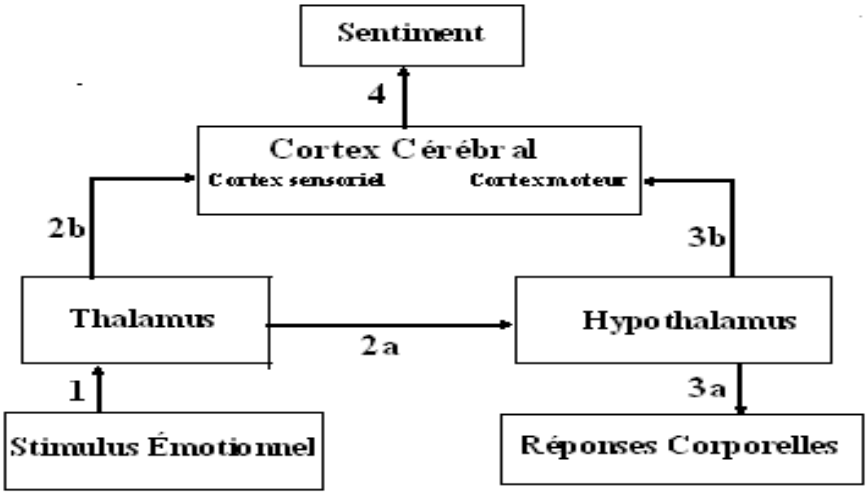


Figure1.10 : Le modèle de Cannon.

L'interaction des messages dans le cortex entre ce qu'est le stimulus et sa signification, émotionnelle résulte en une expérience consciente de l'émotion. Dans cette théorie, les réponses émotionnelles et les sentiments se produisent en parallèle [21].

3.5.3. Modèle de Ledoux

L'information sur les stimuli externes arrive à l'amygdale par une voie directe provenant du thalamus (la voie basse) ou par une voie passant par le cortex (la voie haute). Voir figure 1.11 La voie basse est plus courte et donc plus rapide que celle arrivant du cortex. Mais comme elle court-circuite le cortex, elle ne peut bénéficier du traitement cortical ne fournit donc à l'amygdale qu'une représentation grossière du stimulus [21].

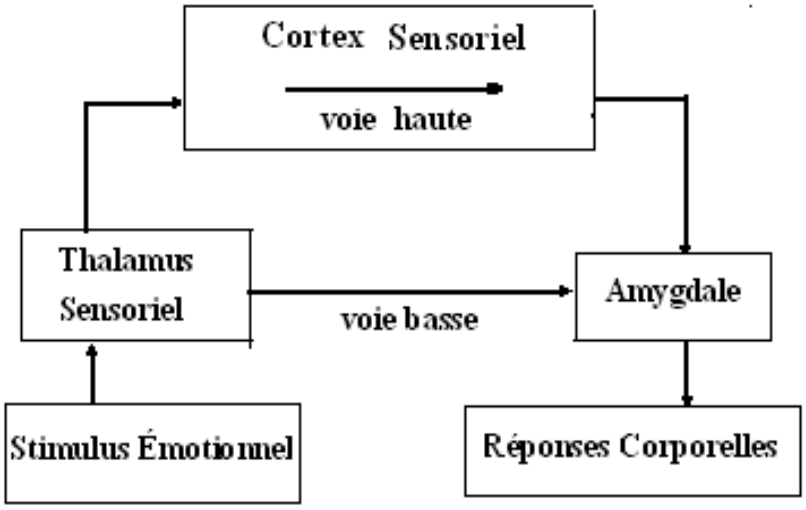


Figure 1.11 : Le modèle de Ledoux.

La voie directe nous permet de commencer à répondre à des stimuli relativement dangereux avant de savoir pleinement de quoi il s'agit. Cela peut être très utile dans les situations dangereuses. La voie haute permet de regarder plus sereinement la situation [21].

Elle fait le lien entre ce qui se passe dans le cortex, la conscience, la possibilité de symboliser les affects et l'organique végétatifs, hormonal et les réactions comportementales.

3.5.4. Modèle de Papez

Papez a élaboré un circuit inspiré de différents résultats expérimentaux. Il existe d'abord une circulation d'information à travers un circuit interconnectant l'hypothalamus et le cortex cingulaire. Il distingue une sous-région de l'hypothalamus appelée corps mamillaires.

Ces corps reçoivent les informations sensorielles en provenance du thalamus, et les relaient au cortex cingulaire par l'entremise du noyau antérieur du thalamus. L'hippocampe reçoit les informations du cortex cingulaire et envoie l'information à l'hypothalamus.

Les émotions peuvent aussi être générées de deux façons, par stimuli sensoriels entrant par le thalamus, ou par des pensées en provenance du cortex cingulaire [22].

3.6. Représentation des émotions

La manipulation des émotions sur une machine soulève de nombreuses problématiques. D'abord au niveau de la représentation des émotions, il s'agit de trouver un formalisme qui soit en accord avec les résultats psychologiques existants ; tout en permettant une manipulation simple sur machine. Ensuite, pour un événement donné, il faut pouvoir déterminer le potentiel émotionnel qui lui est associé [23].

En se fondant sur les travaux en psychologie, certaines mesures considèrent les états affectifs comme des catégories, d'autres comme un construit multidimensionnel [23].

Les approches les plus connues sont l'approche catégorielle ou discrète et l'approche multidimensionnel.

Les sept émotions basiques, largement acceptées par la communauté des psychologues, fournissent un premier ensemble, discret, d’émotions sur lequel se baser.

3.6.1. Approche catégorielle

C'est l'approche la plus répandue, qui consiste à considérer les émotions comme des caractéristiques épisodiques et universelles. Il suffit ensuite d'associer un mot du langage à ces caractéristiques. Le caractère universel des émotions entraîne la définition d'un petit nombre d’émotions basiques [24]. Cette approche fait essentiellement la distinction entrées émotions et propose de les classer sous forme des catégories discrètes. Ainsi les dénominations affectives qui ne trouvent pas leur place dans ces classifications sont considérées comme des mélanges d'émotions primaires [25].

La justification principale de cette approche réside dans le fait que ces émotions basiques sont clairement identifiables chez la majorité des individus, notamment à travers la communication non verbale. Toutefois, leur nombre, le nom qu'il faut leur attribuer et leur caractérisation comme émotion basique, restent des questions ouvertes. L'intérêt principal de

L‘approche catégorielle est qu'une fois que les émotions à traiter sont clairement identifiées, il devient simple de les manipuler, aussi bien pour les hommes que pour les machines [24].

3.6.2. Approche multidimensionnelle

Une deuxième façon de catégoriser l'émotion s'appuie sur un espace continu d’émotions. La perspective dimensionnelle, quant à elle, propose de modéliser toutes les réactions affectives à partir de plusieurs dimensions [25]. Cette approche consiste à considérer les émotions comme un point dans un espace multidimensionnel, et même interpréter la similarité entre divers types d'émotion comme des proximités dans l'espace. En général, deux axes suffisent à représenter un grand nombre d'émotions. Les deux axes de cet *espace multidimensionnel représentent des attributs qui sont, à priori communs à toutes les* manifestations émotionnelles comme la valence ou le plaisir de l'émotion (positif, négatif) et l’activation ou l'excitation de l'émotion (actif, passif) [24].

D'après les théories relevant de cette approche, ces deux dimensions émergent clairement :

3.6.2.1. Le plaisir ou la valence (positif, négatif)

Il forme le fondement de l'expérience affective et donne à cette expérience son caractère spécifique [23].

Cette dimension traduit le degré de bienêtre et de satisfaction.

3.6.2.2. L'activation ou l'excitation (actif/passif)

C'est une composante physiologique caractérisant l'activité physique d'un organisme, elle comprend deux pôles extrêmes (le sommeil et la surexcitation). Elle a été longtemps considérée comme la manifestation essentielle de l'affection pour la seule raison que l'état affectif n'était rien qu'une haute activation [23].

Le vécu émotionnel pouvait être retranscrit au moyen de trois dimensions [23]. Les dimensions plaisir (positif/négatif) et excitation (actif/passif) sont retrouvées, alors que la troisième dimension est la puissance (tension, relaxation) :

3.6.2.3. Puissance (tension, relaxation)

Elle fait référence à la sensation de pouvoir, de contrôle ou d'influence versus un manque de pouvoir ou une incapacité à contrôler ou influencer une situation [23].

Elle semble être moins stable.

L'approche dimensionnelle permet de présenter facilement des émotions nuancées, mais également des transitions entre différents états émotionnels [25]. Voir figure 1 :12

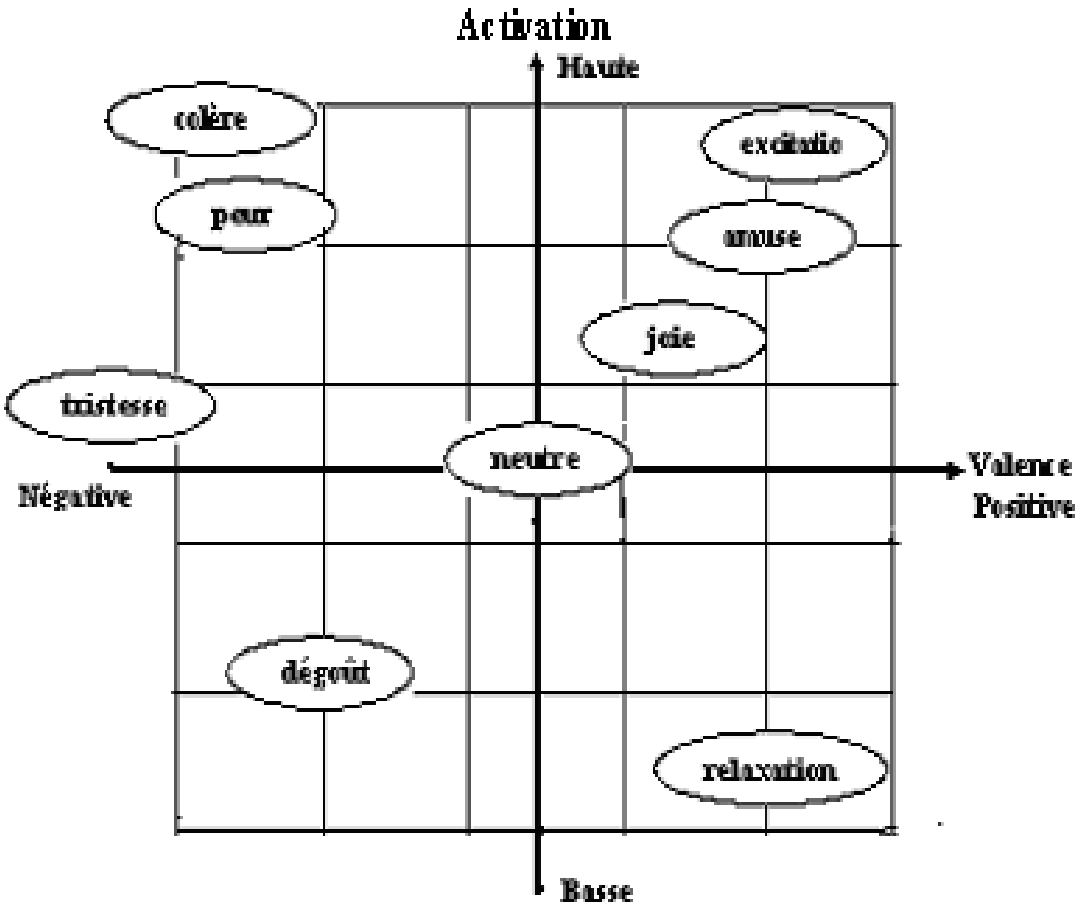


Figure1.12 : La représentation de quelques émotions sur deux axes

En outre, une combinaison de plusieurs émotions primaires permettrait d'expliquer la complexité de ce que l'on éprouve. Les émotions sont comparées à une palette de couleurs, les émotions primaires correspondant aux couleurs primaires, et les émotions plus complexes à un mélange de ces couleurs primaires [25]. Par exemple, le mépris résulte de la colère et du dégoût (voir figure 1.13). De plus, ces émotions varient en intensité (voir figure 1.14).

Figure1.13 : La représentation des émotions mixtes

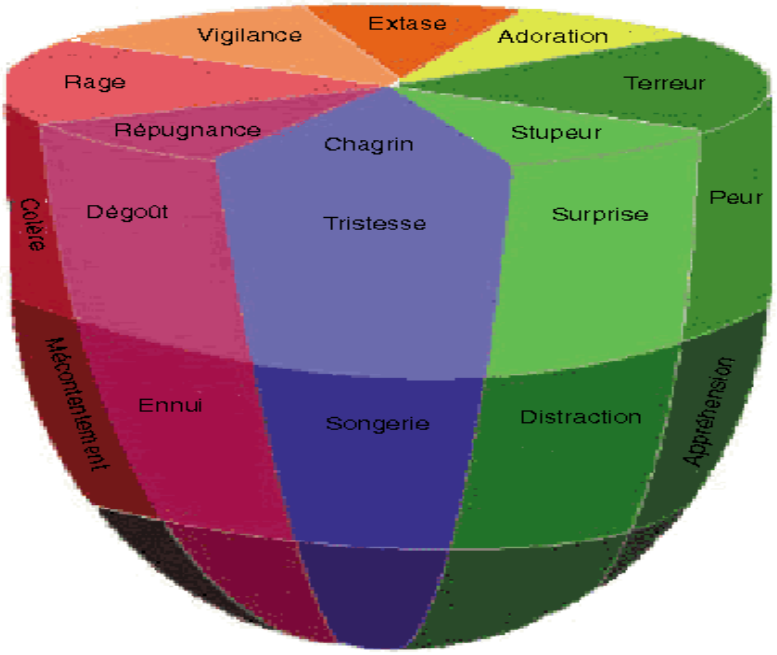


Figure 1.14 : La représentation de diverses émotions selon leurs intensités

Les deux approches, catégorielle et dimensionnelle, loin d'être opposées, sont complémentaires pour l'étude des émotions [25].

3.7. Les domaines d'application de la reconnaissance automatique des émotions

D'une façon générale, la majorité des applications, dans le domaine de la reconnaissance automatique des émotions, visent à remplacer le traditionnel questionnaire qui permet d'avoir le feedback de l'utilisateur. Une analyse des émotions en temps réel permet d'obtenir des résultats plus fiables, et ainsi une amélioration rapide et efficace de l'expérience de l'utilisateur. Cette section traite de différents domaines dans lesquels cette technique est utilisée.

a. Education et e-Learning

Il est particulièrement intéressant d'étudier l'humeur d'une personne durant son apprentissage. On parle ici d'apprentissage affectif (Affective Learning en Anglais), qui est un terme utilisé en psychologie et en pédagogie bien avant son utilisation en informatique, et qui décrit l'influence de l'état émotionnel sur la cognition et l'apprentissage humain [26].

Plusieurs faits ont été établis suite à de nombreuses études sur l'impact des émotions sur l'apprentissage, comme par exemple, qu'un état d'excitation très élevé stimulait la mémoire [26].

La combinaison de l'informatique affective à l'apprentissage affectif représente le fait d'analyser en temps réel, ou non, l'état émotionnel de l'apprenant. On peut ainsi détecter de nombreux signes tels que le stress, l'ennui, la déconcentration ou encore la frustration qu'un enseignant pourrait ne pas remarquer. Avec de telles informations, le processus d'apprentissage peut être optimisé et adapté afin de fournir la meilleure version à l'apprenant.

Cela s'avère d'autant plus utile avec l'expansion de l'apprentissage en ligne (e-Learning) et en direct où l'enseignant et les étudiants interagissent à travers des plateformes connectées et où il est difficile pour l'enseignant de capter et de garder l'attention des étudiants.

