REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre :……. /M2/2023

Mémoire

présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **Systèmes Informatiques, optimisation et décision (SIOD)**

Conception et réalisation d'un système pour assister les malvoyants

---

Par :

Harzallah Hakima

Soutenu le 20 /6/2023, devant le jury composé de :

Mme. Ben Sghier Nadia	MCB	Président
M.Djaber Khaled	MAA	Rapporteur
M.Meadi Mohamed Nadjib	MCB	Examinateur

Année universitaire 2022-2023

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

Année universitaire 2022-2023

Remerciements

Nous remercions tout d'abord **ALLAH** de nous avoir donné le courage d'accomplir ce travail.
Je tiens à exprimer mes remerciements à mon encadreur
M. Khaled Djaber
De m'avoir soutenue et fait confiance durant mon projet avec une
Grande patience.
Avec son expérience dans la recherche et l'enseignement, avec ses conseils,
j'ai pu découvrir le monde de la recherche
scientifique dans le domaine du traitement d'image
Mes remerciements et ma profonde reconnaissance
s'adressent à mesdames les membres de jury
En second lieu, je remercie chaleureusement mes chers parents, Mon père, ma mère, mes frères, et soeur pour leurs sacrifices, aides, soutiens et encouragements et à tous ceux qui de près ou de loin ont
contribués au bon déroulement de ce mémoire.
Je souhaite à présent adresser mes sincères remerciements
à toutes les personnes avec qui j'ai eu la chance
de travailler et à qui j'ai eu l'honneur de
côtoyer avant et pendant mon mémoire, et à tous les enseignants, intervenants de l'Université Mohamed Khider BISKRA.

Hakima Harzallah

Dédicaces

Je dédie ce travail,
Fruit de nombreuses années d'étude à :
à mes chers
*Parents pour leur patience**,***
***Ma chère mère**. Merci pour tes conseils, tes*
Sacrifices, ton soutien et tes encouragements
***Papa** ma Gratitude ne suffit pas à exprimer ce*
qu'il mérite pour tout ses sacrifices depuis ma
naissance, pendant mon
*À mes frères **Badreddine**, **AbdeRahmen**,*
*À ma Sœur **sana.***
À Tous mes enseignants.
*À tous mes amis et spécialement **Rofida,** qui m'ont*
soutenu dans l'accomplissement de cet
Humble travail
À tous mes professeurs et à tous ceux qui se sont
engagés dans ces modestes travaux
À tout ma famille.
Et À tous qui m'ont aide
de près ou de loin pour la réalisation de ce travail.

Hakima Harzallah

Table des matières

Liste des Abréviations

DL: Deep **L**earning

CNN : **C**onvolutional **N**eural **N**etwork

SSD: Single **S**hot MultiBox **D**etector

YOLO: You **O**nly **L**ook **O**nce

RPN : Réseau de **P**roposition de **R**égion

GAN : Réseaux Adversarial Génératifs

RNN :Réseaux neuronaux récurrents

Liste de Figure

➢ Chapitre1 : deep Learning et détection d'Object

➢ Chapitre 2 : Modèle YOLO

➢ Chapitre 3 : conception

➢ Chapitre 4 : Implémentation et Résultats

Résumé

Ce mémoire présente le développement d'un système innovant visant à améliorer l'accessibilité des personnes aveugles. Le projet repose sur la détection d'objets et leur conversion en messages vocaux grâce à des techniques de Deep Learning. L'algorithme de détection choisi est YOLOv5, reconnu pour sa simplicité et sa rapidité de traitement. Parallèlement, la synthèse vocale (TTS) est utilisée pour traduire les informations détectées en voix. Le système offre une solution prometteuse pour aider les personnes aveugles à identifier et à comprendre leur environnement. En utilisant les avancées technologiques dans les réseaux de neurones convolutifs (CNN) et le traitement du langage naturel, ce système vise à favoriser l'indépendance et l'autonomie des personnes aveugles.

Mots clés : détection d'objets, Deep Learning, YOLOv5, synthèse vocale (TTS), accessibilité, personnes aveugles.

الملخص:

تقدم هذه الأطروحة تطوير نظام مبتكر يهدف إلى تعزيز إمكانية التعرف على الاشياء الخاصة بلمكفوفين. يعتمد المشروع على اكتشاف الأشياء وتحويلها إلى رسائل صوتية باستخدام تقنيات التعلم العميق. خوارزمية الكشف المختارة YOLOv5 هي والمعروفة ببساطتها وسرعة معالجتها. في الوقت نفسه ، يتم استخدام التوليف الصوتي (TTS) ، لترجمة المعلومات المكتشفة إلى صوت. يقدم النظام حلاً واعدًا لمساعدة المكفوفين على التعرف على محيطهم وفهمه. ومعالجة اللغة الطبيعية ، يهدف هذا النظام إلى تعزيز (CNN) باستخدام التقدم التكنولوجي في الشبكات العصبية التلافيفية استقلالية واستقلالية الأشخاص المكفوفين.

الكلمات الرئيسية:اكتشاف الكائن,التعلم العميق,YOLOv5,تركيب تحويل النص إلى كلام (TTs),إمكانية الوصول,ضعاف البصر

Abstract

This thesis presents the development of an innovative system aimed at enhancing accessibility for blind individuals. The project is based on object detection and conversion of detected objects into voice messages using Deep Learning techniques. The chosen object detection algorithm is YOLOv5, renowned for its simplicity and fast processing speed. Additionally, Google text-to-speech (TTS) synthesis is employed to translate the detected information into speech. The system offers a promising solution to assist blind individuals in identifying and understanding their environment. By leveraging advancements in convolutional neural networks (CNNs) and natural language processing, this system aims to promote independence and autonomy for the visually impaired.

Keywords: object detection, Deep Learning, YOLOv5, Google text-to-speech synthesis (TTS),accessibility, visually impaired.

Introduction générale

L'accessibilité et l'autonomie des personnes aveugles ou malvoyantes sont des enjeux majeurs dans notre société. Ces individus font face à de nombreux défis au quotidien, tels que la navigation dans des environnements inconnus, l'identification des objets et la participation active à la vie sociale. Les avancées technologiques récentes offrent des opportunités prometteuses pour répondre à ces défis et améliorer la qualité de vie de ces personnes.

Dans ce contexte, ce mémoire de master présente le développement d'un système intelligent de détection d'objets, conçu spécifiquement pour améliorer l'accessibilité des personnes aveugles. Ce système repose sur l'utilisation de la plateforme Raspberry Pi et des algorithmes de détection d'objets basés sur le Deep Learning. L'objectif principal est de permettre aux utilisateurs de détecter et d'identifier les objets qui les entourent en traduisant les informations visuelles en instructions vocales.

La détection d'objets est un domaine de recherche en pleine expansion, offrant des opportunités passionnantes pour résoudre des problèmes pratiques. En exploitant les réseaux de neurones convolutifs (CNN) et les techniques de traitement d'image, ce mémoire explore les différentes approches et méthodes utilisées pour la détection précise des objets dans des environnements réels.

En outre, une attention particulière est portée à l'aspect vocal du système, en utilisant des techniques de synthèse vocale pour traduire les informations détectées en instructions vocales claires et compréhensibles. L'objectif est de fournir aux utilisateurs aveugles ou malvoyants une expérience utilisateur intuitive et enrichissante, leur permettant de percevoir leur environnement avec une plus grande confiance et autonomie.

Structure du mémoire :

Notre mémoire se structure en quatre (04) chapitres, à savoir :

Le premier chapitre donne une présentation générale sur La vision humaine, le Deep Learning et la détection des objets et ses domaines d'application.

Deuxième Le chapitre est consacré à l'étude des différentes architectures et méthodes de Deep Learning pour la détection des objets, en particulier la méthode YOLOv5.

Ensuite Le chapitre 3 nous présentons la conception de notre système de détection et les digrammes UML adoptés.

Le chapitre 4 nous présentons les résultats expérimentaux obtenus par notre système.

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

chapitre1

Deep Learning et détection d' Object

1.1 Introduction

Aujourd'hui, l'intelligence artificielle est un vaste domaine de recherche qui a connu une évolution rapide ces dernières années, principalement grâce aux avancées en matière de matériel de traitement des données. Parmi les domaines les plus en vue de l'IA, le "Deep Learning" se démarque en offrant des techniques de traitement d'images qui sont largement utilisées à travers le monde. Ces techniques sont appliquées dans des domaines tels que la surveillance par caméra, le suivi d'objets, la détection et la classification d'objets, et bien d'autres encore. Parmi ces applications, la détection d'objets revêt une importance particulière, notamment pour les personnes aveugles, car elle leur permet de découvrir et d'identifier des éléments de leur environnement. Dans ce chapitre, nous explorerons certains concepts clés liés à cette problématique**.**

1.2 La vision humaine :

La vision humaine est la perception humaine d'objets distants en ressentant le rayonnement lumineux de l'objet.