La reconnaissance automatique des émotions peut également contribuer à l'apprentissage des enfants de manière plus efficace qu'un adulte, le processus étant plus lourd pour un enfant, car il est plus facilement déconcentré et plus rapidement lassé.

b. Médecine et Psychologie

Des millions de personnes dans le monde souffrent d'autisme. Ce trouble, de nature imprévisible, cause aux personnes qui en souffrent des difficultés d'apprentissage et de communication avec le monde extérieur. Suite à de nombreuses études, il a été prouvé que l'utilisation de robots humanoïdes facilitait de façon significative ces deux processus. Un robot est plus simple à aborder pour le malade qu'un être humain, car il est moins complexe et plus prévisible en ce qu'il s'agit de comportement [27].

Les robots ayant la capacité de reconnaitre automatiquement les émotions peuvent donc être utilisées pour apprendre aux personnes souffrant de ce handicap à comprendre et à exprimer explicitement leurs émotions, à travers des exercices d'apprentissage par exemple. Il est également possible aujourd'hui, grâce à la reconnaissance des émotions par le pouls, de détecter et de prévenir les pics de stress dont souffrent les personnes autistes, et qui sont l'une des principales caractéristiques qui rendent ce trouble imprévisible et difficilement gérable.

Ces crises, qui sont incontrôlables, poussent parfois les personnes qui en souffrent à se faire du mal de façon involontaire, en les détectant ne serait-ce que quelques secondes à l'avance on peut limiter leurs effets et ainsi éviter des drames parfois irréversibles.

c. Personnalisation des sites web

Les sites web récoltent des informations sur leurs utilisateurs. Les derniers sites qu'il a visités, ce qu'il a acheté, ce qu'il a aimé ou bien ce qu'il a regardé, constitue une mine d'or pour les propriétaires du site, car ces informations permettent de créer un profil de l'utilisateur afin de lui fournir un contenu personnalisé qui vise à le faire rester le plus longtemps possible sur le site. Elles servent également à adapter les publicités affichées sur le site web selon chaque utilisateur afin qu'elles aient l'impact souhaitée. Ajouter les émotions à ces informations améliorerait, encore plus, la description de la personnalité de l'utilisateur, et permettrait d'avoir des profils plus complets et plus efficaces [28]

Deux expériences ont été effectuées par [28] dans lesquelles il a utilisé la reconnaissance automatique des émotions afin de déterminer deux points essentiels qui sont : quel impact l'état affectif de l'utilisateur avait sur sa consommation de contenu ? Et quelle publicité convient le mieux à chaque humeur ?

Pour cela, ils ont confronté des utilisateurs, dans différents états affectifs à plusieurs mises en forme du site web, afin qu'ils choisissent celle qui leur convenait le mieux. Ensuite ils

leurs ont présenté différentes publicités durant lesquelles ils étudiaient leurs émotions tout en effectuant un suivi de leurs yeux afin de déterminer non seulement la réaction mais aussi le temps pendant lequel l'utilisateur restait concentré avec la publicité.

4.Conclusion

Dans ce chapitre, nous allons voir quelques concepts de base sur l'expression faciale et ces principales difficultés. Ensuite, nous allons présenter quelques notions relatives aux émotions telles que leur définition et leurs types. Ainsi que, nous expliquerons par la suite les modèles du processus émotionnel. Enfin, nous présenteront quelques domaines d'application de la reconnaissance automatique des émotions.

Chapitre 02

L’apprentissage automatique et le deep learning

1. Introduction

L’apprentissage profond (Deep Learning : DL) est un concept d'Intelligence Artificielle (IA), dérivé de l’apprentissage machine (Machine Learning : ML) [29] (voir Figure 2.1). L’IA correspond à un ensemble de technologies et outils qui permettent de simuler l’intelligence et accomplir des tâches que les humains exécutent de manière intuitive et quasi automatique tel que la perception, la compréhension et la prise de décision [30]. Le DL est l’une des raisons qui ont conduit au progrès de l’IA, et qui fait penser que finalement, il existe une possibilité pour l’IA de devenir plus réaliste. Le DL s’intéresse à la reconnaissance des formes et à l’apprentissage à partir des données en se basant sur les modèles de réseaux de neurones artificielles.

Dans ce qui suit, nous allons rappeler les concepts de ML. Nous présentons ensuite le modèle du réseau de neurones qui est à la base de DL. Nous présentons ensuite les différentes architectures profondes de réseaux de neuronal à convolution profond (Deep-CNN).

2. Intelligence Artificielle

L'intelligence artificielle est l'étude de la manière dont les ordinateurs peuvent effectuer des tâches intelligentes qu’est dans le passé, ne pouvaient être réalisées que par des humains. [31]

Autre définition dit que L'intelligence artificielle est un ensemble de plusieurs technologies et théories informatiques et aussi un Domaines d'intérêt pour la pensée. La logique et l'intelligence visent à concevoir des programmes capables de résoudre des problèmes et de traiter le langage ainsi que d’exécuter autres tâches qu’était exclusive à l’être humain [32].

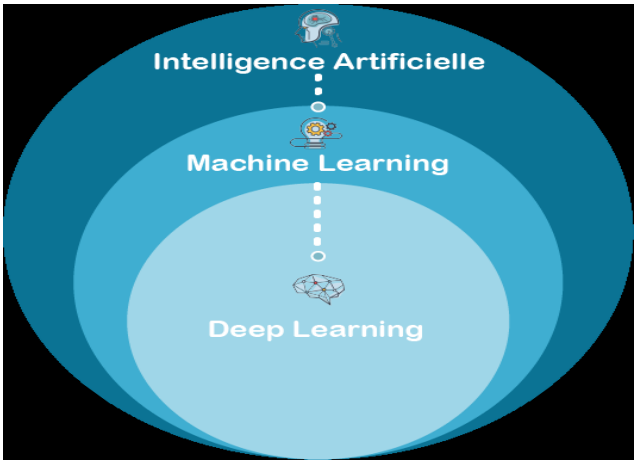


Figure 2.1 : La relation entre IA et ML et DL [33].

3. Apprentissage automatique « machine-learning »

3.1. Définition

La théorie de ML est un domaine qui croise les aspects mathématiques et informatiques découlant de l'apprentissage itératif [34]. Une définition formelle de ML a été proposée par T. Mitchell [35], lequel affirme qu'un programme informatique est dit capable d'exécuter une tâche lorsqu'il peut améliorer ses performances avec l'expérience. C'est grâce à l'apprentissage qu'il pourra apprendre à exécuter de nouvelles tâches et acquérir de nouvelles compétences.

L'apprentissage nécessite une expérience sous forme d'une base de données que le modèle analyse. Cet apprentissage d'une tâche se fait à l'aide d'une fonction de coût calculée sur une base de données d'apprentissage, distincte de celle utilisée pour le test, afin de mesurer les performances.

Lors de l'apprentissage, les paramètres peuvent être ajustés pour optimiser le modèle afin de réduire la fonction coût sur la base de données test. Le fait que ces performances soient mesurées sur des bases de données différentes cela implique une capacité de généralisation d'un modèle dit d'apprentissage [34].

3.2. Les types d'apprentissage automatique

On distingue usuellement au moins trois types d'apprentissage machine : l'apprentissage par renforcement, l'apprentissage supervisé et l'apprentissage non supervisé. :

3.2.1. Apprentissage supervisé

Supposant on donne des exemples étiquetés, comme des images de lettres manuscrites avec le nom de la lettre correspondante (étiquettes *a*, *b*, *Z*…). L'apprentissage consiste alors à construire une fonction capable de déterminer la lettre de l'alphabet à laquelle se rapporte chaque image. Cette forme d'apprentissage a fait des progrès considérables ces dernières années [36].

3.2.2. Apprentissage non supervisé

Lorsque les gens font référence à des systèmes capables d'apprendre par eux-mêmes, ils font référence à l'apprentissage non supervisé. Dans l'apprentissage non supervisé, l'algorithme d'apprentissage ne reçoit pas d'étiquettes pour les données, laissant l'algorithme trouver la structure à partir de l'entrée.

Comme les données ne sont pas étiquetées, il n'y a pas d'évaluation de la précision de la structure produite par l'algorithme.

Cela inclut le regroupement, la réduction de la dimensionnalité et l'apprentissage des règles d'association. L'algorithme peut ne jamais trouver la bonne sortie, mais modéliser la structure sous-jacente des données [37].

3.2.3. L'apprentissage par renforcement

Un algorithme d'apprentissage automatique par renforcement apprend de l'environnement s'il obtient de bons résultats, il reçoit une récompense, et l'objectif est de maximiser la récompense [38].

L'algorithme reçoit un retour d'information concernant les récompenses et les punitions au fur et à mesure qu'il avance dans le problème. L'apprentissage par renforcement permet de décider de la meilleure action suivante en fonction de son état actuel et en apprenant les comportements qui maximiseront la récompense [37].

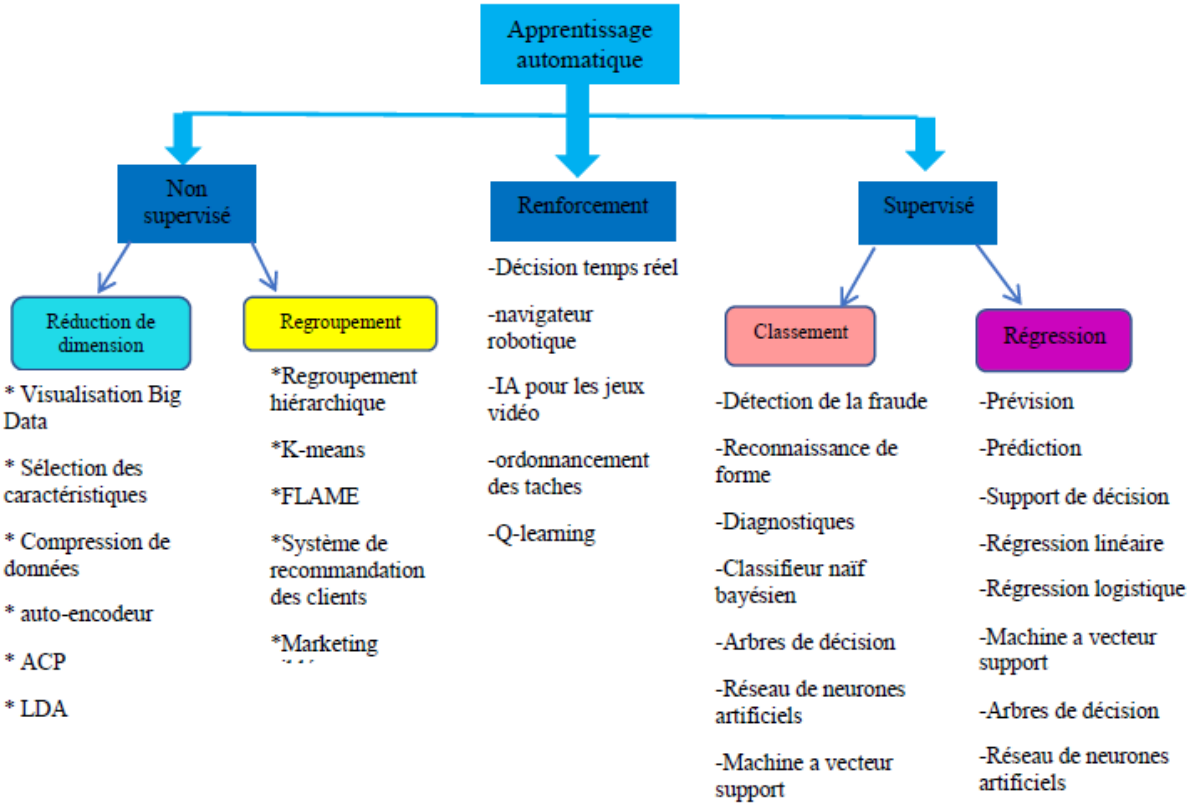


Figure 2.2 : les différents types de l'apprentissage automatique

4. Apprentissage profond « Deep Learning »

L'apprentissage en profondeur est une branche d'apprentissage automatique, qui utilise plusieurs couches de traitement non linéaires pour apprendre des représentations de fonctions utiles directement à partir de données. Les modèles de Deep Learning peuvent atteindre un très haut niveau de précision dans le cadre de la classification d'objets. L'entraînement des modèles s'effectue sur un vaste jeu de données labellisées et sur des architectures de réseaux de neurones contenant de nombreuses couches. Les modèles de Deep Learning sont bâtis sur le même modèle que les perceptrons multicouches précédemment décrits. Cependant, les couches intermédiaires sont plus nombreuses [39]. Ainsi que le deep learning permet de faire l'extraction des caractéristiques aussi.

Bien que l'apprentissage approfondi ait été théorisé pour la première fois dans les années

1980, il existe deux raisons principales pour lesquelles il est devenu très utile :

L'apprentissage approfondi nécessite de grandes quantités de données marquées.

Par exemple, le développement de voitures sans conducteur nécessite des millions d'images et des milliers d'heures de vidéo.

L'apprentissage approfondi nécessite un pouvoir informatique important.

Les GPU haute performance possède une architecture parallèle efficace pour l'apprentissage en profondeur. Cela permet réduire le temps de formation pour un réseau d'apprentissage en profondeur de semaines en heures ou moins [40].

Le Deep Learning peut s'appliquer à de nombreux problèmes : Classification d'images,

Reconnaissance vocale, et le traitement du langage naturel.

4.1 Les réseaux de neurones profonds

Les réseaux de neurones peu profonds (avec une seule couche cachée) requirent un ensemble de caractéristiques extraites manuellement à partir de données brutes en utilisant par exemple des descripteurs tel que SIFT, HOG et mises sous formes de vecteur a une dimension 1D. L’extraction de ces caractéristiques demande de bonnes connaissances sur ces données brutes et sur la tâche d’apprentissage, ainsi qu’un travail d’ingénierie pour adapter les méthodes d’extraction. Cette opération est relativement coûteuse à la mise en place, ainsi qu’une mauvaise extraction des caractéristiques mène à de très mauvaises performances en termes d’apprentissage.

L’idée des architectures profondes consiste alors à intégrer cette extraction de caractéristiques, normalement faite "à la main", par un processus d’apprentissage dans les premières couches du réseau de neurones, en traitant ainsi des données a deux dimensions 2D, à savoir des images, plutôt que des données à une dimension 1D. Cela met en jeu une capacité à traiter de grandes quantités d’information. Le terme profond réfère donc au nombre de couches (plus de 2 couches cachées) des réseaux de neurones profonds entre l’entrée et la sortie.