La vision englobe tous les processus physiologiques et cognitive-mentaux par lesquels la lumière émise ou réfléchie par l'environnement détermine les détails des représentations sensorielles, telles que la forme, la couleur, la texture, le mouvement, la distance et le relief. [1]

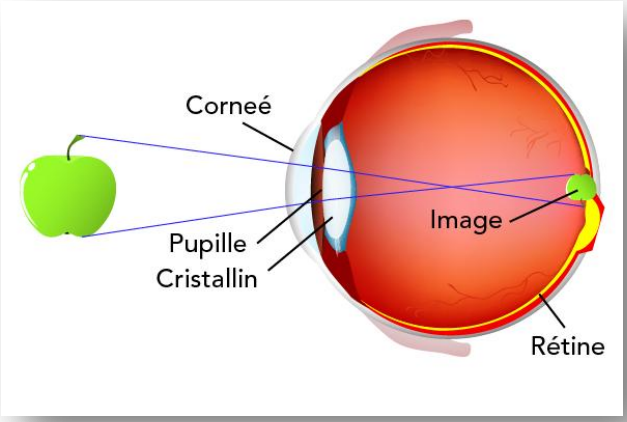


Figure1. 1 : la vision humaine

1.2.1 L'importance de la vision :

La vision joue un rôle très important dans de nombreux domaines :[2]

- L'information : la vision nous apporte la majorité des informations ; elle nous sert à connaitre le monde qui nous entoure (reconnaissance des objets, des visages, interprétation correcte des scènes visuelles, etc.…)

- La communication : la vision est le support primordial à la communication. A la fois émetteur (je regarde) et récepteur (je capte le regard de l'autre), la vision nous permet de décoder les relations humaines, ce qui explique l'importance de son rôle social.

- Les gestes de la vie quotidienne : en effet, la précision de nos gestes, comme se servir à boire par exemple, relève d'un travail de coordination entre l'oeil et la main.

- Les déplacements : l'altération de la vision joue à la fois sur la difficulté à analyser correctement notre environnement, sur la détection des obstacles mais aussi sur notre équilibre.

  Enfin, la vision joue aussi un rôle dans :

- La régulation de la durée et de la qualité de nos phases de vigilance (jour) et de sommeil (nuit).

- influant sur notre humeur et notre état psychologique

1.2.2 Champ visuel :

Le champ de vision est l'espace que l'œil immobile peut voir. Il est mesuré à l'aide d'une coupole Goldmann illustrée ci-dessous. Chez le sujet sain, le champ visuel monoculaire couvre 90° en temporal, 60° en nasal, 70° en bas et 60° en haut . [1]

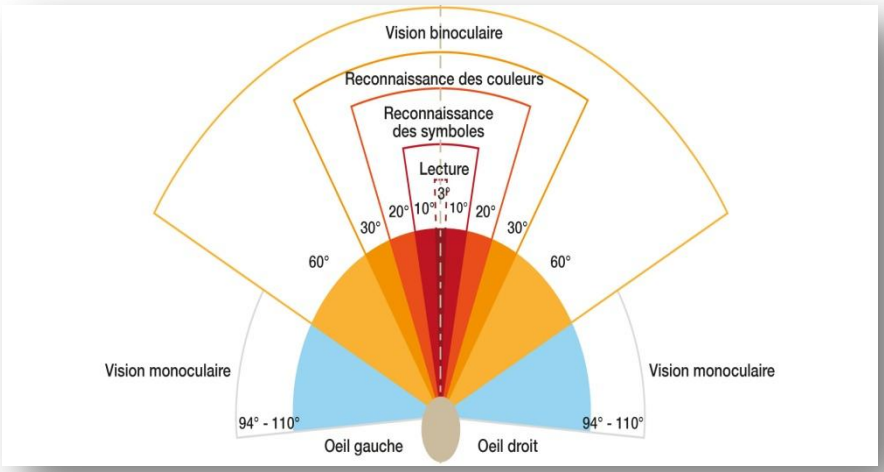


Figure1. 2: Le champ de vision

1.3 Définition malvoyants :

La Fédération européenne des aveugles propose une définition de travail : « Une personne déficiente visuelle est une personne dont la déficience visuelle empêche une ou plusieurs des activités suivantes : lecture et écriture (vision de près), vie quotidienne (vision intermédiaire), communication (vision de près et vision de loin intermédiaire), compréhension de l'espace et du mouvement (vision de loin), s'engager dans des activités qui nécessitent une attention visuelle prolongée"

Une personne est considérée comme malvoyante si son acuité visuelle après correction est comprise entre 4/10e et 1/20e, OU si son champ visuel est compris entre 10 et 20 degrés. [3]

Figure1. 3: Définition malvoyants

1.3.1 Souffrance des malvoyants :

Les malvoyants ont des difficultés à se déplacer en ville pour les transports en commun et repérage préalable du parcours. En plus, ils ont des difficultés au quotidien : la lecture des textes, reconnaissance d'objets…etc. La chose la plus importante est les interactions sociales car ils ne peuvent pas reconnaître les personnes ce qui leurs faisant se sentir inférieurs et marginalisés. [2]

1.4 Qu'est-ce que l'apprentissage ?

apprentissage est un concept vaste qui fait partie intégrante de intelligence artificielle (IA). Il permet aux machines de reconnaître des objets en se basant sur leurs expériences de détection antérieures. objectif principal de apprentissage est de comprendre la structure des données et de les représenter une manière compréhensible et utilisable. En général, il existe deux types apprentissage : supervisé et non supervisé. [4]

1.5 Deep Learning :

1.5.1 Concept :

Le Deep Learning , également connu sous le nom apprentissage profond, est une des principales technologies du Machine Learning. Il utilise des réseaux de neurones pour capturer des modèles complexes. Ces réseaux sont inspirés de la structure et du fonctionnement du cerveau humain. Cette technologie permet aux systèmes intelligence artificielle accomplir des tâches humaines telles que la reconnaissance visuelle objets réels ou la compréhension de la parole. Le Deep Learning fait partie une classe algorithmes de Machine Learning. [4]

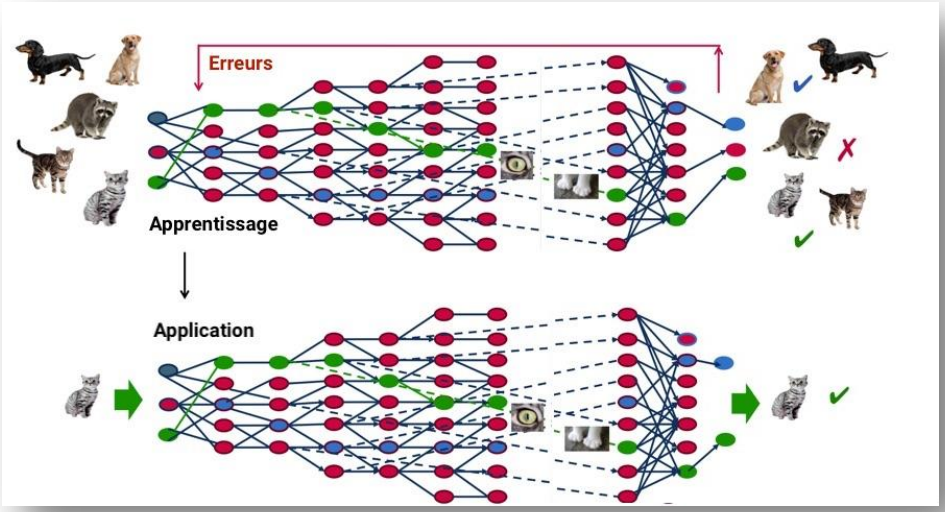


Figure1. 4: le concept apprentissage profond [4]

1.5.2 Les Architecture du Deep Learning :

Les trois grandes architectures de réseaux profonds selon (Patterson & Gibson, 2017) : [5]

1.5.2.1 Réseaux Adversarial Génératifs (GAN) : [6]

Les réseaux antagonistes génératifs , ou Generative Adversarial Networks (GAN), sont une classe algorithmes apprentissage profond utilisés pour générer de nouvelles données à partir un ensemble de données existant. Ils se composent de deux réseaux distincts : un réseau génératif et un réseau discriminatif.

Le réseau génératif a pour objectif de produire de nouveaux échantillons de données, tandis que le réseau discriminatif évalue les données générées et détermine si elles sont réalistes.

Les GAN fonctionnent en entraînant simultanément ces deux réseaux. Le réseau génératif prend une entrée et essaie de générer une sortie réaliste, tandis que le réseau discriminatif compare les données générées avec les données réelles pour les distinguer.

L' objectif des GAN est de parvenir à un point où le réseau génératif produit des données si réalistes que le réseau discriminatif ne peut plus les différencier des

données réelles. Cependant, application des GAN présente le défi de trouver équilibre entre les modèles génératif et discriminatif afin obtenir des résultats optimaux.

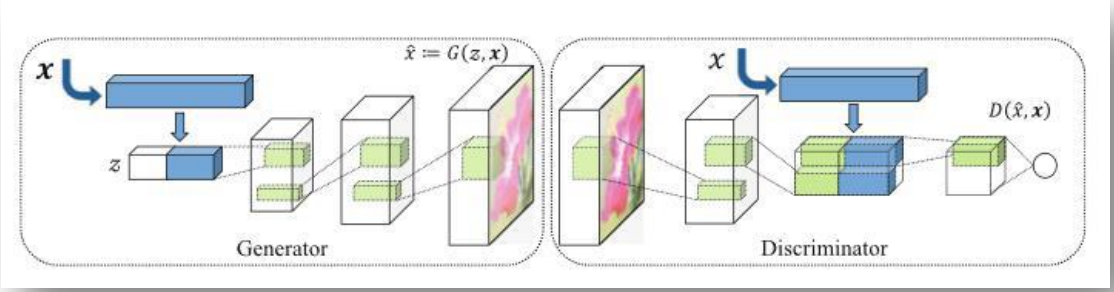


Figure1. 5: Les réseaux adversaires génératifs

1.5.2.2 Réseaux Neuronaux Convolutifs (CNN) :

Les réseaux neuronaux convolutifs (CNN), également appelés ConvNets , sont un type de réseaux neuronaux à anticipation bien adaptés aux tâches liées au domaine de la vision par ordinateur, notamment à la reconnaissance d'objets [7]

Les réseaux neuronaux convolutifs fonctionnent en ingérant et en traitant de grandes quantités de données dans un format de grille, puis en extrayant des caractéristiques granulaires importantes pour la classification et la détection. Les CNN se composent généralement de trois types de couches : une couche convolutive, une couche de mise en commun et une couche entièrement connectée. Chaque couche a un objectif différent, exécute une tâche sur les données ingérées et apprend des quantités croissantes de complexité. [8]

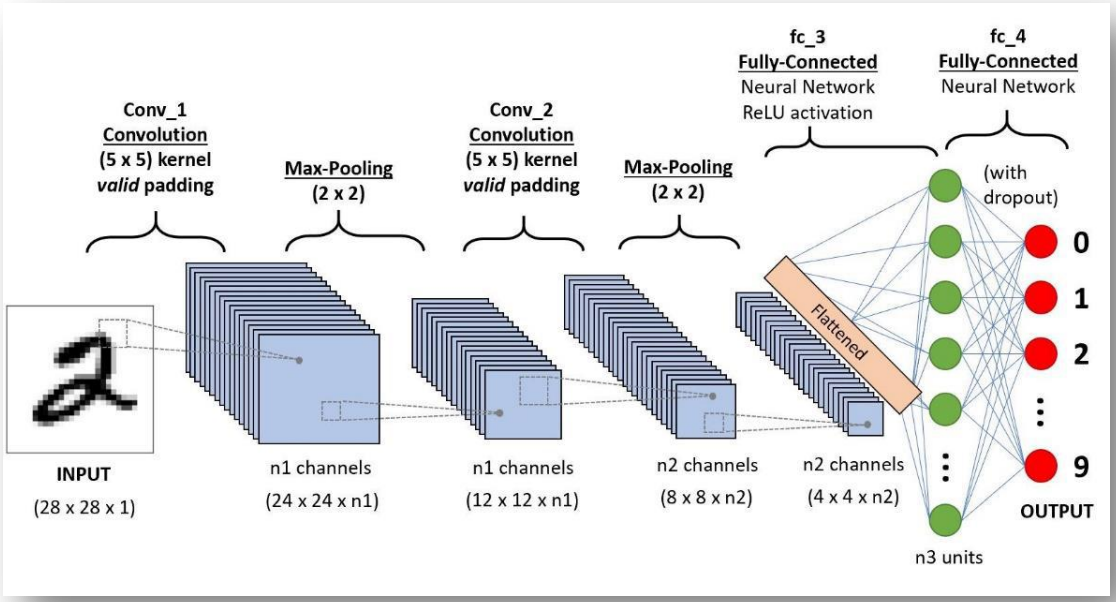


Figure1. 6: Les réseaux neuronaux convolutifs [9]

a) La couche Convolution :

Une couche convolutive est parfois appelée couche d'extraction de caractéristiques car c'est au niveau de cette couche que les caractéristiques de l'image sont extraites.

Tout d'abord, une partie de l'image est connectée à la couche Convolution pour l'opération de convolution, et le résultat de l'opération de produit scalaire du champ récepteur (qui est de la même taille que le filtre dans l'image d'entrée) et le filtre est l'entier unique de la sortie. Nous faisons ensuite glisser le filtre vers le champ récepteur suivant de la même image d'entrée et refaites la même chose .Cette opération est répétée encore et encore à travers le même processus jusqu'à ce que l'image entière soit numérisée. [10]

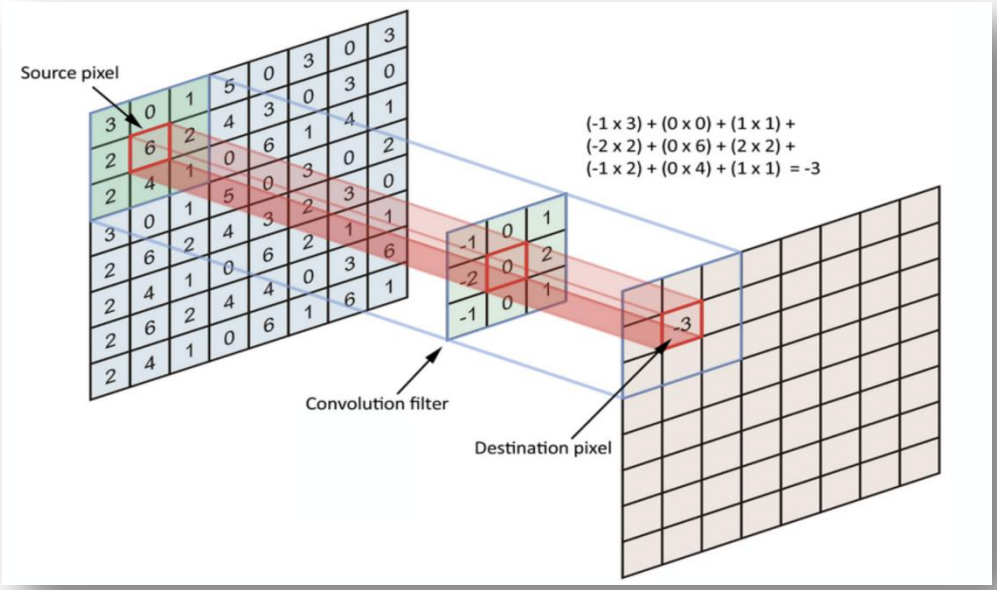


Figure1. 7: **Une couche convolutive [11]**

b) Couche de mise en commun (Pooling) :

Une couche de regroupement (POOL) est une opération de sous-échantillonnage, généralement appliquée entre deux couches convolutives.