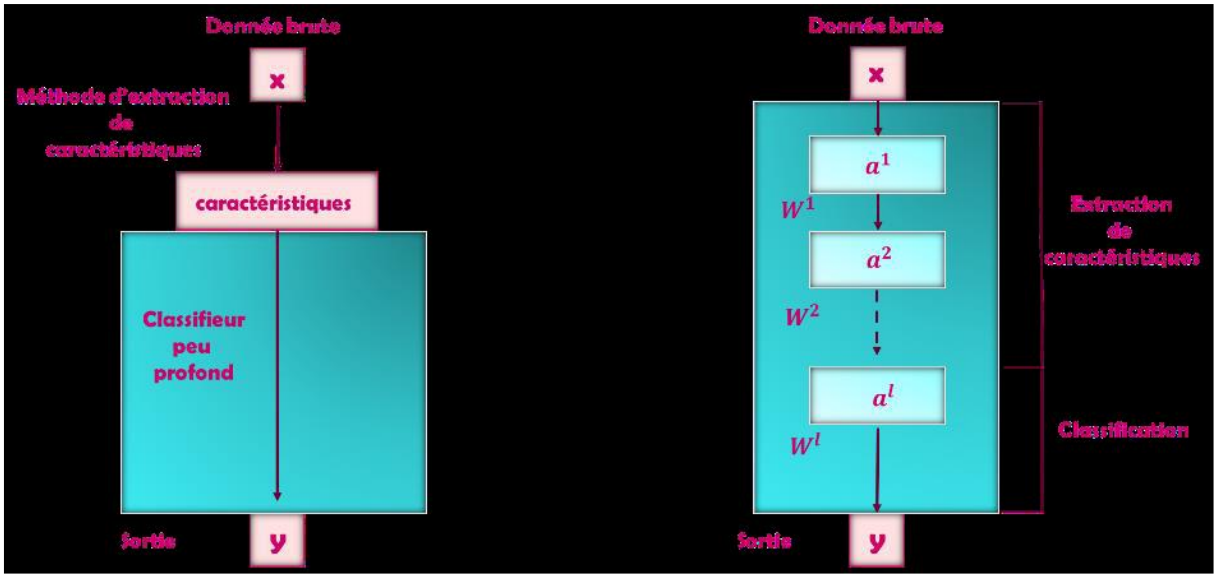


Figure 2.3 : La différence entre l'apprentissage automatique classique (à gauche) et l'apprentissage profond (à droite). La zone en bleu est la zone d'apprentissage [41].

4.2. Les réseaux de neurones convolutionnels

4.2.1. Définition

Un réseau neuronal conventionnel (CNN) est un type de réseau neuronal artificiel utilisé dans la reconnaissance et le traitement d'images et spécifiquement conçu pour traiter les données de pixels.

Conventionnelle neural network sont de puissants systèmes de traitement d'images, d'intelligence artificielle (IA)qui utilisent un apprentissage approfondi (deep learning) pour effectuer des tâches à la fois génératives et descriptives, souvent à l'aide de **Machine Vision** qui inclut la reconnaissance d'images et de vidéos, ainsi que des systèmes de recommandation et le traitement du langage naturel(NLP) [42].

4.2.2 Architecture d'un réseau de neurone convolutionnels

Une architecture CNN typique comprend généralement des couches alternées de convolution et de mise en commun, suivies d'une ou plusieurs couches entièrement connectées à la fin. Dans certains cas, une couche entièrement connectée est remplacée par une couche de mise en commun de la moyenne globale. En plus des différentes fonctions de mappage, différentes unités de régulation telles que la normalisation et le dropout des lots sont également incorporées pour optimiser les performances du CNN. La disposition des composants du CNN joue un rôle fondamental dans la conception de nouvelles architectures

et l'obtention de meilleures performances. Cette section aborde brièvement le rôle de ces composants dans une architecture CNN [43].

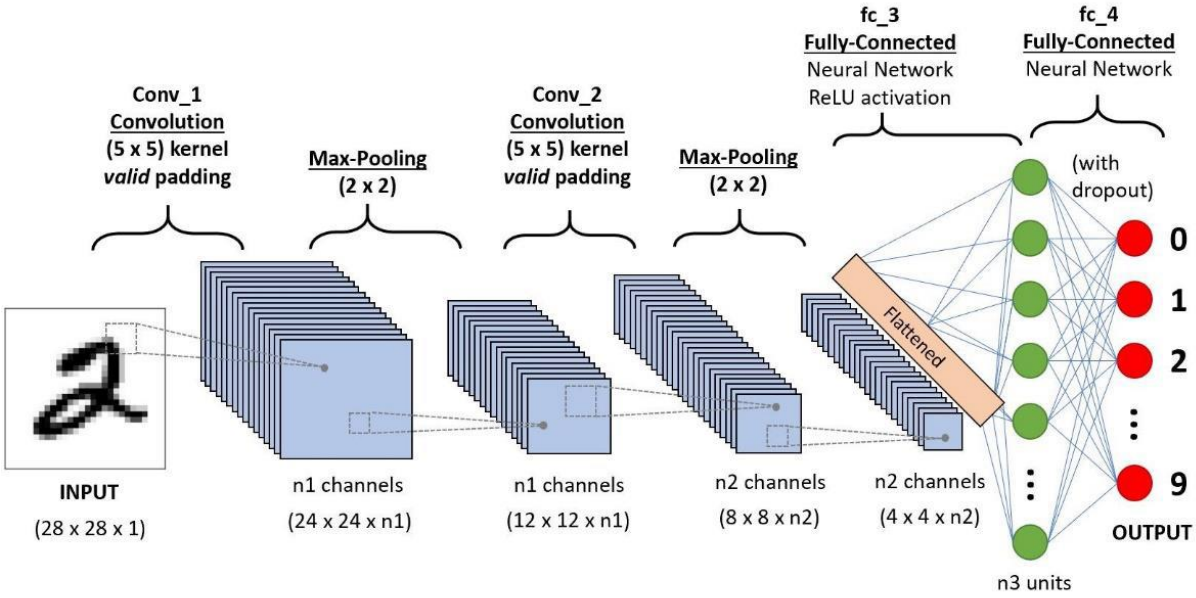


Figure 2.4 : architecture d'un réseau de neurone convolutionnel [44].

- **La couche Convolution**

La couche de convolution est parfois appelée couche d'extraction de caractéristiques, car les caractéristiques de l'image sont extraites dans cette couche.

Tout d'abord, une partie de l'image est connectée à la couche Convolution pour effectuer une opération de convolution et calculer le produit scalaire entre le champ récepteur (c'est une région locale de l'image d'entrée ayant la même taille que celle du filtre) et le filtre Le résultat de l'opération est un entier unique du volume de sortie. Ensuite, nous faisons glisser le filtre sur le champ récepteur suivant de la même image d'entrée par une foulée et refaisons la même opération. Cette opération est répétée par le même processus encore et encore jusqu'à ce que toute l'image soit parcourue [45].

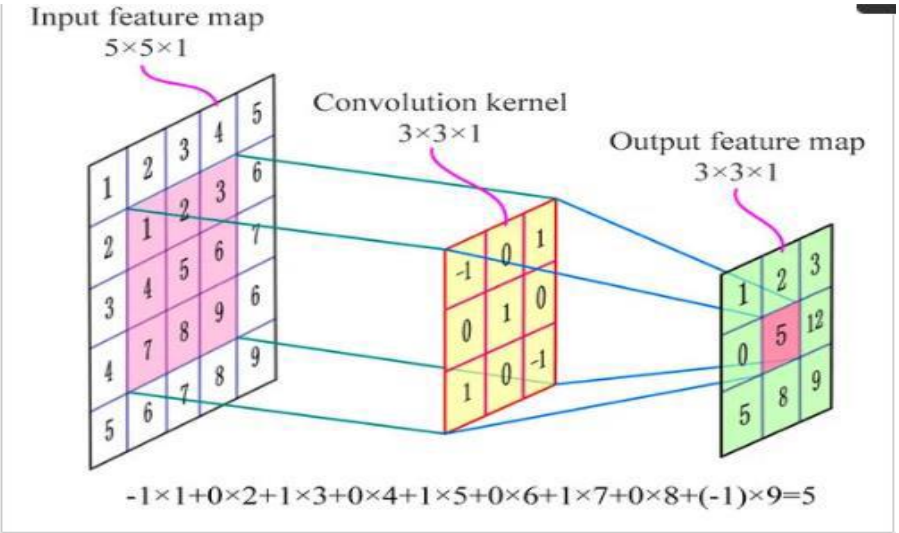


Figure 2.5 : La couche de convolution.

- **Couche de mise en commun (Pooling)**

La couche de mise en commun (POOL) est une opération de sous-échantillonnage, généralement cette opération est appliquée entre deux couches de convolution.

Sa fonction est de réduire progressivement la taille de la carte de fonctionnalités (matrice de convolution) pour réduire les paramètres et les calculs réseau, tout en conservant les informations importantes [46].

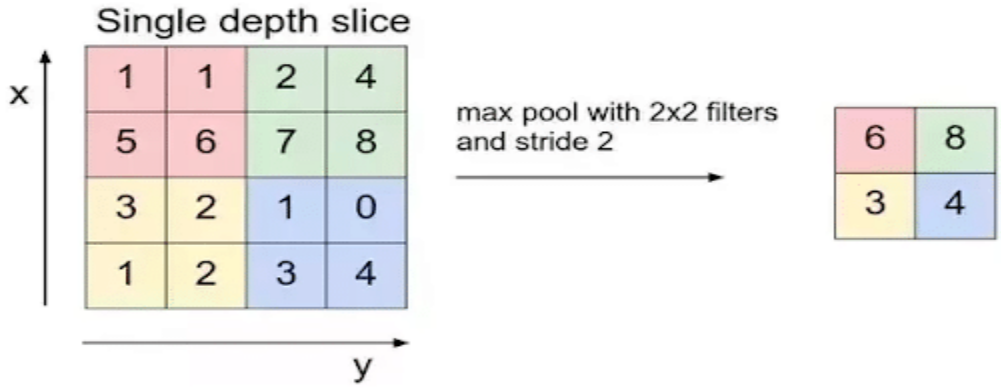


Figure 2.6 : la couche pooling

➢ **Couche entièrement connecté (Fully Connected)**

Les couches totalement connectées font les mêmes tâches que celles des ANN standard et tenteront de produire des notes de classe à partir des activations, pour les utiliser pour la classification. Il est également suggéré d'utiliser Relu entre ces couches pour améliorer les performances. [47]

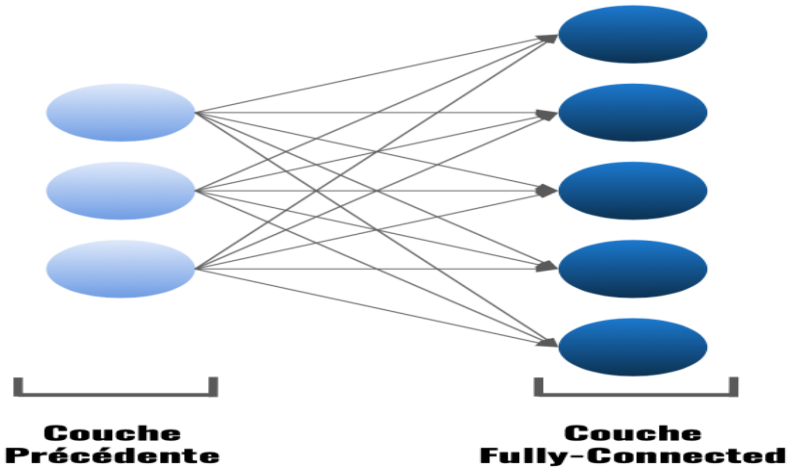


Figure 2.7 : Couche entièrement connecté (Fully Connected) [48].

➢**Couche de sortie (output layer)**

La couche de sortie, c'est la dernière couche de réseaux qui contient les neurones qui identifient les classes de modèle, donc le nombre de neurones à cette couche dépende du nombre de classes.

Il y a plusieurs demains qui utiliser le CNN pour résoudre les problèmes et parmi ces demains en a la détection d'objet et classification d'image, dans le schéma suivant nous allons voir les meilleurs model dans les deux demains précèdent :

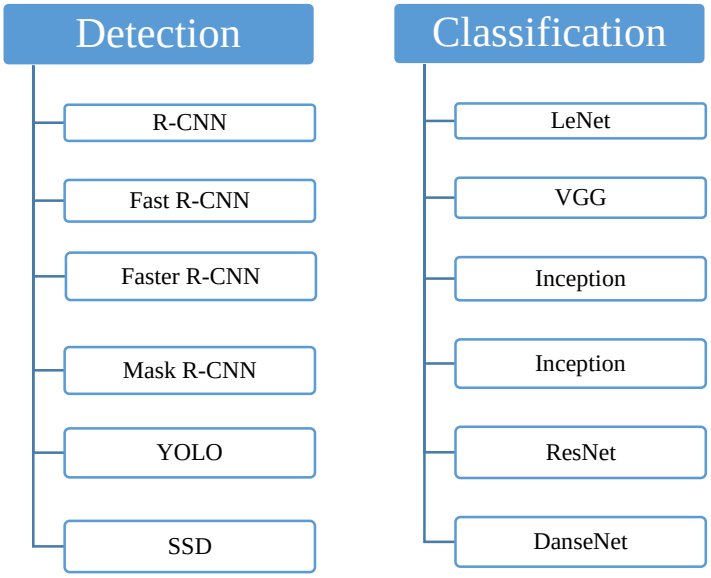


Figure 2.8 : schéma des modèles de détection et classification par CNN.

4.2.3 : Réseau neuronal à convolution profonde (Deep-CNN)

4.2.3.1 Qu'est-ce que le Deep-CNN

Dans les réseaux neuronaux, le réseau neuronal convolutionnel (ConvNets ou CNNs) est l'une des principales catégories pour faire de la reconnaissance d'images, de la classification d'images. Détection d'objets, reconnaissance de visages, etc.

Les classifications d'images CNN prennent une image en entrée, la traitent et la classent dans certaines catégories (par exemple, chien, chat, tigre et lion). Les ordinateurs voient une image d'entrée comme un réseau de pixels et cela dépend de la résolution de l'image. En fonction de la résolution de l'image, il verra h x w x d (h = hauteur, w = largeur, d = dimension).

Par exemple, une image de 6 x 6 x 3 matrice de RVB (3 se réfère aux valeurs RVB) et une image de 4 x 4 x 1 matrice d'image en niveaux de gris.

Techniquement, les modèles CNN profonds font passer chaque image d'entrée pour la former et la tester à travers une série de couches de convolution avec des filtres (Kernels), des couches de pooling, des couches entièrement connectées (FC) et appliquent la fonction Soft max pour classer un objet avec des valeurs probabilistes entre 0 et 1 [49].

4.2.3.2 Architecture et principales opérations utilisées dans un Deep-CNN

La figure ci-dessous (figure2.9) est un flux complet de Deep-CNN pour traiter une image d'entrée.

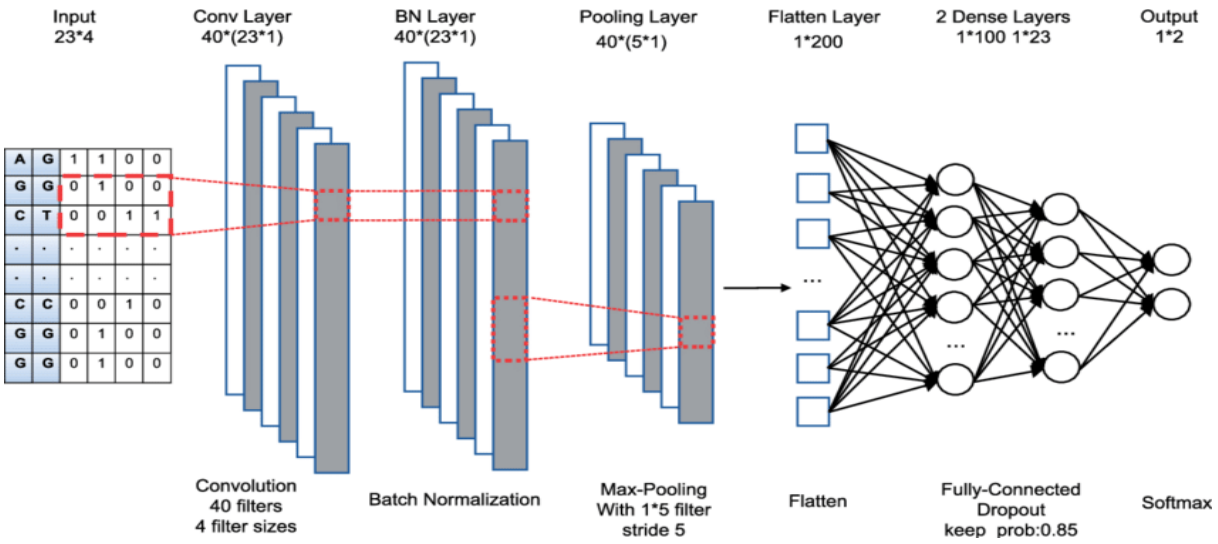


Figure 2.9 : Exemple d'architecture Deep-CNN

Tout d'abord, tous les termes utilisés dans cette architecture seront définis afin d'expliquer son fonctionnement.