Son rôle est de réduire progressivement la taille de la feature map (matrice de convolution) pour réduire les paramètres et les calculs du réseau, tout en conservant les informations importantes [10]

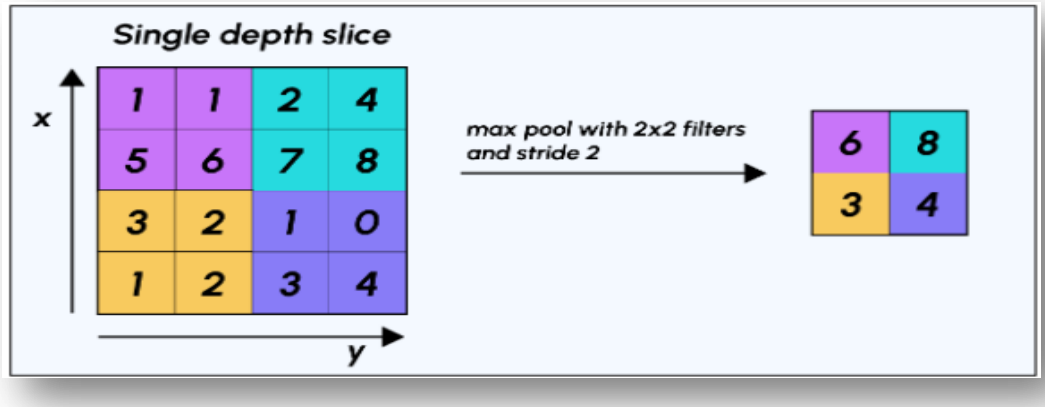


Figure1. 8: Une couche de Pooling [12]

c) Couche entièrement connecté (Fully Connected) :

Une couche entièrement connectée dans un réseau neuronal est une couche où toutes les entrées d'une couche sont connectées à chaque fonction d'activation de la couche suivante .Dans un modèle d'apprentissage automatique, la dernière couche est une couche entièrement connectée qui utilise les données extraites par la couche précédente pour former la sortie finale [13]

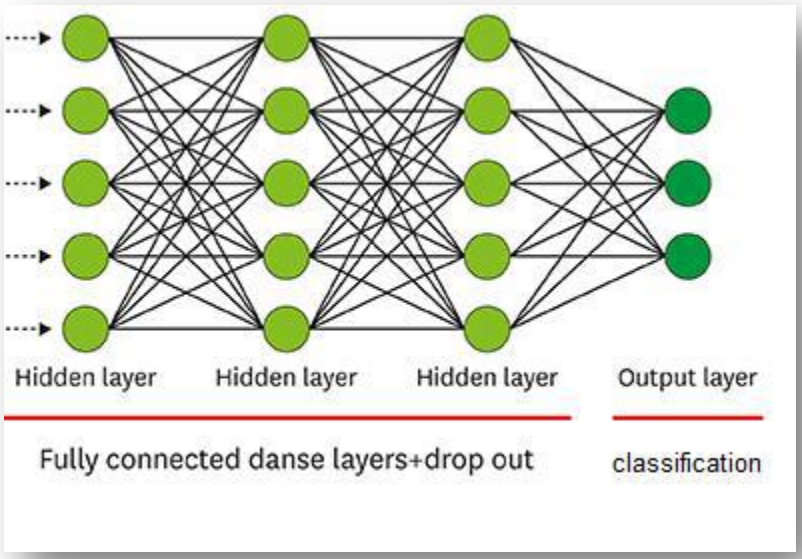


Figure1. 9: Une couche entièrement connectée [14]

1.5.2.3 Réseaux neuronaux récurrents (RNN) :

Un réseau neuronal récurrent (RNN) est un réseau dans lequel les connexions entre les unités forment des boucles dirigées. En fait, il peut y avoir des connexions plus complexes entre les unités cachées .Grâce à ces connexions récurrentes, les RNN peuvent traiter des données séquentielles, telles que des vidéos et des phrases vocales. [10]

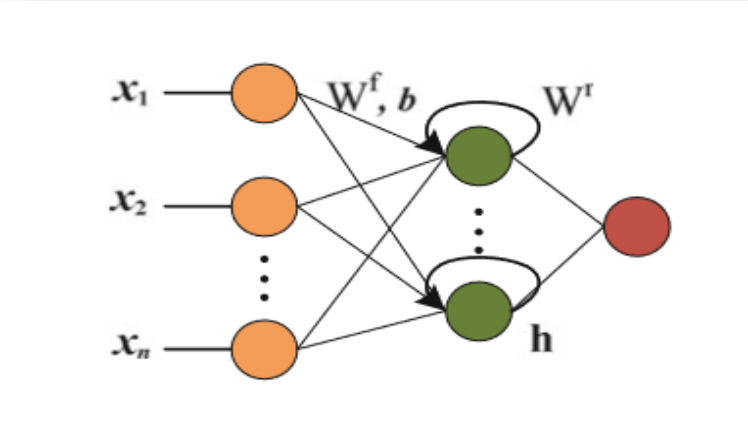


Figure1. 10: Un réseau neuronal récurrent [10]

En général, considéré comme les réseaux de neurones supervisés comme appartenant au domaine machine Learning plutôt que Deep Learning.

1.5.3 L'avantage et d'inconvénient du Deep Learning :

Deep Learning présente plusieurs avantages et inconvénients : [15]

- ✓ Avantage :
  - La système peut être plus performante que les méthodes de machine Learning traditionnelles.
  - Le prix est bas par rapport aux autres méthodes, car tous les détails n'ont pas besoin d'être correctement planifiés.
  - Le système est rapide et peut gérer des données volumineuses
- ✓ Inconvénient :
  - Il nécessite beaucoup de données pour s'entraîner, ce qui peut être un défi pour les utilisateurs disposant de petites bases de données.
  - Les styles d'apprentissage en profondeur sont complexes et peuvent être difficiles à comprendre ou à reproduire s'ils ne sont pas correctement formés.

1.1.1 Application du Deep Learning :

Le Deep Learning est utilisé dans de nombreux domaines comme : [4]

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

1.6 Définition d'objet :

Dans le domaine de la vision par ordinateur, le concept d'objet constitue un élément clé, puisque nos recherches portent toujours sur le concept d'objet. Un objet désigne une zone d'une image qui se caractérise par sa texture, sa forme, sa couleur, sa direction de dégradé, voire son mouvement, et qu'il représente. En d'autres termes, nous pouvons dire que les objets d'une séquence d'images représentent des pixels qui appartiennent au premier plan ou à l'arrière-plan de la scène. Il existe deux types d'objets : les objets fixes et les objets mobiles. [16]

1.7 Détection des objets :

La détection d'objets est une technique de vision par ordinateur qui permet d'identifier et de localiser des objets dans des l'image numérique. Il détecte les caractéristiques des objets et ignore tout le reste.la détection d'objet dessine une boîte qui se connecte autour des objets détectés, nous permettant de trouver ces objets dans une situation donnée. [16]

Figure1. 11: La détection d'objets

1.7.1 Pour quoi la détection d'objet ?:

La détection d'objets est un domaine de recherche vaste et très important, car la recherche actuelle vise à créer des méthodes qui se rapprochent des compétences des humains dans la compréhension et l'observation d'objets et la reconnaissance .La détection d'objet elle est parmi les application pratique les plus intéressantes dans la vie courante, il est utilisé dans la surveillance du trafic, la surveillance Object pour les aveugles, cette dernière est utilisé dans les entreprise pour détecter les produits bien fait et les mal fait [14]

1.7.2 principe de détection d'objet :

Le principe de la détection d'objets est le suivant : pour une image donnée, on recherche les régions de celle-ci qui pourraient contenir un objet puis pour chacune de ces régions découvertes, on l'extrait et on la classe à l'aide d'un modèle de classification d'image. Les régions de l'image d'origine ayant de bons résultats de classification sont conservés et les autres jetés. Ainsi, pour avoir une bonne méthode de détection d'objets, il est nécessaire d'avoir un algorithme solide de détection des régions et un bon algorithme de classification d'images [16]

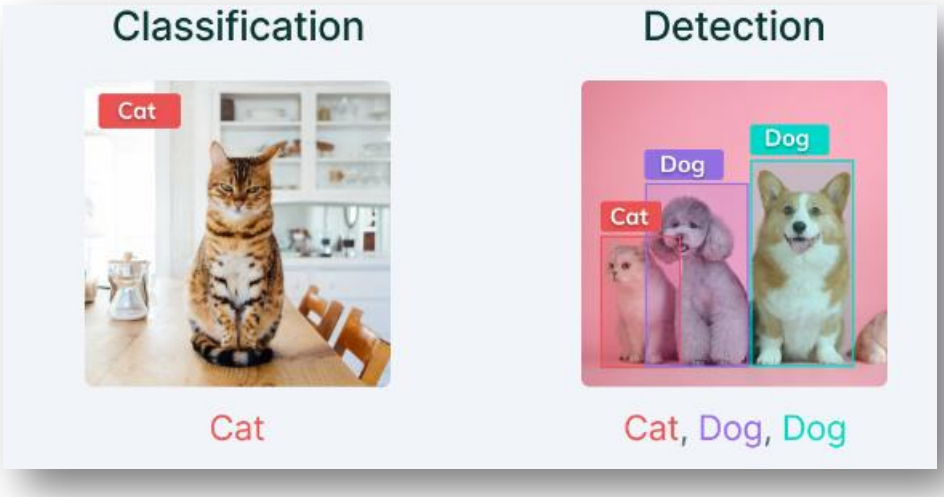
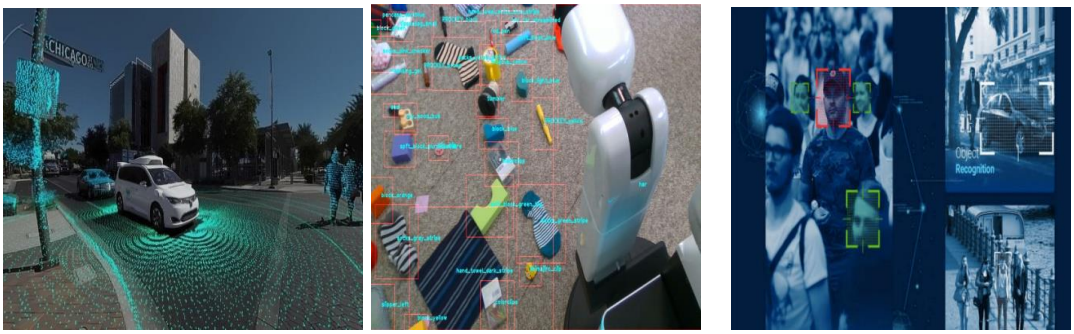


Figure1. 12: Le principe de la détection d'objets

1.7.3 Domaine d'application :

Récemment, de nombreuses études ont montré l'efficacité de la détection d'objets et de ses travaux qui sont très importants dans différents domaines et nombreux, y compris : [14]

- Robotique
- L'analyse d'images médicales
- La vidéosurveillance
- Militaire
- La voiture autonome
- Et la détection de la somnolence des conducteurs sur l'autoroute afin d'éviter les accidents peut être obtenue par la détection d'objets aussi.



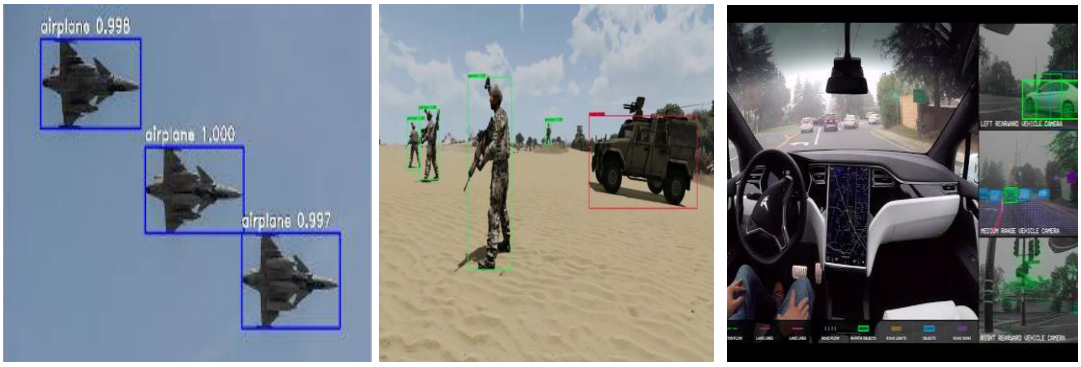


Figure1. 13: Domaine d'application

1.8 Etat de l'Art sur la détection d'objets basés sur le Deep Learning:

actuellement, les cardères détection d'Object basés sur l'apprentissage approfondi peuvent être principalement divisés en deux type : [14]

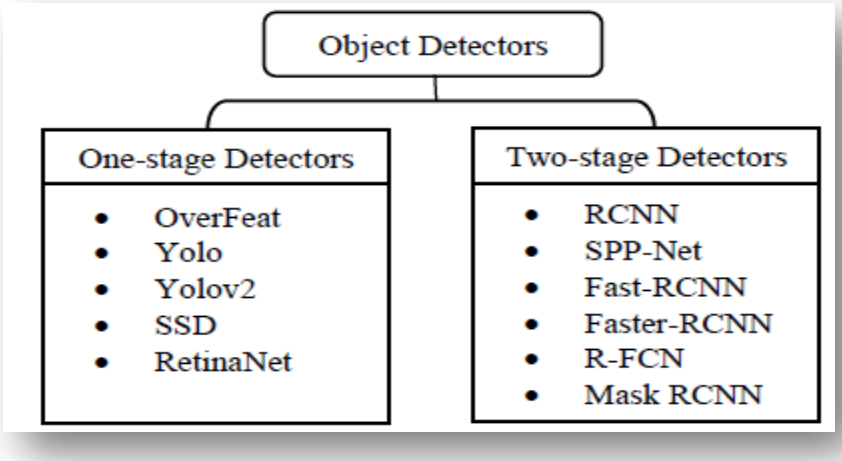


Figure1. 14: le type de détection d'Object basés sur l'apprentissage approfondi [14]

❖ Détecteurs à deux étages :

Les détecteurs à deux étages ont une plus grande précision et de meilleures performances, et rapportent de meilleurs résultats idéaux dans la détection d'objets que les détecteurs à un étage, mais ils sont généralement plus lents que les détecteurs à un étage car ils ont deux étages : le premier étage utilise un réseau de proposition de région, et le classement final et la régression dans la deuxième étape [14]

❖ Détecteurs à un étage

Les détecteurs à un étage sont plus rapides et mieux adaptés aux applications de détection d'objets en temps réel, mais fonctionnent relativement mal par rapport aux détecteurs à deux étages . Parce que les détecteurs à un étage n'ont pas d'étape de génération de proposition distincte (ou génération de proposition d'apprentissage). Dans ce cas, les étapes de pré-détection et de classification des régions d'intérêt sont combinées en une seule étape de détection gérée par un seul réseau de neurones. [14]

1.9 Conclusion :

En conclusion, l'intégration de l'apprentissage profond a apporté une révolution dans le domaine de la détection d'objets en vision par ordinateur. Les détecteurs à deux étapes se distinguent par leur précision supérieure, bien qu'ils puissent être plus lents, tandis que les détecteurs à une étape offrent une vitesse accrue au détriment de performances légèrement inférieures. Ces avancées ouvrent de nouvelles perspectives dans des domaines tels que la surveillance et la reconnaissance d'objets. Notre choix est tombé sur les détecteurs à une étape vu les exigences de notre application. Dans le chapitre suivant, nous expliquons nos choix et notre conception.

chapitre2

Modèle Yolo

2.1 Introduction :

Ce chapitre se concentre sur l'algorithme révolutionnaire YOLO (You Only Look Once), qui a réinventé la détection d'objets en vision par ordinateur. Nous plongerons dans les détails de son architecture et de son fonctionnement, ainsi que dans les avancées significatives qu'il a apportées en termes de vitesse et de précision. L'algorithme YOLO a ouvert de nouvelles perspectives dans la détection d'objets en temps réel, et son impact est visible dans de nombreux domaines, de la surveillance à la conduite autonome. Nous explorerons également les défis et les améliorations récentes de YOLO.

2.2 Algorithme de Modèle de détection :

Au cours des dernières années, la détection d'objets basée sur l'apprentissage en profondeur a connu des avancées significatives en termes de robustesse par rapport aux méthodes traditionnelles. Dans cette section, nous aborderons les méthodes de détection les plus couramment utilisées, à savoir Faster R-CNN, YOLO (You Only Look Once) et SSD (Single Shot MultiBox Detector). Nous examinerons les principes fondamentaux de chaque méthode, leurs performances et leurs domaines d'application privilégiés. Cette analyse comparative nous permettra de mieux comprendre les forces et les limites de chaque approche, et d'identifier les facteurs clés à considérer lors du choix d'une méthode de détection d'objets en fonction des besoins spécifiques d'une tâche donnée.

2.2.1 Faster R-CNNs :

Faster R-CNN est l'une des méthodes de détection d'objets les plus populaires et performantes. Elle fait partie de la série R-CNN développée en 2014 par Ross Girshick et son équipe. Améliorée avec Fast R-CNN, elle a ensuite abouti à Faster R-CNN. [17]

Le processus de Faster R-CNN commence par passer l'image en entrée à travers un réseau neuronal convolutif (CNN) pour obtenir une carte de caractéristiques des objets présents. Cette carte est ensuite utilisée par un réseau de proposition de région (RPN) pour générer des propositions de régions. Le RPN utilise une méthode d'apprentissage en profondeur pour prédire les régions les plus susceptibles de contenir des objets d'intérêt. [17]

Les caractéristiques extraites par le CNN et les cadres de délimitation des objets pertinents sont ensuite utilisés pour générer une nouvelle carte de caractéristiques en effectuant une mise en commun des régions d'intérêt (RoI). Les régions regroupées passent ensuite par des couches entièrement connectées pour prédire les coordonnées des zones des objets et leurs classes. [17]

Les résultats obtenus par Faster R-CNN dépassent considérablement ceux des R-CNN et Fast R-CNN originaux, comme le montre la figure ci-dessous, illustrant les gains significatifs de performance. Cette méthode est largement utilisée dans de nombreux domaines pour sa précision et sa robustesse dans la détection d'objets.