- **Input :** Représente l'information brute qui est introduite dans le réseau. Il s'agit de l'image d'entrée.
- **Convolution Layer :** C'est la première couche qui extrait les caractéristiques d'une image d'entrée. La convolution préserve la relation entre les pixels en apprenant les caractéristiques de l'image à l'aide de petits carrés de données d'entrée. Il s'agit d'une opération mathématique qui prend deux entrées telles que la matrice d'image et un filtre ou noyau (figure 2.10) [49].

- An image matrix (volume) of dimension **(h x w x d)**
- A filter **(f_h x f_w x d)**
- Outputs a volume dimension **(h - f_h + 1) x (w - f_w + 1) x 1**

Figure 2.10 : La matrice d'image multiplie la matrice de noyau ou de filtre.

Considérons une matrice 5 x 5 dont les valeurs des pixels de l'image sont 0, 1 et une matrice de filtre 3 x 3 comme indiqué ci-dessous (figure 2.11) [49].

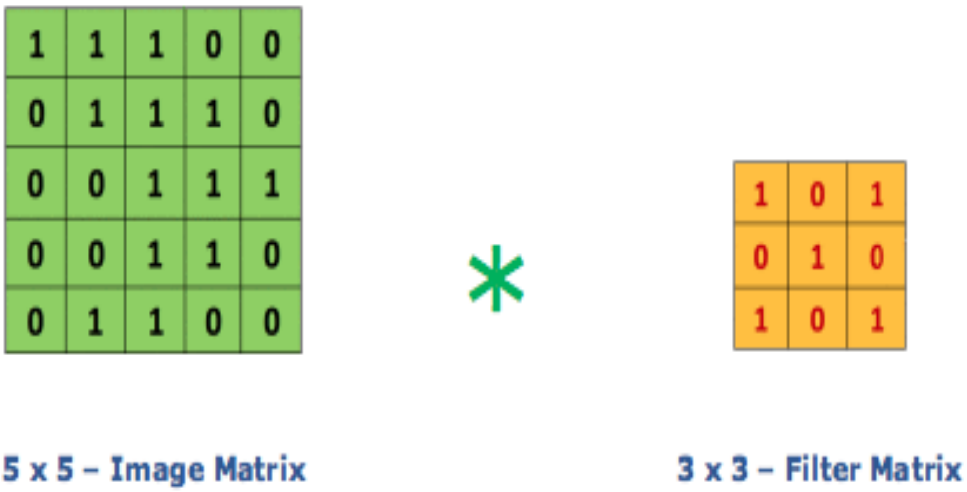


Figure 2.11 : La matrice d'image multiplie le noyau ou la matrice de filtre avec la convolution de l'image 5 x 5.

La matrice est multipliée par une matrice de filtrage 3 x 3, appelée "Feature Map", dont la sortie est illustrée ci-dessous (Figure 2.12) [49].

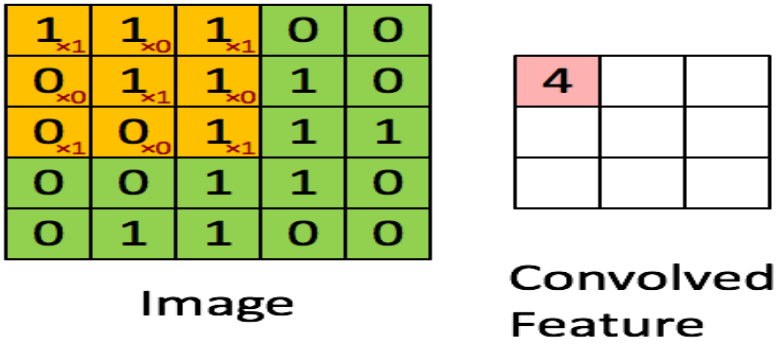


Figure 2.12 : Matrice de sortie 3 x 3

- **ReLu Layer** : L'unité linéaire rectifiée ou ReLu, n'est pas un composant distinct du processus des réseaux neuronaux convolutifs. Il s'agit d'une étape supplémentaire à l'opération de convolution (Figure2.13).

Le but de l'application de la fonction de redressement est d'augmenter la non-linéarité de nos images, car les images sont naturellement non linéaires.

Toute image contient de nombreuses caractéristiques non linéaires (par exemple, la transition entre les pixels, les bordures, les couleurs, etc.) Le redresseur sert à briser encore plus la linéarité afin de compenser la linéarité que nous pourrions imposer à une image lorsque nous la soumettons à l'opération de convolution [50].

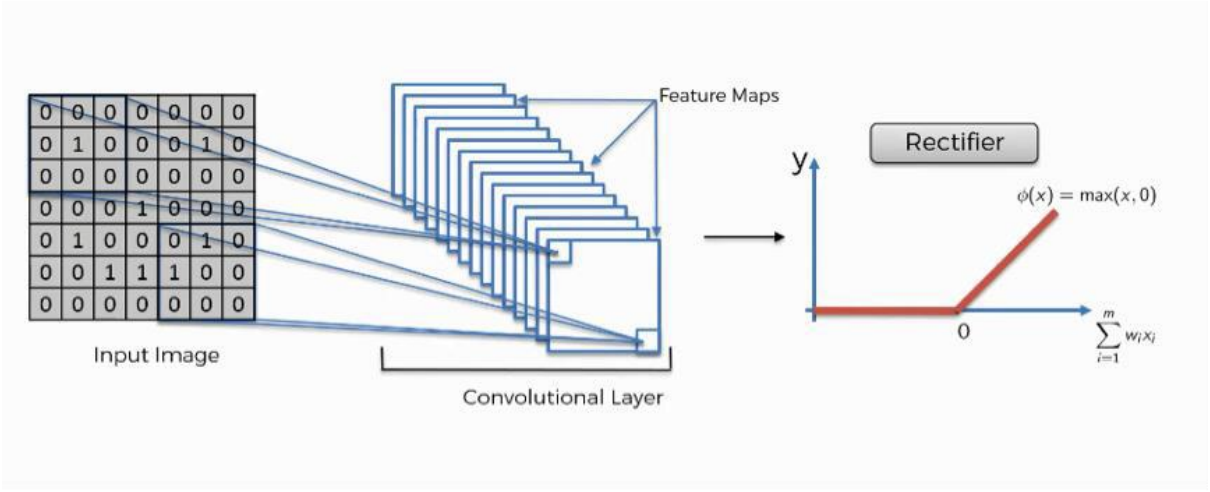


Figure 2.13 : La fonction de redresseur.

- **Batch Normalisation :** La normalisation par lots fonctionne de manière très similaire dans les réseaux neuronaux convolutionnells. Dans les convolutions, nous avons des filtres partagés qui suivent les cartes de caractéristiques de l'entrée (dans les images, la carte de caractéristiques est généralement la hauteur et la largeur).

Ces filtres sont les mêmes sur chaque carte de caractéristiques. Il est donc raisonnable de normaliser la sortie, de la même manière, en le partageant sur les cartes de caractéristiques.

En d'autres termes, cela signifie que les paramètres utilisés pour normaliser sont calculés avec chaque carte de caractéristiques. Dans une norme de lot ordinaire, chaque caractéristique aurait une moyenne et un écart type différents. Ici, chaque carte de caractéristiques aura une moyenne et un écart-type uniques, utilisés pour toutes les caractéristiques qu'elle contient [51].

- **Stride :** La stride est le nombre de pixels décalés sur la matrice d'entrée. Lorsque le stride est de 1, les filtres sont déplacés d'un pixel à la fois. Lorsque le stride est de 2, les filtres sont déplacés de 2 pixels à la fois et ainsi de suite. La figure ci-dessous (figure 2.14) montre que la convolution fonctionne avec un stride de 2.

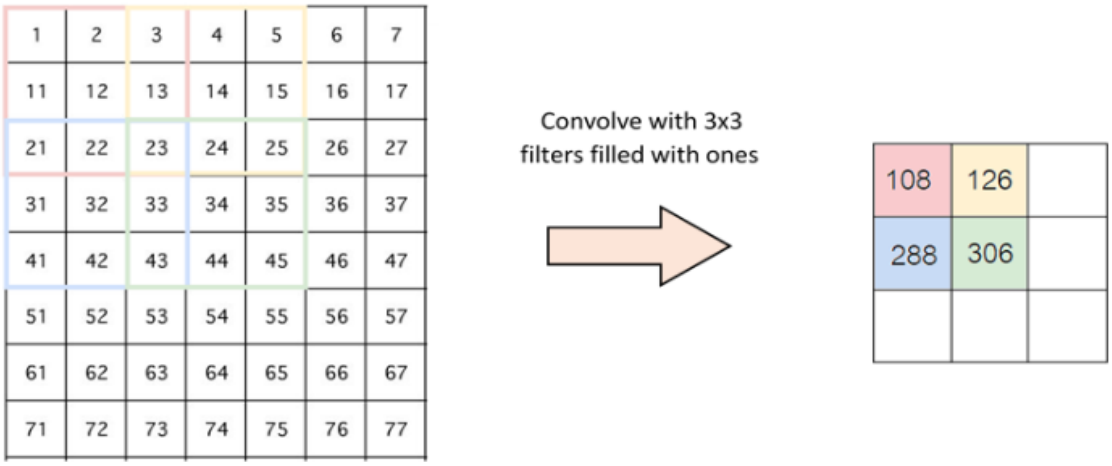


Figure 2.14 : Stride de 2 pixels

- **Max Pooling :** La mise en commun maximale est une opération de mise en commun qui sélectionne l'élément maximal de la région de la carte de caractéristiques couverte par le filtre. Ainsi, la sortie après la couche de max-pooling serait une carte de caractéristiques contenant les caractéristiques les plus importantes de la carte de caractéristiques précédente [52].

- **Flatten :** Flatten est la fonction qui convertit la carte de caractéristiques regroupées en une seule colonne qui est transmise à la couche entièrement connectée. Dense ajoute la couche entièrement connectée au réseau neuronal [53].

- **Fully Connected Layer :** La couche que nous appelons la couche FC, nous avons aplati notre matrice en vecteur et nous l'alimentons dans une couche entièrement connectée comme un réseau neuronal.

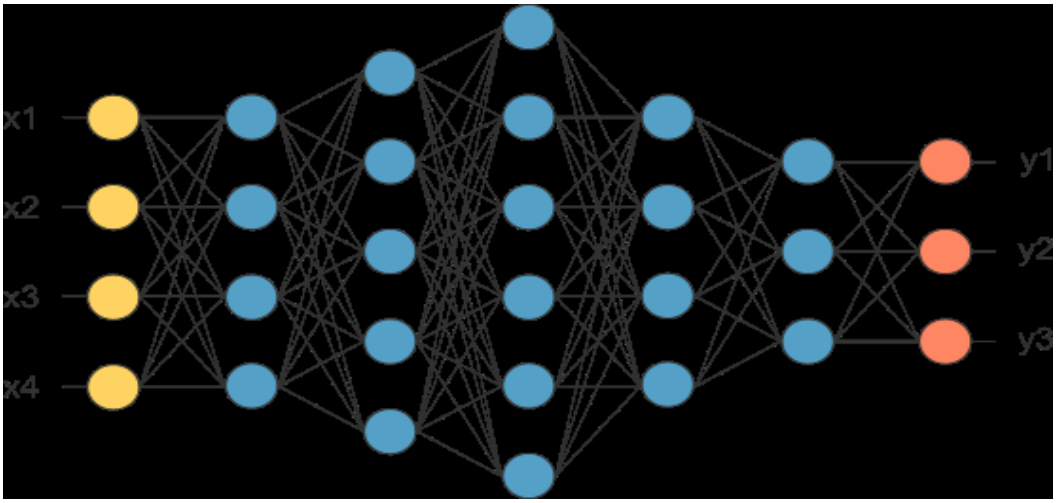


Figure 2.15 : Couche FC après la couche de mise en commun.

La matrice de la carte des caractéristiques sera convertie en vecteur (x1, x2, x3 ...). Avec les couches entièrement connectées, nous avons combiné ces caractéristiques ensemble pour créer un modèle. Enfin, nous avons une fonction d'activation telle que soft max ou sigmoïde pour classer les sorties comme chat, chien, voiture, camion etc. [54].

- **Dropout** : Le Dropout est implémenté par couche dans un réseau neuronal. Il peut être utilisé avec la plupart des types de couches, telles que les couches denses entièrement connectées, les couches convolutionnells et les couches récurrentes telles que la couche du réseau de mémoire à long terme.

Le Dropout peut être implémenté sur une ou toutes les couches cachées du réseau ainsi que sur la couche visible ou d'entrée. Il n'est pas utilisé sur la couche de sortie (figure 2.16).

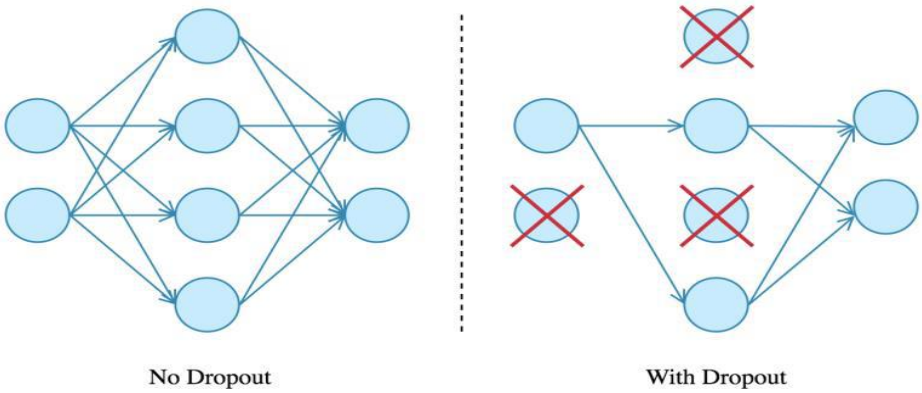


Figure 2.16: Example de Dropout.

5. YOLO (You Only Look Once)

Proposé par J. Redmon et al [54], est l'un des algorithmes de détection et de classification d'objets en temps réel les plus puissants. Le YOLO possède une architecture très simple, ce qui le rend extrêmement rapide. On l'appelle ainsi car contrairement aux algorithmes de détection et de classification d'objets cités précédemment, qui examinent successivement plusieurs régions de l'image pour trouver les objets qui sont présents, puis faire plusieurs prédictions sur chacune des régions, le YOLO a changé cela en raisonnant au niveau de l'image globale. Plutôt que d'utiliser la méthode en deux étapes pour la classification et la localisation de l'objet, YOLO applique un CNN unique pour la classification et la localisation de l'objet simultanément. Il analyse l'image une seule fois en la découpant en S× S cellules. Si le centre d'un objet tombe dans une cellule, cette cellule est "responsable" de la détection de l'existence de cet objet. Chaque cellule prédit :

1) L'emplacement des fenêtres englobantes :

-Les coordonnées de la fenêtre sont définies par un tuple de 4 valeurs, (centre (x ; y), largeur w, hauteur h). De plus, x, y, w et h sont normalisés par la largeur et la hauteur de l'image, et donc leurs valeurs varient entre (0, 1)

2) Un score de confiance :

-Il indique la probabilité que la cellule contienne un objet : *Pr* (contenant un objet) x IoU (pred, vérité) ; où *Pr*= probabilité et IoU = interaction sur union.