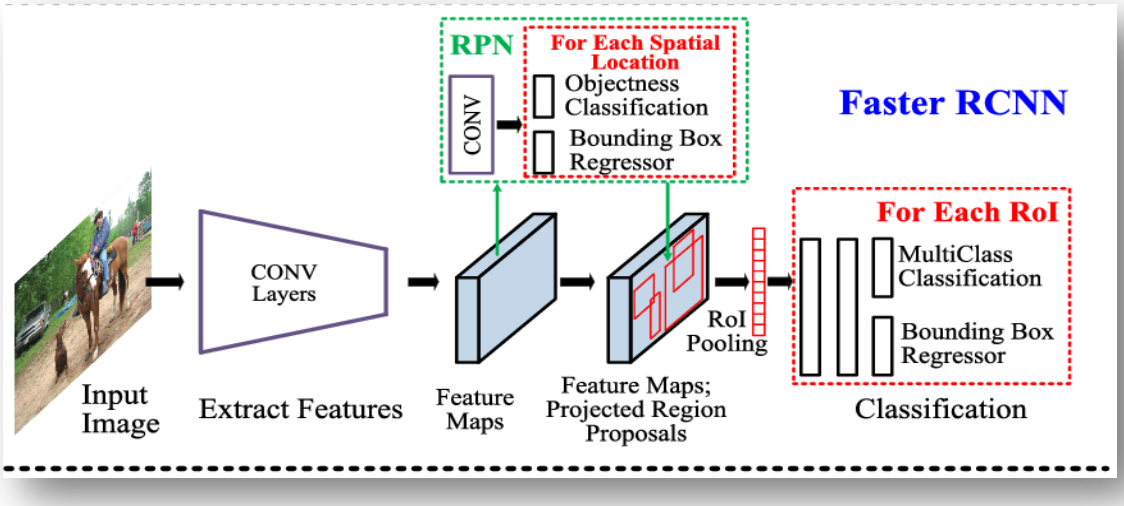


Figure2. 1 : le méthode de détection Faster R-CNN [14]

2.2.2 Single Shot Detectors(SSD) :

Le SSD en anglais Single Shot MultiBox Detector présenté par Liu et al, SSD est un modèle de détection d'objets, le SSD est basé sur l'utilisation de réseaux convolutif qui produisent plusieurs boîtes englobant de différentes tailles fixes et évaluent la présence de l'instance de classe d'objets dans ces boîtes, suivies d'une étape de suppression non maximale pour produire le détections finales. Le modèle SSD fonctionne comme suit : chaque image d'entrée est divisée en grilles de différentes tailles et à chaque grille, la détection est effectuée pour différentes classes et différents rapports d'aspect. [18]

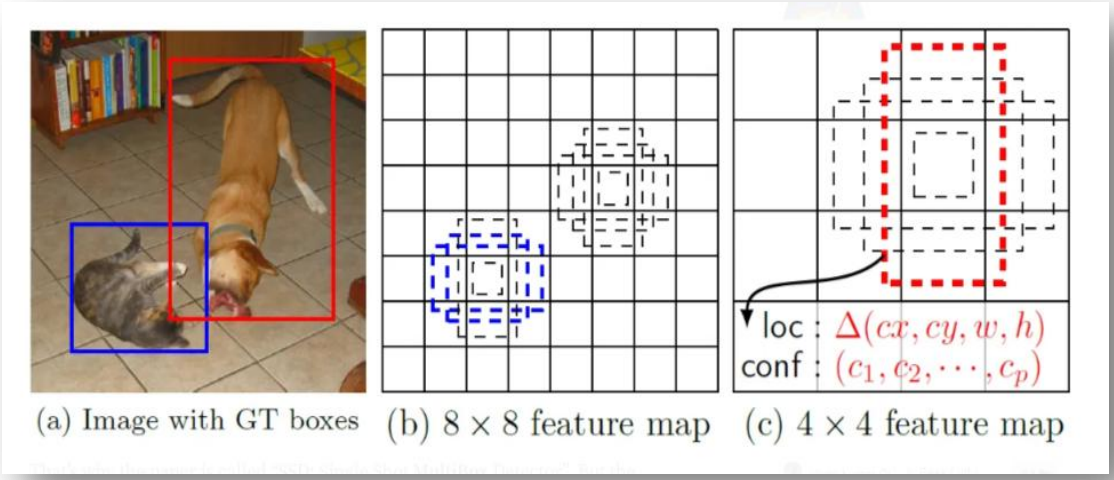


Figure2. 2: la méthode de détection Single Shot Detectors (SSD) [18]

2.2.3 You Only Look Once(YOLO) :

Proposé par J. Redmon et al , c'est l'un des algorithmes de détection et de classification d'objets en temps réel les plus puissants. YOLO a une architecture très simple qui le rend très rapide. Il est ainsi nommé car contrairement aux algorithmes de détection et de classification d'objets mentionnés précédemment, qui examinent séquentiellement plusieurs régions d'une image pour trouver les objets présents, puis font plusieurs prédictions pour chaque région, YOLO change cela en raisonnant un peu au niveau de l'image globale. Au lieu d'utiliser une approche en deux étapes pour la classification et la localisation des objets, YOLO applique simultanément un seul CNN pour la classification et la localisation des objets. [19]

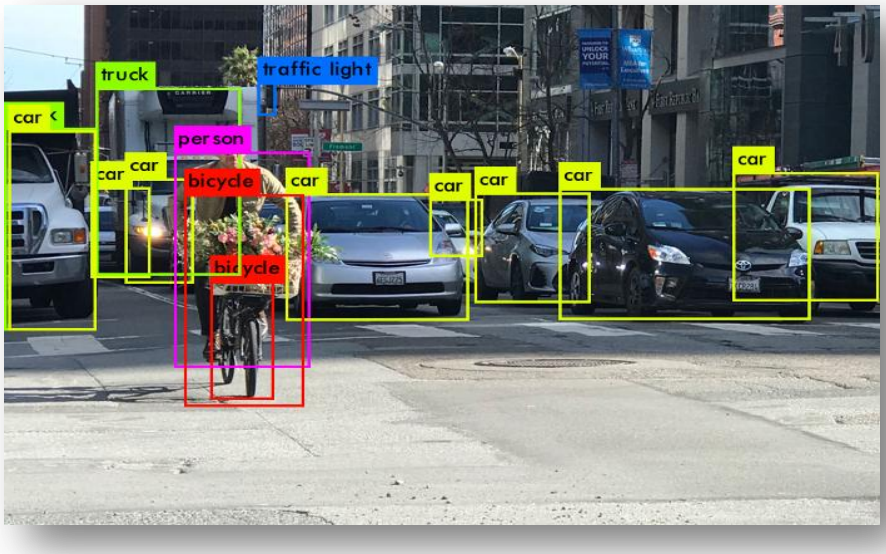


Figure2. 3: la méthode de détection You Only Look Once(YOLO)

a) **Intersection sur Union (IoU) :**

L'intersection sur l'union (IoU) est la métrique d'évaluation de facto utilisée dans la détection d'objets. Il est utilisé pour déterminer les vrais positifs et les faux positifs dans un ensemble de prédictions. Lors de l'utilisation de l'IoU comme métrique d'évaluation, un seuil de précision doit être choisi . Il compare la boîte prédite à la boîte détectée, et la zone peut être calculée comme suit : [10]

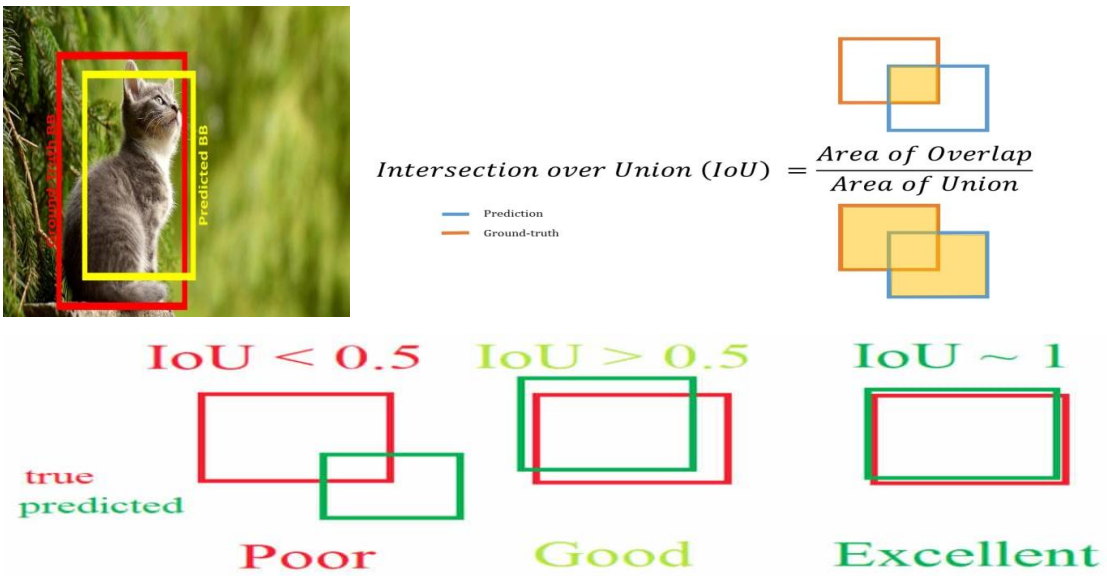


Figure2. 4: Intersection sur Union (IoU)

b) **Boîte d'ancrage (Anchor Box) :**

Cellules de grille l'idée de diviser les images en cellules de grille est unique dans YOLO en définissant les grilles comme S x S. Si le centre de l'objet est à quelle grille de cellules la grille de cellules prédira l'objet. [20]