3) Une probabilité de classe d'objet conditionnée par l'existence d'un objet dans la boîte de délimitation :

- Si la cellule contient un objet, elle prédit une probabilité que cet objet appartienne à chaque classe Ci,=1,..., : Pr (l'objet appartient à la classe Ci | contenant un objet). À ce stade, le modèle ne prédit qu'un seul ensemble de probabilités de classe par cellule, quel que soit le nombre de fenêtres englobantes.

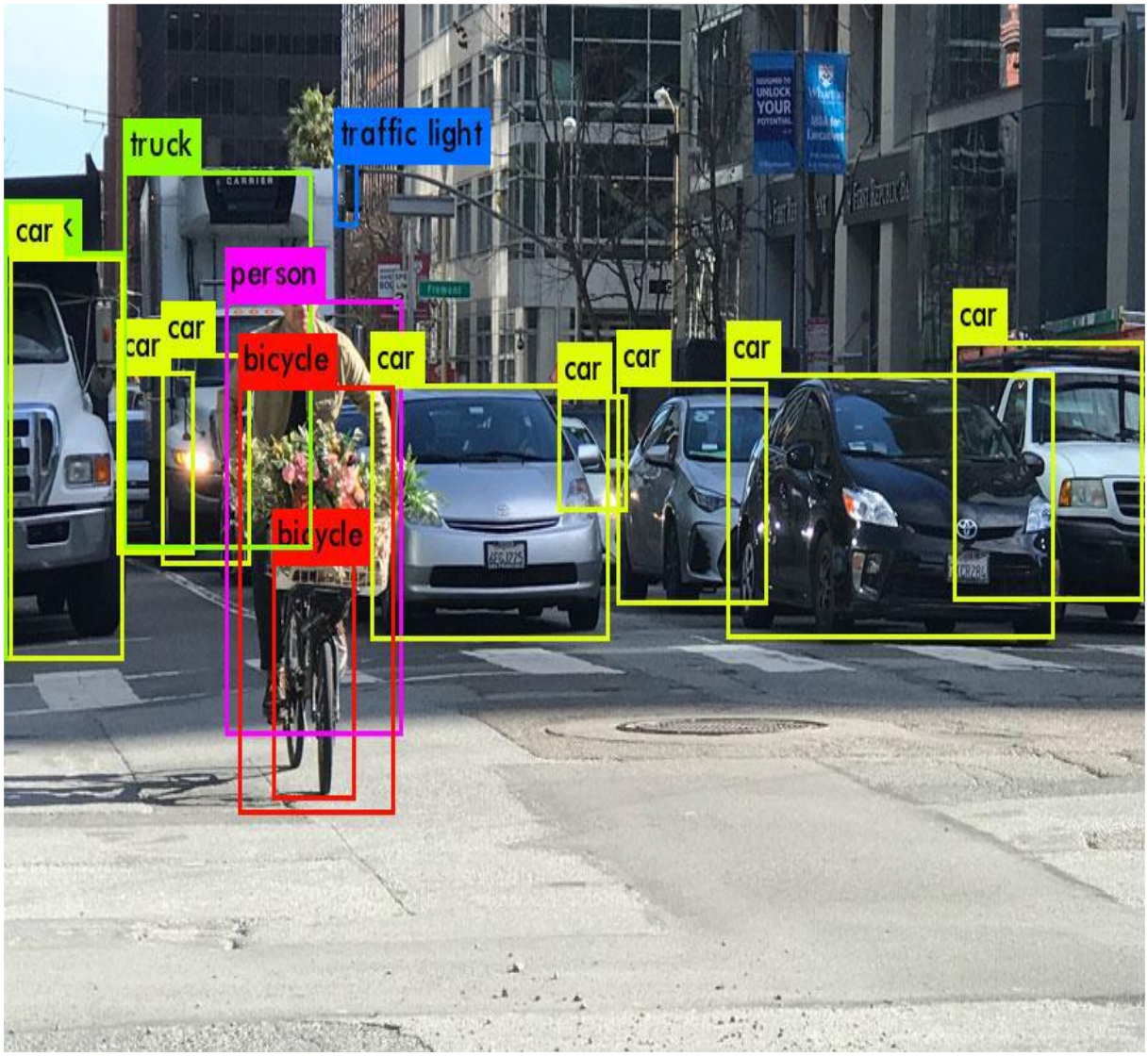


Figure 2.17 : Exemple de prédiction avec YOLO.

5.1. Le model yolov4

Comme tous les modèles de détection d'objet yolov4 composé de trois parties :

Backbone: CSPDarknet53, Neck: SPP, PANet, Head: Même que YOLOv3.

Dans la figure suivant nous allons illustrer l'architecture de yolov4 on détaille :

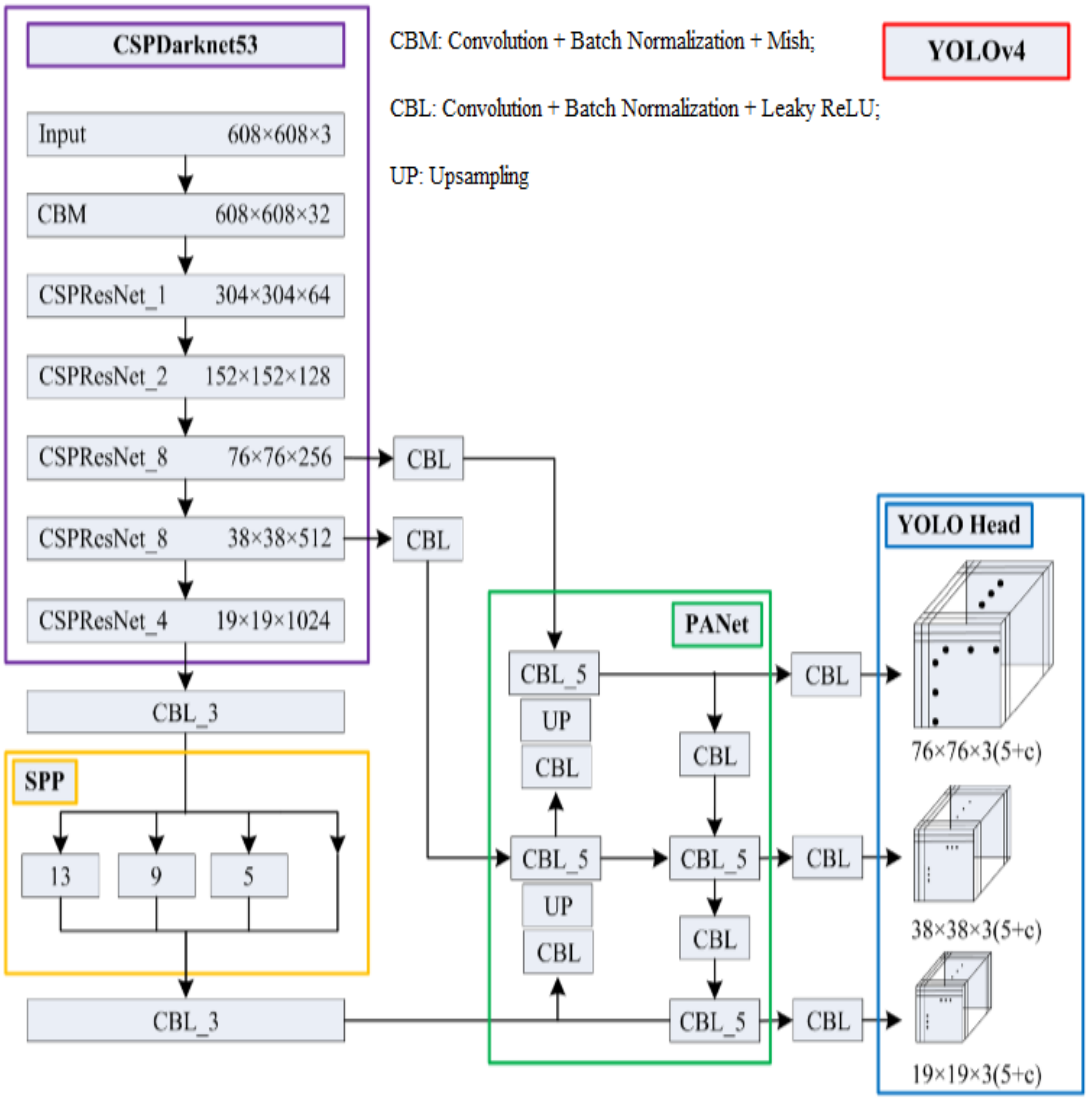


Figure 2.18 : Architecture de model yolov4.

6. Conclusion

Aujourd’hui, l’apprentissage profond a prouvé sa grande efficacité et a permis des avancées considérables dans de nombreux domaines.

Dans ce chapitre, nous avons présenté un aperçu des méthodes de détection d’images basées sur l'apprentissage en profondeur. Nous avons commencé par structure générale d’un modèle de détection base sur Deep Learning et passé en revue les méthodes de détection d'images les plus connus et les plus utilisées avec ses architectures. Après nous avons vu la définition de modèle YOLO avec son architecture.

Chapitre 03

Conception et Implémentation

1. Introduction

D’après les modelés de détection nous avons vu le modèle YOLO avec ses différents algorithmes, alors nous avons choisir ce modelé pour notre conception, pour résoudre notre problème qui se résume en la création d'un système de détection des émotions faciale. Nous avons choisi le modèle yolov4, que nous avons reconfiguré selon nos besoins, puis ré-entraîner sur classes sélectionnés pour notre contexte à partir de la base de donnée.

Utilisation du langage de programmation Python dans l'environnement de programmation libre de Google Colab. Ce chapitre se concentre sur la conception de notre approche ainsi que les détails de notre modèle choisi : yolov4 avec l’expérimentation et les résultats.

2. Conception

2.1. Schéma de conception

Notre système se compose de 3 modules principaux :

- Prétraitement de données
- Apprentissage de modèle yolov4 sur nos données choisies
- Tests et résultats

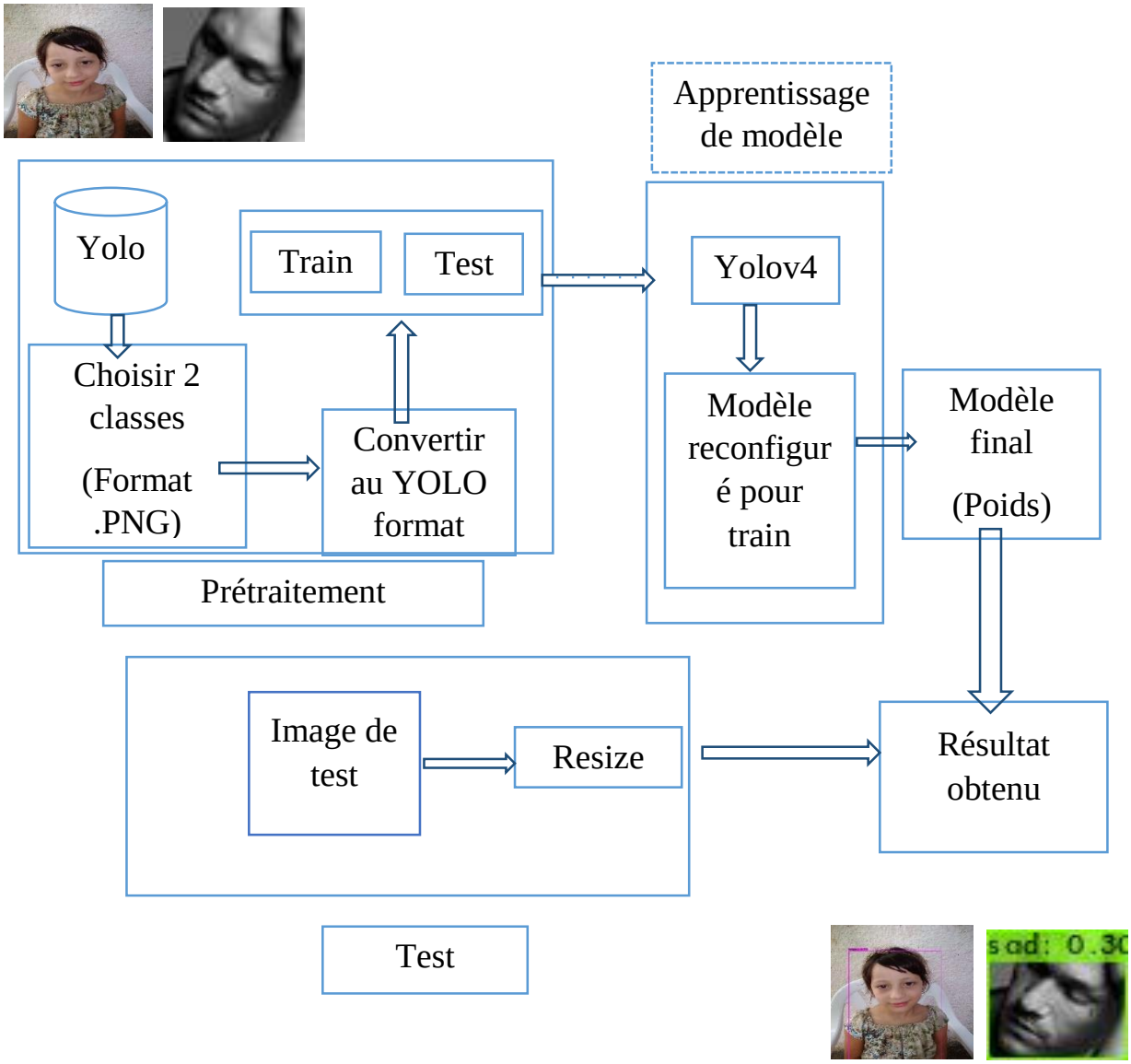


Figure3.1 : architecture générale de notre système de reconnaissance

2.1.1. Le choix de la base de données

Afin de développer un model basée sur YOLOv4 il est nécessaire de disposer une base de données que nous divisons en deux parties : la première partie pour effectuer l'apprentissage et la deuxième pour tester le réseau obtenu et déterminer ses performances.

Dans notre cas, le data set est acquis à partir d'un site Web bien connu (Kaggle), il contient un total de 2000 images qui sont classées dans 2 classes : happy, sad.

- **Kaggle :**

Kaggle, une filiale de Google LLC, est une communauté en ligne de spécialistes des données et de l'apprentissage automatique. Kaggle permet aux utilisateurs de trouver et de publier des ensembles de données, d'explorer et de construire des modèles dans un environnement de science des données basé sur le Web, de travailler avec d'autres scientifiques des données et ingénieurs d'apprentissage automatique, et de participer à des concours pour résoudre des défis de science des données.