Les boîtes d'ancrage sont l'algorithme de YOLO qui sépare les objets si plusieurs centres d'image se trouvent dans la même cellule de grille. . Elles permettent de séparer et de localiser précisément les objets en leur assignant des rectangles prédéfinis de tailles et de formes spécifiques. Cela garantit que chaque objet est correctement identifié et localisé, même lorsque les objets se chevauchent ou sont proches les uns des autres. Les boîtes d'ancrage sont un élément clé de l'algorithme YOLO pour une détection précise des objets dans les images. [20]

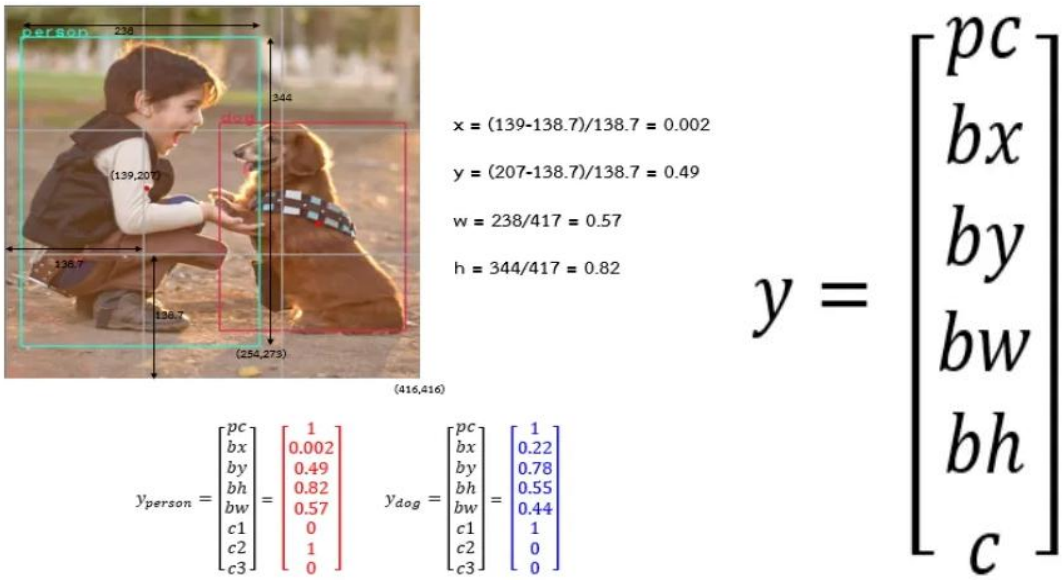


Figure2. 5: Boîte d'ancrage (Anchor Box)

Donc en générale la formule si on divise l'image en une grille S × S et, pour chaque cellule de la grille, il prédit B boîtes de délimitation, la confiance pour ces boîtes et les probabilités de classe C. Ces prédictions sont encodées sous la forme d'un tenseur $S \times S \times (B * 5 + C)$. [20]

c) **Suppression non maximale :**

Cette étape est la dernière étape de l'algorithme de détection, elle est utilisée pour détecter le même objet dans l'image à travers plusieurs boîtes englobant , cette technique est utilisée pour "supprimer" les boîtes englobant improbables au lieu de ne garder que les meilleures .Ensuite le processus de cette technique se déroule en 5 étapes : [21]

Étape 1 : Sélectionnez la case avec le score d'objectivité le plus élevé

Étape 2 : Ensuite, comparez le chevauchement (intersection sur union) de cette boîte avec d'autres boîtes

Étape 3 : Supprimer les cadres de délimitation avec chevauchement (intersection sur union) > 50 %

Étape 4 : Passez ensuite au score d'objectivité le plus élevé suivant

Étape 5 : Enfin, répétez les étapes 2 à 4

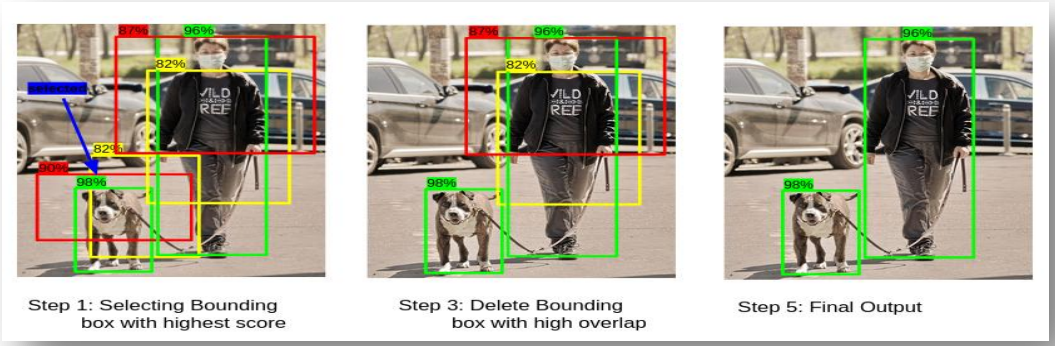


Figure2. 6: Suppression non maximale [21]

2.2.3.1 Avantages et inconvénients de YOLO : [22]

✓ Avantages :

L'avantage de l'algorithme YOLO est qu'il est très rapide car au moment de l'exécution, l'image n'a été exécutée qu'une seule fois sur CNN, ce qui rend YOLO beaucoup plus rapide que Faster R-CNN et peut être s'exécuter en temps réel.

✓ Inconvénients :

Yolo à certain limitation

- Rappel relativement faible par rapport à Faster R_CNN .
- Erreur de localisation plus importante
- Difficulté à détecter les objets proches en raison des limites des boîtes englobantes par grille.

- Difficulté à détecter les petits objets en raison de la résolution spatiale limitée

2.3 Le model yolov5 :

YOLOv5 est l'une des dernières versions de la famille YOLO et propose plusieurs avancées en matière de détection d'objets. Son modèle regroupe une série de modèles à différentes échelles, formés sur l'ensemble de données. YOLOv5 intègre l'assemblage de modèles, qui combine plusieurs modèles lors du processus de prédiction. Il utilise également l'augmentation du temps de test, qui applique des modifications aléatoires aux images de test telles que des retournements et des rotations. De plus, YOLOv5 intègre l'évolution des hyperparamètres, qui optimise les hyperparamètres à l'aide d'un algorithme génétique pour améliorer les performances. [23]

2.3.1 Architecteur yolov5 :

Généralement architecture de Yolo se compose de trois pièces principales : [24]

- **Colonne vertébrale (Backbone)** : Un réseau de neurones convolutifs qui agrège et forme des caractéristiques d'image à différentes granularités.
- **Cou (Neck)** : Une série de couches pour mélanger et combiner les caractéristiques de l'image pour les transmettre à la prédiction.
- **Tête (Head)** : Consomme les caractéristiques du cou et effectue des étapes de prédiction de boîte et de classe .