Kaggle a démarré en 2010 en proposant des concours d'apprentissage automatique et offre désormais également une plateforme de données publique, un banc de travail en nuage pour la science des données et une formation à l'intelligence artificielle. Ses principaux collaborateurs sont Anthony Gold bloom et Jeremy Howard. Nicholas Gruen était le président fondateur, suivi par Max Levchin. Une levée de fonds a eu lieu en 2011 valorisant l'entreprise à 25 millions de dollars. Le 8 mars 2017, Google a annoncé qu'il rachetait Kaggle

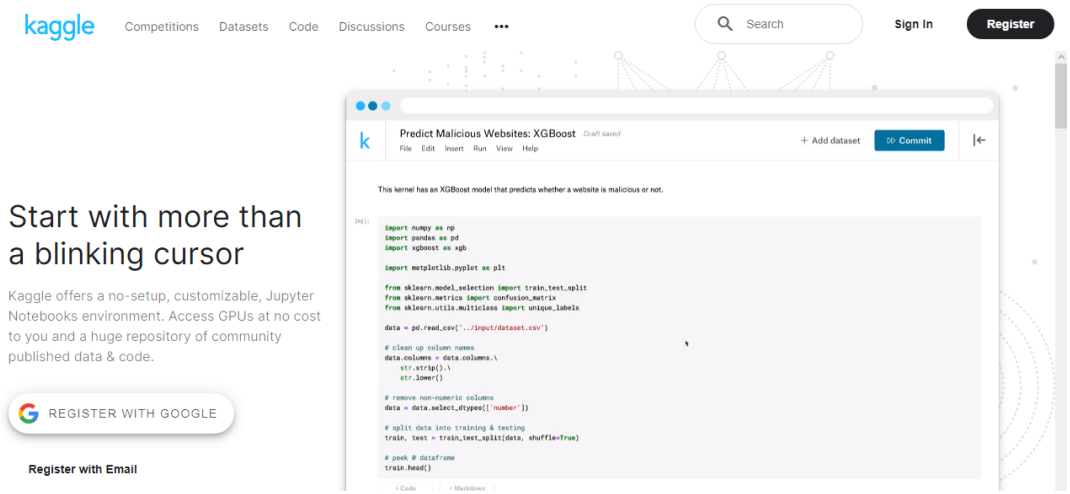


Figure 3.2 : Kaggle

2.1.2. Prétraitement de données

1) Choisir 02 classes

Cette étape consiste à appliquer quelques traitements sur les images en format PNG tel que le choix des deux classes (happy, sad) après conversionl'annotation de chaque image au format d'annotation Yolo.

La création des 02 classes de format PNG (happy, sad), consiste à lire chaque fichier d'annotation (. Png). Après on teste si la personne est happy ou non (sad).

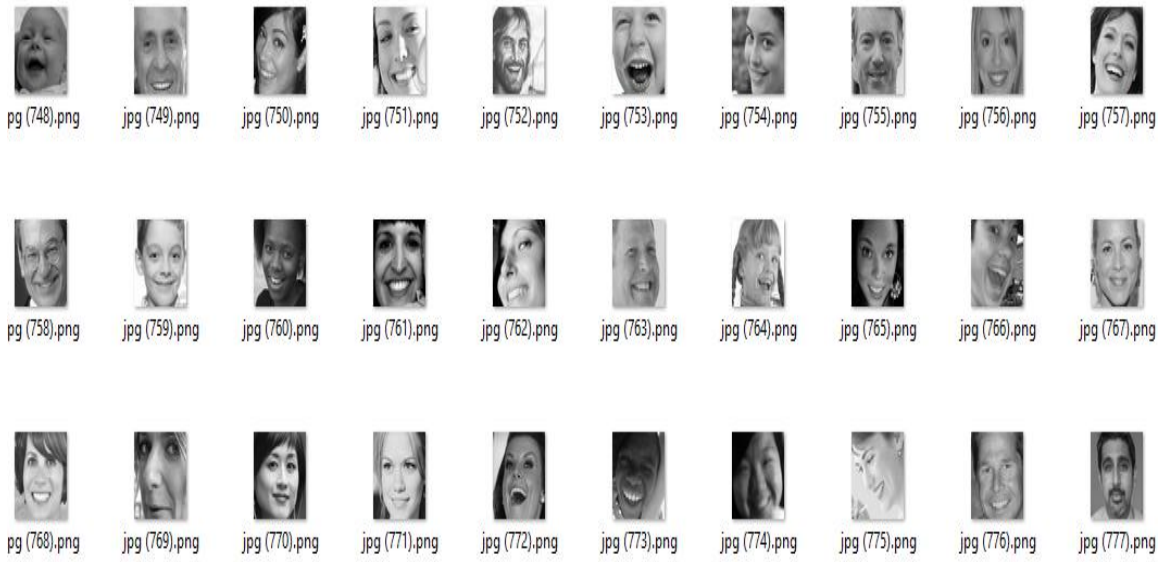


Figure 3.3 : Exemple d'images de la base de données de la classe happy

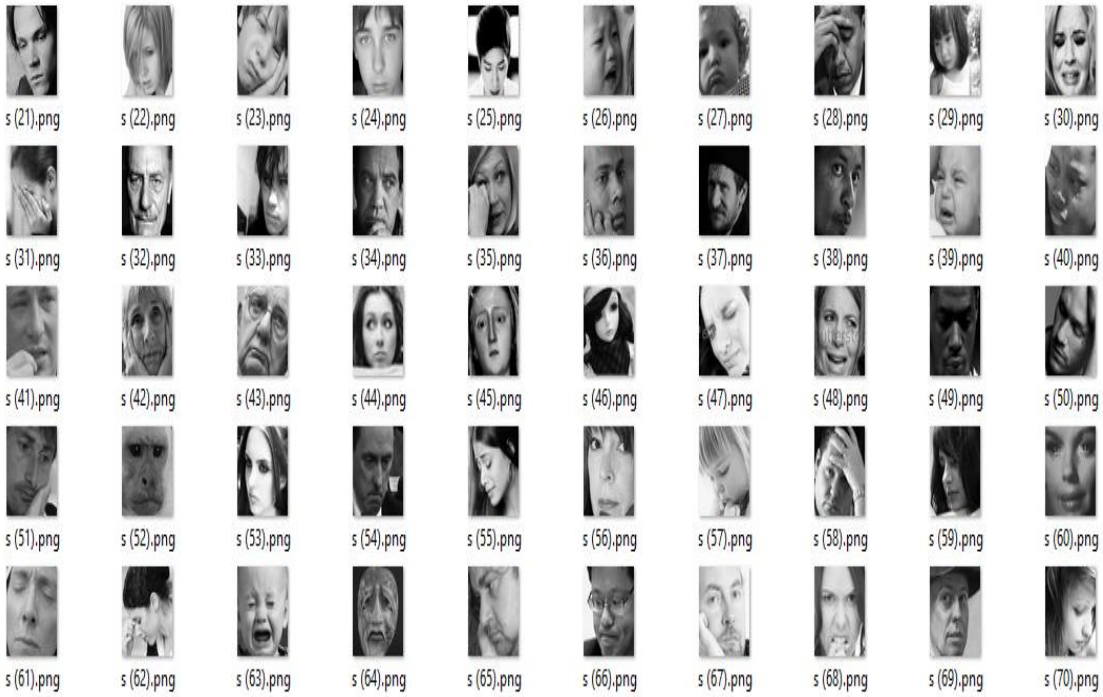


Figure 3.4 : Exemple d'images de la base de données de la classe sad

Labelling tool :

LabelImg est un outil graphique d'annotation d'images. Il est écrit en Python et utilise Qt pour son interface graphique.

Les annotations sont enregistrées sous forme de fichiers XML au format PASCAL VOC, le format utilisé par ImageNet. En outre, il supporte également les formats YOLO et CreateML [55].

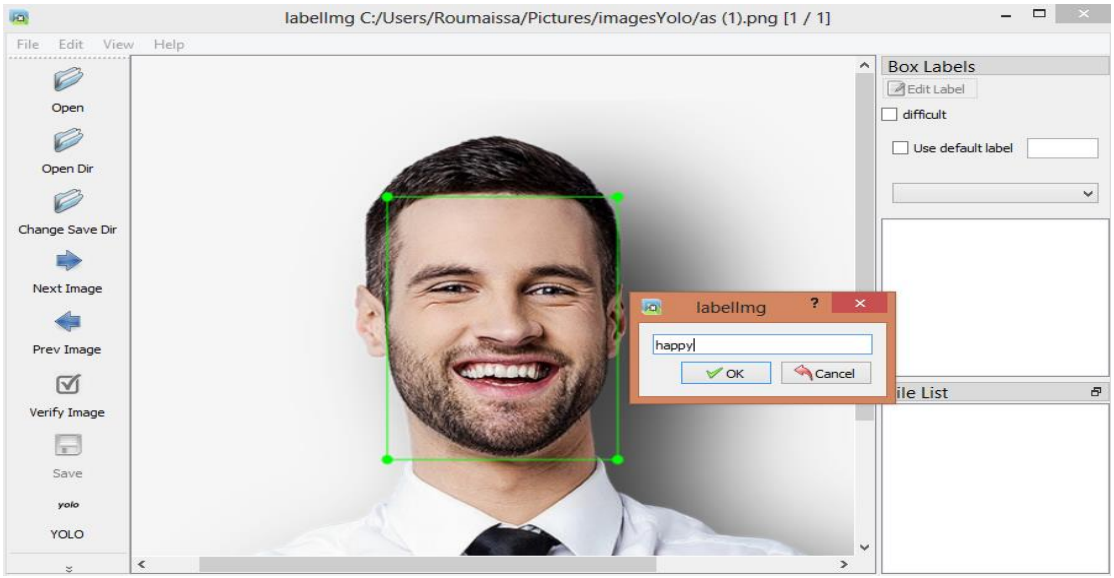


Figure 3.5 : LabelImg tool.

2) Conversion au format yolo

La deuxième étape consiste à créer l'annotation des sous-classes choisis sous le format yolo, Les valeurs de la boîte englobant de model yolo sont normalisées entre 0 et 1. Chaque fichier d'annotation (im (3). Txt) contient une ligne représente une émotion qui est représenté par numéro de la classe et boîte d'englobante dans l'image. Une ligne a le format suivant :

<Numéro de classe> <centre_x> <centre_y> <largeur> <hauteur>

Numéro de classe : L'index de la classe dans la liste des classes

Centre_x : La valeur x du centre normalisé de la boîte englobante

Centre_y : Valeur normalisée du centre y de la boîte englobante

Largeur : La valeur de la largeur normalisée de la boîte englobante.

Hauteur : La valeur de la hauteur normalisée de la boîte englobante

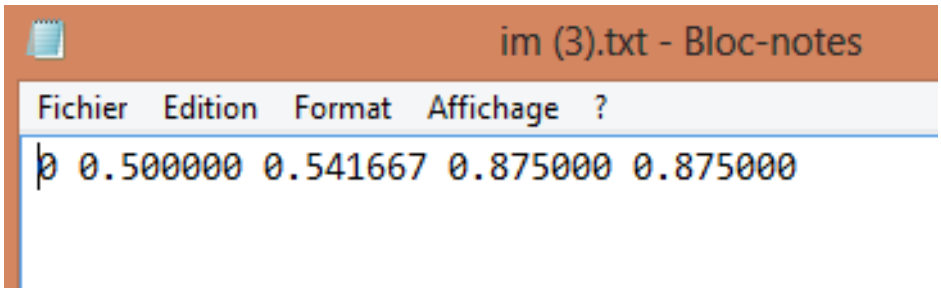


Figure 3.6 : Exemple d'annotation de la base de données.

3) Diviser la base (train et test)

La dernière étape dans ce module c'est la division des données en deux parties, un pour l'apprentissage (train.txt), et autre pour le test (test.txt). Ces deux fichiers contiennent les chemins (paths) des images de l'apprentissage et de test. Dans notre cas nous avons choisi 80% pour l'apprentissage et 20% pour test.

```
path_list = []
for current_dir, dirs, files in os.walk('.'):
  for f in files:
    if f.endswith('.png'):
      file_loc = image_path + '/' +f
      path_list.append(file_loc + '\n')
path_list_test = path_list[:int(len(path_list)* 0.20)]
path_list = path_list[int(len(path_list)* 0.20):]
with open('train.txt', 'w') as train:
  for i in path_list:
    train.write(i)

with open('test.txt', 'w') as test:
  for i in path_list_test:
    test.write(i)
i = 0
with open(image_path + '/' + 'classes.names', 'w') as cls, \
     open(image_path + '/' + 'classes.txt', 'r') as text:
     for l in text:
      cls.write(l)
      i += 1

with open(image_path + '/' + 'image_data.data', 'w') as data:
    data.write('classes =' + str(i) + '\n')
    data.write('train =' + image_path + '/' + 'train.txt' + '\n')
    data.write('test =' + image_path + '/' + 'test.txt' + '\n')
    data.write('names =' + image_path + '/' + 'classes.names' + '\n')
    data.write('backup = backup')
```

Figure 3.7 : Partie du code pour la division des data

2.1.3. Apprentissage de model yolov4

Ce module contient deux étapes principales qui sont :

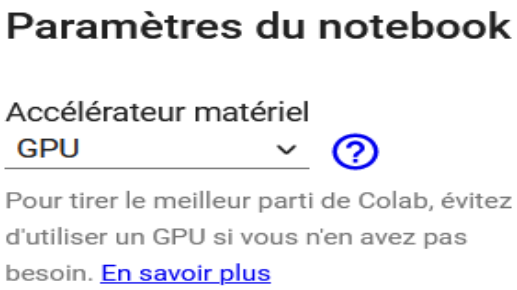
1) Télécharger le modèle yolov4

Cette étape consiste à connecter google Colab avec notre google drive, ensuite télécharger le fichier darknet qui contient le model yolov4 avec le command :

```
[ ]  !git clone https://github.com/AlexeyAB/darknet
```

Ensuite, nous avons préparé l'enivrement google Colab par la configuration de certains paramètres. Il s'agit de :

- Changer les paramètres du notebook, et choisir le matériel GPU pour le l'apprentissage de model.



- Accéder au fichier darknet, puis faire des changements dans le makefile pour activer OPENCV et le GPU et CUDA

```
[4]  %cd darknet/
     !sed -i 's/OPENCV=0/OPENCV=1/' Makefile
     !sed -i 's/GPU=0/GPU=1/' Makefile
     !sed -i 's/CUDNN=0/CUDNN=1/' Makefile
     !sed -i 's/CUDNN_HALF=0/CUDNN_HALF=1/' Makefile
     !sed -i 's/LIBSO=0/LIBSO=1/' Makefile
```

Après la configuration de Colab, nous avons téléchargé la base de données dans un dossier appeler **« amaniEtChaima »**. Ensuite, nous avons créé deux fichier **« classes.name »** qui

contient les noms de nos classes, et **« image_data.data »** qui contient les chemins **(paths)** au fichier **« train.txt »** et « **test.txt »** et **« classes.name »** et l'emplacement du répertoire **« backup »** pour enregistrer les poids **(weights)** du modèle au cours de l'entrainement. Télécharger les poids pré-entraînés de YOLOv4, au lieu d'entraîner un modèle à partir de zéro, nous avons utilisé les poids « **yolov4.conv.137 »** qui a été entraînés jusqu'à 137 couches convolutives.

2) Préparer le modèle pour le train

Cette étape consiste à changer les paramètres dans le fichier de configuration « Yolov4-train.cfg » de modèle pour adapter notre base de données choisi, les paramètres qui sont changés étaient calculés comme suite :

Le nombre de batch et subdivision (batch = 64, subdivision = 16) ou bien une valeur multiple à 32.

Définir la taille du modèle (width=608, height=608)

Changer la ligne max_batches en (classes*2000 donc 2*2000 =4000)

Changer les étapes de la ligne à 80% et 90% de max_batches (steps=1600,1900).

Remplacez la ligne classes=80 par notre nombre d'objets dans chacune des 3 couches [yolo] : classes = 2

Changer [filters=255] en filters = (classes + 5) x 3 dans les 3 [convolutions] avant chaque couche [yolo]. On doit garder à l'esprit qu'il doit seulement s'agir de la dernière convolution avant chacune des couches yolo (Filtres = (2+5) *3=21)).

L'étape suivante est d'exécuter la commande « ! make » pour construire les fichiers dans darknet, ensuite le command « ! Chmod +x. /darknet » pour autoriser à utiliser le fichier darknet.exe qui permet de lancer l'apprentissage.

L'apprentissage du modèle yolo lancé par une commande a besoins de paramètres comme arguments d'entrer qui sont le nom de fichier « image_data.data » et le fichier de configuration (*. Cfg), les poids pré-entraînés (pour la première fois on utilise yolov4.conv.137) et après on utilise toujours les best. Weights. Ce changement de nom du fichier des poids est dû aux coupures et aux interruptions causés par internet et par la politique de colab qui arrête toutes les ressources chaque fois à mi- nuit.

```
!./darknet detector train data/image_data.data cfg/yolov4_train.cfg /content/drive/MyDrive/amani/darknet/backup/yolov4_train_last.weights  -dont_show
```

L’utilisation du flag ‘dont show’ car il n'y a pas d'écran d'affichage dans google Colab pour tracer le graph de mAP dans l’apprentissage. La valeur de perte affichée à chaque itération de l’apprentissage, et la valeur de mAP elle est calculée dans chaque 4 étape (epoch).

```
 1701: 1.852833, 1.852833 avg loss, 0.000130 rate, 9.993431 seconds, 54432 images, -1.000000 hours left
Loaded: 0.000049 seconds
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 139 Avg (IOU: 0.000000), count: 1, class_loss = 0.000000, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 150 Avg (IOU: 0.000000), count: 1, class_loss = 0.000004, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 161 Avg (IOU: 0.922732), count: 2, class_loss = 0.934315, iou_loss = 0.106768, total_l
 total_bbox = 190, rewritten_bbox = 0.000000 %
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 139 Avg (IOU: 0.000000), count: 1, class_loss = 0.000000, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 150 Avg (IOU: 0.000000), count: 1, class_loss = 0.000004, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 161 Avg (IOU: 0.795943), count: 2, class_loss = 0.922352, iou_loss = 0.100397, total_l
 total_bbox = 192, rewritten_bbox = 0.000000 %
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 139 Avg (IOU: 0.000000), count: 1, class_loss = 0.000000, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 150 Avg (IOU: 0.000000), count: 1, class_loss = 0.000004, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 161 Avg (IOU: 0.750642), count: 3, class_loss = 1.504336, iou_loss = 0.252472, total_l
 total_bbox = 195, rewritten_bbox = 0.000000 %
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 139 Avg (IOU: 0.000000), count: 1, class_loss = 0.000000, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 150 Avg (IOU: 0.000000), count: 1, class_loss = 0.000004, iou_loss = 0.000000, total_l
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 161 Avg (IOU: 0.838036), count: 2, class_loss = 1.007441, iou_loss = 0.122287, total_l
 total_bbox = 197, rewritten_bbox = 0.000000 %
v3 (iou loss, Normalizer: (iou: 0.07, obj: 1.00, cls: 1.00) Region 139 Avg (IOU: 0.000000), count: 1, class_loss = 0.000000, iou_loss = 0.000000, total_l
```

Figure 3.8 : l’affichage du processus de l’apprentissage.

Le résultat de l’apprentissage obtenu avec mAP@0.5 c’est 82%. Dans le tableau suivant nous avons les résultats obtenus pour chaque classe :

Class	*Précision*	*Rappel*	*F1-score*	*Average IoU*	*MAP@0.5*
Happy	**0.97**	**0.96**	**0.96**	**77.56 %**	**97.45%**
Sad	**0.75**	**0.51**	**0.53**	**59.58%**	**66.52%**

Tableau 3.1. Les résultats de l'apprentissage pour les deux classes

3.L'implémentation

Pour ré-entrainer le modèle yolov4 sur les sous-classes la base de données, nous avons utilisé l'environnement libre de google Colab aussi des libraires opencv en langage python.

3.1. Présentation des outils de développement

3.1.1. Matériel

Le matériel utilisé est un PC personnel HP avec un 4GB capacité mémoire, et un processeur Intel® Celeron® N4000 CPU @ 1.10GHz, avec Windows 10 édition intégrale, service pack 1 64 bits type système.

3.1.2. Logiciel

3.1.2.1. Google-Colaboratory

Pendant de nombreuses années, Google a développé un outil de développement appelé Colaboratory (Google colab). Aujourd'hui, Google a rendu Colaboratory gratuit pour une utilisation publique.

Google colab est un environnement de bloc-notes gratuit qui fonctionne entièrement dans le nuage.

Permet, de modifier des documents, de la même manière que travailler avec Google Docs. Colab prend en charge de nombreuses bibliothèques d'apprentissage automatique populaires qui peuvent être facilement chargées dans le carnet de notes. Avec Google colab, il suffit de quelques lignes de code pour importer un ensemble de données d'images, entraîner un classificateur d'images et évaluer le modèle.

Les notebooks Colab exécutent le code sur les serveurs en nuage de Google, ce qui signifie que les utilisateurs peuvent exploiter la puissance du matériel Google, notamment les GPU et

les TPU, quelle que soit la puissance de leur machine. En utilisant un langage de programmation soit : C++ ou Python [56].

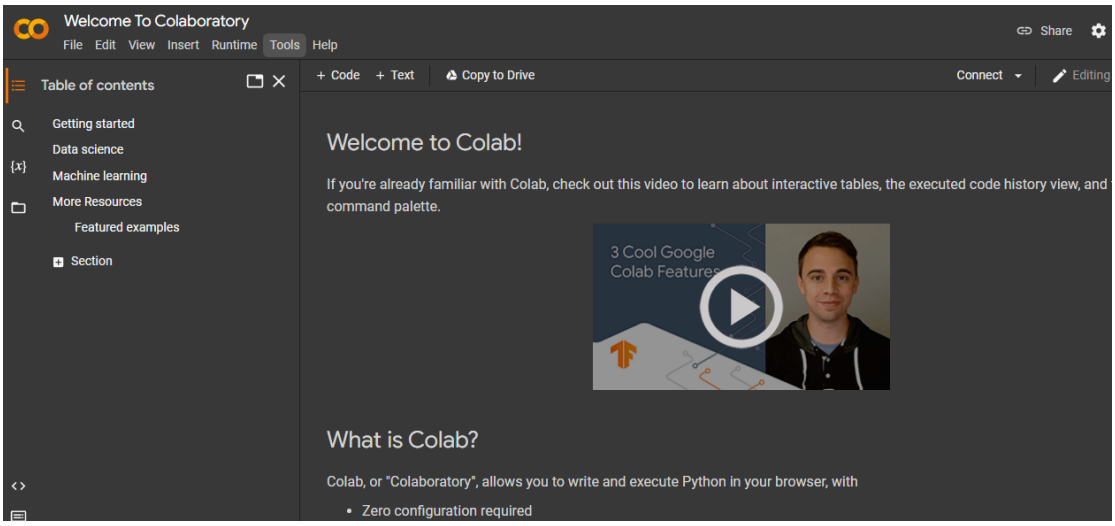


Figure 3.9 : L`environnement Google Colaboratory.

3.1.2.1.1. Pourquoi utiliser Google Colab

Il y a plusieurs raisons pour l'adopter Google Colab au lieu d'une simple instance de Jupyter Notebook.

1) Bibliothèques préinstallées

La distribution Anaconda de Jupyter Notebook est livrée avec plusieurs bibliothèques de données préinstallées, telles que Pandas, NumPy, Matplotlib, ce qui est génial. Google Colab, quant à lui, fournit encore plus de bibliothèques d'apprentissage automatique préinstallées, telles que Keras, TensorFlow et PyTorch.

2) Sauvegardé dans le Cloud

Lorsque vous choisissez d'utiliser un simple carnet Jupyter comme environnement de développement, tout est enregistré sur votre machine locale. Si vous êtes prudent en matière de confidentialité, il s'agit peut-être d'une fonctionnalité que vous préférez. Toutefois, si vous souhaitez que vos carnets soient accessibles à partir de n'importe quel appareil avec une simple connexion Google, alors Google Colab est la solution. Tous vos carnets Google Colab sont enregistrés sous votre compte Google Drive, tout comme vos fichiers Google Docs et Google Sheets.

3) Collaboration

La fonction de collaboration est une autre fonctionnalité intéressante de Google Colab. Si vous travaillez avec plusieurs développeurs sur un projet, l'utilisation de Google Colab notebook est idéale. Tout comme vous collaborez sur un document Google Docs, vous pouvez coder avec plusieurs développeurs à l'aide d'un bloc-notes Google Colab. En outre, vous pouvez également partager votre travail terminé avec d'autres développeurs.

4) Utilisation gratuite des GPU et TPU

Google Research vous permet d'utiliser ses GPU et TPU dédiés pour vos projets personnels d'apprentissage automatique. D'après mon expérience, pour certains projets, l'accélération des GPU et TPU fait une énorme différence, même pour de petits projets.

C'est l'une des principales raisons pour lesquelles je code tous mes projets éducatifs sur Google Colab. En outre, comme il utilise les ressources de Google, les opérations d'optimisation du réseau neuronal ne perturbent pas mes processeurs et mon ventilateur de refroidissement ne s'emballe pas [57].

3.1.2.2. Python3

Python est un langage de programmation interprété, orienté objet, de haut niveau et doté d'une sémantique dynamique. Ses structures de données intégrées de haut niveau, combinées au typage dynamique et à la liaison dynamique, le rendent très attrayant pour le développement rapide d'applications, ainsi que pour une utilisation en tant que langage de script ou de colle pour connecter des composants existants.

La syntaxe simple et facile à apprendre de Python privilégie la lisibilité et réduit donc le coût de la maintenance des programmes. Python supporte les modules et les packages, ce qui encourage la modularité des programmes et la réutilisation du code. L'interpréteur Python et la bibliothèque standard étendue sont disponibles gratuitement sous forme de source ou de binaire pour toutes les principales plates-formes et peuvent être distribués librement.

En général, les programmeurs tombent amoureux de Python en raison de la productivité accrue qu'il procure. Comme il n'y a pas d'étape de compilation, le cycle édition-test-débogage est incroyablement rapide. Le débogage des programmes Python est facile : un bogue ou une mauvaise entrée ne provoquera jamais une erreur de segmentation. Au contraire, lorsque l'interpréteur découvre une erreur, il lève une exception. Si le programme n'attrape pas l'exception, l'interpréteur imprime une trace de la pile.

Un débogueur au niveau de la source permet d'inspecter les variables locales et globales, d'évaluer des expressions arbitraires, de définir des points d'arrêt, de parcourir le code ligne par ligne, et ainsi de suite. Le débogueur est écrit en Python même, ce qui témoigne de la puissance introspective de

Python. D'un autre côté, le moyen le plus rapide de déboguer un programme est souvent d'ajouter quelques instructions d'impression au code source : le cycle rapide édition-test-débogage rend cette approche simple très efficace [58].

3.1.2.3. OpenCV

OpenCV (Open Source Computer Vision Library) est une bibliothèque logicielle open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception artificielle dans les produits commerciaux. Étant un produit sous licence BSD, OpenCV permet aux entreprises d'utiliser et de modifier facilement le code. La bibliothèque compte plus de 2500 algorithmes optimisés, ce qui inclut un ensemble complet d'algorithmes de vision par ordinateur et d'apprentissage automatique classiques et de pointe. Ces algorithmes peuvent être utilisés pour détecter et reconnaître des visages, identifier des objets, classer des actions humaines dans des vidéos, suivre les mouvements de la caméra, suivre des objets en mouvement, extraire des modèles 3D d'objets, etc. [59]. Dans le processus de l'apprentissage nous avons utilisé la version 3.2.0.

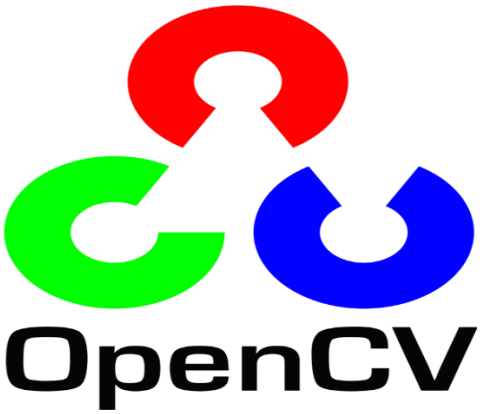


Figure 3.10 : Logo de openCV

3.1.2.4. Darknet

Figure 3.11 : Logo de darknet

Darknet est un cadre pour former des réseaux neuronaux, il est open source et écrit en C/CUDA et sert de base à YOLO. Darknet est utilisé comme cadre pour l'entraînement de YOLO, ce qui signifie qu'il définit l'architecture du réseau. [60]

3.1.2.5. You only look once (YOLO)

Est un système de détection d'objets en temps réel à la pointe de la technologie. Sur un Pascal Titan X, il traite les images à 30 FPS et a un mAP de 57,9% sur COCO test-dev. [61]

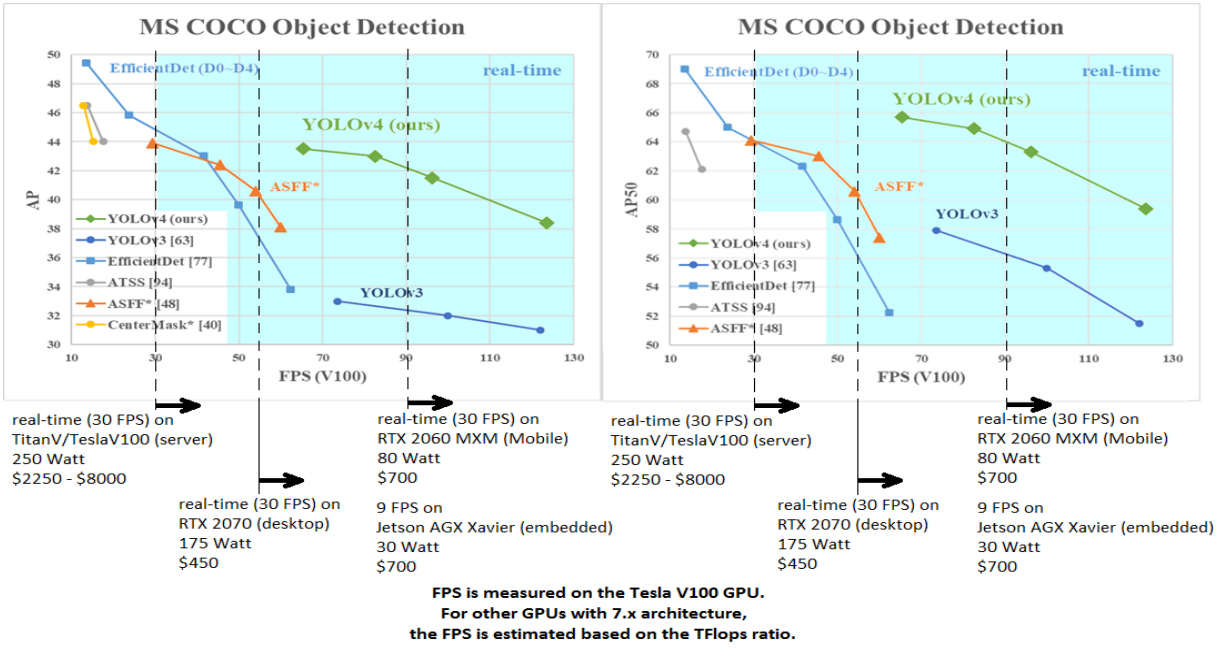


Figure 3.12 : comparaison entre les déférents model YOLO.

3.1.2.6. Chargement de model

Dans cette étape nous utiliserons la fonction prédéfinie de opencv « cv2.dnn.readNet » pour charger le réseau en mémoire. Elle détecte automatiquement la configuration et le cadre en fonction du nom du fichier spécifié. Dans notre cas, il s'agit d'un fichier «. Weight » Nous avons créé une liste qui contient les labels des objets spécifiques que nous avons choisi (9 sous classes). Après nous avons récupéré les couches de notre modèle pour obtenir les couches de sortie. Par la fonction « getUnconnectedOutLayers ())

3.1.2.7. Prétraitement de l'image de test

Nous utilisons la fonction « cv2.dnn.blobFromImage » qui renvoie un blob qui est notre image d'entrée après soustraction de la moyenne, normalisation et échange de canaux.

➢ Pour la soustraction de la moyenne nous utilisons « np.mean » puis on soustrait chaque pixel de l'image par cette moyenne.

➢ Après avoir effectué la soustraction de la moyenne, nous pouvons mettre à l'échelle nos images par un certain facteur. Cette valeur est par défaut de `1.0` (c'est-à-dire, pas de mise à l'échelle) mais nous pouvons également fournir une autre valeur (1/255).

➢ Resize de l'image, Nous fournissons ici la taille spatiale attendue par le réseau neuronal convolutif, dans notre cas nous utilisons (608*608).

3.1.2.8. La Détection

Après les prétraitements de l'image nous passons à la phase de détection.

Cette étape consiste à :

• Passez image blob de résultat de prétraitement dans l'algorithme de model yolo par « net.setInput (blob) ».

• Utilisez « net.forward () » pour transmettre le blob à la couche de sortie que nous avons générer dans la première phase de chargement de modèle et générer le résultat.

Le résultat de la détection c'est tous les boites englobantes détecter à la dernière couche de modèle (sur les 3 scale) qui passent le degré de confiance >0.5 avec le nom de la classe et le degré de confiance. Nous utilisons la fonction NMS dans opencv « cv2.dnn.NMSBoxes » pour effectuer la Suppression Non-Maximale. On utilise un seuil de score et un seuil de NMS comme arguments.

3.1.3. Test et résultat

Dans cette partie nous allons tester/valider les performances du model yolov4 réentrainé sur les deux sous classes sélectionnés à partir des photos PNG.

Ce test est établi sur des données choisies aléatoirement sur le dossier test, et on a obtenu des résultats.

(a) (b)

(c) (d)

(e) (f)

Figure 3.13 : Le Résultat Obtenu de Détection par Yolov4

3.1.4. Discussion

A partir des résultats obtenus (tableau 3.1 et figure 3.13) nous avons vu que la performance du modèle yolo sur la détection des différentes réactions dans l'image est plus précise surtout dans les classes avec un mAP supérieur à 80% (heureux ou triste) et dans le test il donne un bon résultat de détection quelle que soit la position du visage.

- Les points forts

➢ Le Système est puissant dans une image de complexité élevée (plusieurs class comme la figure 3.12 (d), mauvaise qualité (e), mauvaise éclairage (figure 3.13).

➢ Le Système capable à travaille sur des images ou bien vidéo/cam avec une résolution (608) (en gardant un bon rapport temps/fps).

➢ La Plupart des boites englobantes cadrent bien les visages.

La précision (acuraccy) du modèle est en général très intéressante par apport aux modèles de détection en temps réel.

- Les problèmes

Problème de limite de l'utilisation de GPU.

➢ Chaque fois il y'a initialisation des fichiers des poids sur darknet à cause de la politique de colab (l'historique n'est pas sauvegardé).

➢ Si on lance le train deuxième fois il ne continue pas l'apprentissage sur le dernier arrêt.

- Le problème des bases de données, soit trop gros, soit problème d'étiquetage.
- L'apprentissage prend des temps incroyables (plusieurs itérations = des semaines).

4.Conclusion

Nous avons présenté dans ce chapitre la conception et l'implémentation de l'approche de détection des émotions basée sur le Deep Learning. Pour cela, nous avons utilisé le réseau de neurones convolutif (YOLOv4) qui a été pré-entrainé jusqu'à 137 couches de convolution et on a réentrainer sur 2 classes de la base de données de format PNG pour que nous assurons une meilleure précision avec le moins de temps. Pour une détection de plus des visages (émotions), il faut que vous entraîniez l'algorithme sur votre propre BDD.

Conclusion générale

Conclusion générale

La capacité de l'ordinateur à reconnaitre les expressions faciales de la personne constitue un nouveau défi pour la recherche scientifique moderne. D'autant plus que la communication entre humain et les appareils électroniques a augmenté considérablement, alors les chercheurs cherchent à développer des programmes intelligents capables de comprendre l'état affectif d'un être humain.

Notre travail traite la reconnaissance automatique des émotions à partir des expressions faciales, dans le but de suivi la productivité des employées, patients autistes, et patients en soins intensifs. Pour réaliser ce système, nous nous sommes basées sur les réseaux de neurones convolutifs profonds (Deep-CNN). Ces derniers sont utiles pour la modélisation de n'importe quel problème difficile à décrire avec des modèles physiques et mathématiques en raison de la capacité des réseaux de neurones d'apprendre par des exemples. Nous avons utilisé le modèle CNNyolov4 parce qu'il a prouvé sa performance de la détection d'objet en temps réel.

Pour l'entrainement du modèle, nous avons utilisé une base de données qui nous divisons en deux parties la première pour l'apprentissage sur les deux classes (happy, sad) et la deuxième partie pour le test (résultat). Nous avons obtenu des résultats satisfaisants de détection.

Pour conclure, les objectifs tracés dans ce travail ont été atteints, mais il reste toujours des perspectives et des améliorations possibles peuvent encore être réalisé dans le futures, tels que :

- Tester notre modèle sur d'autres bases des données plus volumineuses pour donner plus de performance à notre système.
- Utiliser un autre modèle de détection d'objet en temps réel tel que le modèle SSD et faire une comparaison avec d'autre modèle de détection en temps réel.
- Changer le backbone de yolov4 et utiliser le MobileNet et comparer le résultat obtenu avec le Darknet.
- Utiliser les algorithmes d'amélioration de la qualité des images pour améliorer la détection.

Références bibliographiques

Références bibliographiques

[1] National Science and Technology Concil (NSTC). Face recongnition.Comittee on Technology, page 10, 7 Aout 2006.

[2] Cabal (Christian), 2003. Rapport sur les méthodes scientifiques d'identification des personnes à partir de données biométriques et les techniques de mise en oeuvre. Office parlementaire d'évaluation des choix scientifiques et technologiques. Enregistré à la présidence de l'Assemblée nationale Le 16 juin 2003 sous N° 938. 70 pages R. Brunelli and T. Poggio, "Face recognition: Features vs. templates," IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 10, pp. 1042–1053, Oct. 1993.

[3] [en ligne] reconnaissance faciale

< https://labiometrie.wordpress.com/2017/02/12/reconnaissance-faciale/ >).

[4] F. Perronnin and J.-L. Dugelay. "Introduction à la biométrie – Authentification des individus par traitement audio-vidéo". Traitement du signal, Vol. 19, No. 4, 2002).

[5] Dorian Dozolme. La détection d'une expression faciale incongrue par rapport `à un modèle de situation émotionnel : un d´défi neurocognitif ? Neurosciences. Université Paris Sud - Paris XI, 2014.

[6] BETTAHAR. A et SABER. F, « Extraction des caractéristiques pour l'analyse Biométrique d'un visage ». (Thèse de magister).

[7] Walid Hizem. Capteur intelligent pour la reconnaissance de visage. PhD thesis, Evry, Institut national des télécommunications, 2009.

[8] Cheng-Chin Chiang, Wen-Kai Tai, Mau-Tsuen Yang, Yi-Ting Huang, and Chi-Jaung Huang. A novel method for detecting lips, eyes and faces in real time. Real-time imaging, 9(4) :277–287, 2003.

[9] Faiza Abdat. Reconnaissance automatique des émotions par données multimodales : expressions faciales et signaux physiologiques. Université de Metz, France, 2010.

[10] Mathieu Van Wambeke, Benoıt Macq, and Christian Van Brussel. Reconnaissance et suivi de visages et implémentation en robotique temps-réel. Université Catholique de Louvain Ecole Polytechnique de Louvain. Mémoire de fin d'études, 2010.

[11] Sofiane Boudjellal. Détection et identification de personne par méthode biométrique. PhD thesis, Université Mouloud Mammeri, 2012.

[12] Souhila Guerfi Ababsa. Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2d/3d. Evry-Val d'Essonne, 2008.

[13] F. Khalfi. Reconnaissance automatique des émotions par données multimodales : expressions faciales et des signaux physiologiques. PhD thesis, Université Paul Verlaine de Metz, France, 2010.

[14] E. Couzon and F. Dorn. Les émotions : développer son intelligence émotionnelle. ESF editeur, 2011.

[15] Karen L Schmidt and Jeffrey F Cohn. Human facial expressions as adaptations: Evolutionary questions in facial expression research. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 116(S33):3–24, 2001.

[16] Charles Darwin. 1965. the expression of the emotions in man and animals. London, UK: John Marry, 1872.

[17] Brais Martinez, Michel F Valstar, Bihan Jiang, and Maja Pantic. Automatic analysis of facial actions: A survey. IEEE transactions on affective computing, 2017.

[18] Ying-Li Tian, Takeo Kanade, and Jeffrey F Cohn. Facial expression analysis. In Handbook of face recognition, pages 247–2.

[19] Mira Jeong and Byoung Chul Ko. Driver's facial expression recognition in real-time for safe driving. Sensors, 18(12) :4270, 2018.

[20] Laurence Vidrascu 2007. Analyse et détection des émotions verbales dans les interactions orales, page 09.

[21] J. Besson "Le travail organique en analyse psycho-organique". Colloque internationnal de Sigulda, Lettonie. Association d'analyse psycho-organique. Juillet, 2007

[22] " Physiologie humaine " groupe médicale, 2 éme édition scherwood, 2004. P 124-126

[23] S.J. Chung "L'expression et la perception de l'émotion extraite de la parole spontanée : évidences du coréen et de l.anglais" Thése de doctorat, Institut de Linguistique et Phonétique

Générales et Appliquées, Université de la Sorbonne Nouvelle, France. Soutenu en Juin, 2000. P 40-43.

[24] C. Maaoui, A. Pruski "A comparative study of SVM kernel applied to emotion recognition from physiological signals", IEEE Transactions on Neural Networks. Laboratoire d'Automatique et des Systemes Cooperatifs, Universite de Metz. France, 2008. P 1-2.

[25] A. Rivière, B. Godet "L'affectiveComputing : rôle adaptatif des émotions dans l'interaction Homme – Machine" Rapport, Université Charles de Gaulle, Lille, France. 2003. P 9-12; 33-38.

[26] Affective computing and affective learning--methods, tools and prospects. EduAction. Electronic Education Magazine, 1(5), 16–31.

[27] 3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2158–2167.https://doi.org/10.1109/CVPR.2018.00230.

[28] Human-Computer Systems Interaction: Backgrounds and Applications 3. Advances in Intelligent Systems and Computing, 300, 51–62. https://doi.org/10.1007/978-3-319-08491-6.

[29] Deep Learning Yann LeCun, Yoshua Bengio Geoffrey Hinton. 2015.

[30] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning.

[31] Caiming.Z et Yang.L "Study on artificial intelligence: The state of the art and future prospects" ScienceDirect, Journal of Industrial Information Integration page 01 (23) 8 Mai 2021.

[32] Marco. A « Comme Exigence Partielle de la Maitrise en Philosophie » Mémoire présenté à l'université du Québec à Trois-Rivières, page 04, Mars 1992.

[33] https://www.bial-r.com/2019/05/22/comprendre-le-machine-learning-et-le-deep-learning. Consulter le 26/04/2022.

[34] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. http://www.deeplearningbook.org. MIT Press, 2016.

[35] Tom MITCHELL, Machine learing, McGraw-Hill Science/Engineering/Math, 1997.

[36] Encyclopædia Universalis « Apprentissage Profond ou Deep Learning » https : //www.universalis.fr/encyclopedie/apprentissage-profond-deep-learning/1-differentstypes-d-apprentissage-machine, consulter le 06/05/2022.

[37] Arjun Panesar « Machine Learning and AI for Healthcare» page 74-75-76 Coventry, UK 2021.

[38] Tanay Agrawal « Hyperparameter Optimization in Machine Learning » page 03 Bangalore, Karnataka, India 2021.

[39] [En ligne]. Available : https://fr.mathworks.com/discovery/deep-learning.html. Consulter le 06/05/2022.

[40] [En ligne]. Available : https://www.mathworks.com/discovery/deep-learning.html. Consulter le 06/05/2022.

[41] Corentin HARDY. Contribution au développement de l'apprentissage profond dans les systèmes distribués. Université RENNES 2019.

[42] actualiteinformatique. Disponible à l'adresse https://actualiteinformatique.fr/intelligence-artificielle/défintion-convolutionel-network consulter le 25/05/2022.

[43] Asifullah Khan et autre « A Survey of the Recent Architectures of Deep Convolutional Neural Networks » Publié dans Artificial Intelligence Review DOI, 21 avril 2020.

[44] https://www.analyticsvidhya.com/blog/2020/10/what-is-the convolutional-neural-network-architecture. Consulter le 29/05/2022.

[45] Louam Abdelhak Bilal « Deep Learning basé sur les méthodes de réduction pour la reconnaissance de visage » Mémoire De Master Sciences et Technologies Télécommunication Réseaux et Télécommunication, Université Mohamed Khaider de Biskra, 2019, page 14.

[46] Trad Houssem Eddine « La détection d'objet avec OpenCV et deep learning » Mémoire De Master Sciences et Technologies Electronique Réseaux Télécommunication, Université Mohamed Khider de Biskra, 30 septembre 2020, page 14.

[47] Oulmi Mehdi et Kaloune Salim « Classification d'objets avec le Deep Learning » Mémoire de Master En Informatique, Université Akli Mohand Oulhadj de Bouira, 2018, page 30.

[48] https ://inside-machonelearning.com/cnn-couche-de-convolution.consulter le 29/05/2022.

[49] Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptuel compréhension du réseau neuronal convolutif-une approche d'apprentissage profond. Procedia computer science, 132, 679-688.

[50] Réseau neuronal convolutif (CNN) : étape 1(b)- couche ReLU.

https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-1b-relu-layer. Consulter le 02/06/2022.

[51] Normalisation par lots dans les réseaux de neurones convolutifs.

https://www.baeldung.com/cs/batch-normalization-cnn. Consulté le 02/06/2022.

[52] Cnn-introduction à la couche de mise en commun. https://www.geeksforgeeks.org. Consulter le 02/06/2022.

[53] a-beginners-guide-to-convolutional-neural-networks-cnn. https://heartbeat.fritz.ai.

Consulter le 03/06/2022.

[54] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, RealTime Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788. doi: 10.1109/CVPR.2016.91.

[55] https://pypi.org/project/labelImg Consulter le 03/06/2022.

[56] Machine learning. Disponible à l'adresse https://colab.research.google.com/notebooks/intro.ipynb#scrollTo=OwuxHmxllTwN Consulter le 03/06/2022.

[57] 4 Reasons Why You Should Use Google Colab for Your Next Project. Disponible à l'adresse https://towardsdatascience.com/4-reasons-why-you-should-use-google-colab-for-your-next-projectb0c4aaad39ed Consulter le 04/06/2022.

[58] What is Python ? Executive Summary Disponible à l'adresse https://www.python.org/doc/essays/blurb/ Consulter le 03/06/2022.

[59] About - OpenCV https://opencv.org/about dernier consulter le 03/06/2022.

[60] https://findanyanswer.com/what-is-yolo-darknet Consulter le 03/06/2022.

[61] https://findanyanswer.com/What is Yolo you only look once Consulter le 03/06/2022.