L'architecture de modèle YOLOv5 se composé comme le montre la figure :

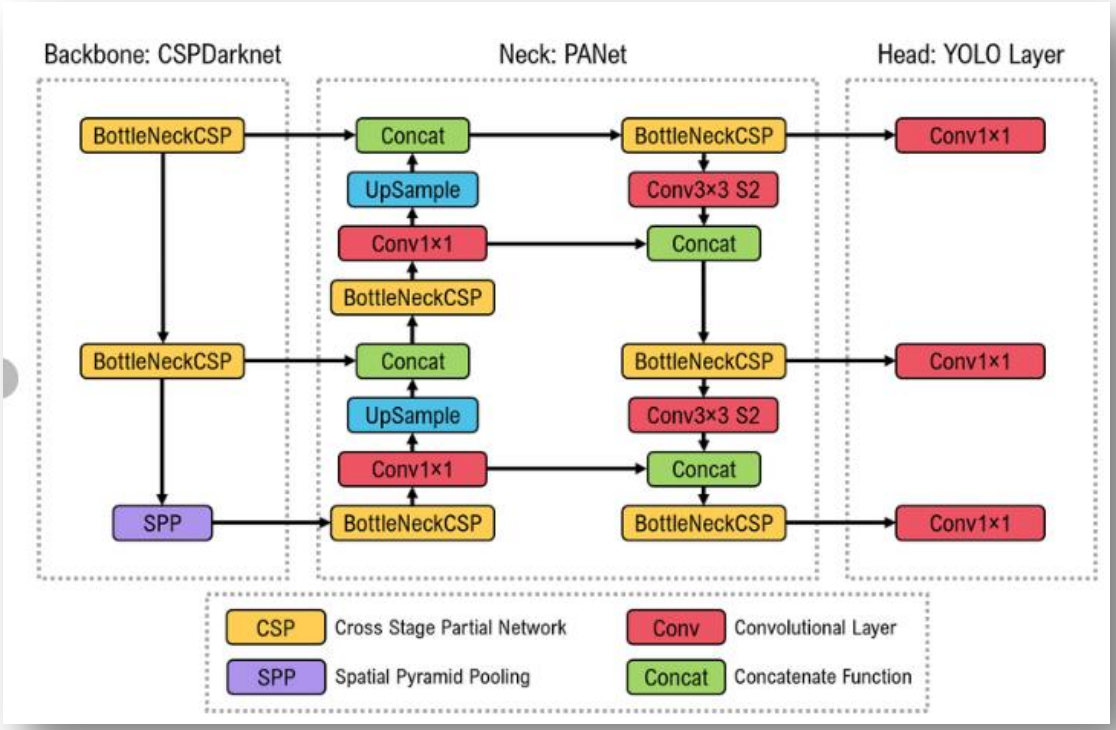


Figure2. 7: L'architecture de modèle YOLOv5 [25]

2.3.1.1 Colonne vertébrale (Backbone) :

a. Cross Stage Partial Network :

Le modèle CSP (Cross Stage Partial Darkent) est dérivé de l'architecture DenseNet, qui prend l'entrée précédente et la concatène avec l'entrée actuelle avant d'entrer dans une couche dense. [24]

DenseNet est conçu pour connecter des couches dans des réseaux de neurones très profonds dans le but d'atténuer le problème du gradient de fuite. [26]

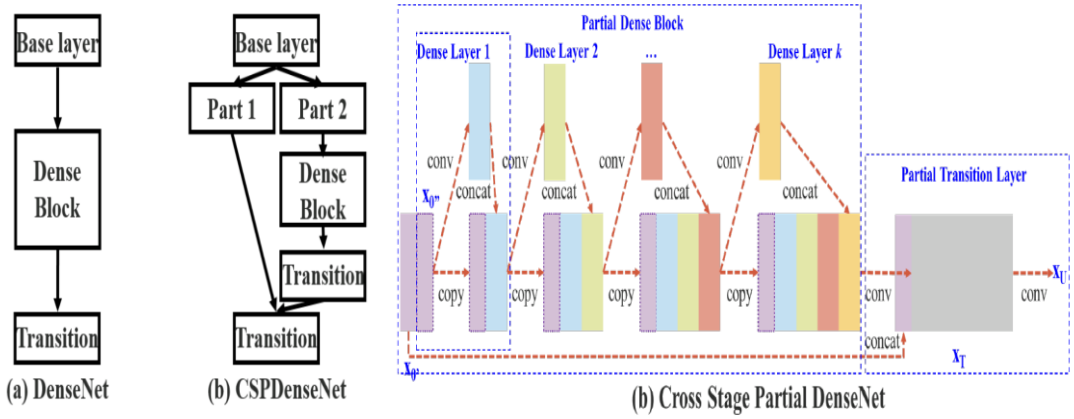


Figure2. 8: Cross Stage Partial Network [26]

DenseNet : Chaque étage d'un DenseNet contient un bloc dense et une couche de transition, et chaque bloc dense est composé de k couches denses. La sortie de la ith couche dense sera concaténée avec l'entrée de la ith couche dense, et le résultat concaténé deviendra l'entrée de la (i + 1) th couche dense. Les équations montrant le mécanisme mentionné ci-dessus peuvent être exprimées : [26]

$$x1 = w1 * x0$$

$$x2 = w2 * [x0, x1]$$

$$xk = wk * [x0, x1, \ldots\ldots, xk-1]$$

où $*$ représente l'opérateur de convolution, et [x0, x1, ...] signifie concaténer x0, x1, ..., et wi et xi sont respectivement les poids et la sortie de la i th couche dense. Si on utilise une backpropagation pour mettre à jour les poids, les équations de mise à jour des poids peuvent s'écrire : [26]

$$w1 = (w1, g0)$$

$$w2 = (w2, g0, g1) \quad (3.2)$$

$$wk = (wk, g0, g1\ldots\ldots, gk-1)$$

Où f est la fonction de mise à jour du poids, et gi représente le gradient propagé à la ith couche dense. Nous pouvons constater qu'une grande quantité d'informations de gradient est réutilisée pour mettre à jour les poids de différentes couches denses. Il en résultera que les différentes couches denses apprendront à plusieurs reprises les informations de gradient copiées. [26]

CSP-DenseNet, dans lequel chaque étape comprend un module Dense local et une couche de transition locale. Dans le module Dense local, la carte d'entités de la couche de base est divisée en deux parties via le canal x=[x ',x"] ,où x'' est directement connecté à la couche Transition à la fin de la scène, x' est connectée à la couche Transition à travers tout le module Dense. La sortie de la couche dense [x",x1,. . . .xk] passera par une couche de transition, et la sortie , xt , sera épissée avec x'', xu est ensuite sortie à travers une autre couche de transition. Les paramètres de l'équation de propagation vers l'avant et CSP-DenseNet sont mis à jour comme suit : [27]

$$x_K = w_K * [x_{0''}, x_1, \ldots, x_{k-1}]$$

$$x_T = w_T * [x_{0'}, x_1, \ldots, x_k]$$

$$x_U = w_U * [x_{0''}, x_T]$$

$$w_k = (w_k, g_0, g_1 \ldots\ldots, g_{k-1})$$

$$w_T = (w_T, g_0, g_1 \ldots\ldots, g_k)$$

$$w_U = (w_U, g_0, g_T)$$

CSP maintient les fonctionnalités par propagation, encourage le réseau à réutiliser les fonctionnalités et réduit le nombre de paramètres réseau, aidant à préserver les fonctionnalités à granularité fine pour les transférer plus efficacement vers des couches plus profondes .Considérant que l'ajout d'un trop grand nombre de couches convolutives densément connectées entraînera une diminution de la vitesse de détection, seul le dernier bloc convolutifs capable d'extraire les caractéristiques sémantiques les plus riches du réseau fédérateur Darkent-53 est amélioré en tant que bloc dense. [26]

b. CSPDarknet :

CSPDarknet53 est un réseau neuronal convolutif et une colonne vertébrale pour la détection d'objets qui utilise Darkent-53. Il utilise une stratégie CSPN et pour diviser la carte des caractéristiques de la couche de base en deux parties, puis les fusionne via une hiérarchie à plusieurs étapes. L'utilisation d'une stratégie de division et de fusion permet un flux plus dégradé à travers le réseau. [28]

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure2. 9: CSPDarknet [10]

c. Spatial Pyramid Pooling :

Le rôle de la structure SPP [31] dans le réseau YOLOv5 est d'implémenter un vecteur de caractéristiques de taille fixe en tant que sortie de couche entièrement connectée pour les images avec des entrées de tailles différentes. La structure SPP utilise trois noyaux de convolution de différentes tailles, 3, 5 et 9, pour extraire les caractéristiques via l'opération de mise en commun maximale, améliorer la capacité d'expression des caractéristiques du graphe de caractéristiques et améliorer le champ de réception du réseau. La structure SPP est illustrée à la figure 7a, qui effectue d'abord 1 × 1, 3 × 3, 5 × 5 et 9 × 9 opérations de regroupement maximum sur les données transférées à partir de la fonction d'activation de normalisation convolutives (Convolution + Banch Normalization + SiLU, CBS) en parallèle puis les connecte à la structure CBS par concaténation splicing. La structure CBS réalise la fusion des caractéristiques et termine l'opération d'extraction des caractéristiques. [29]

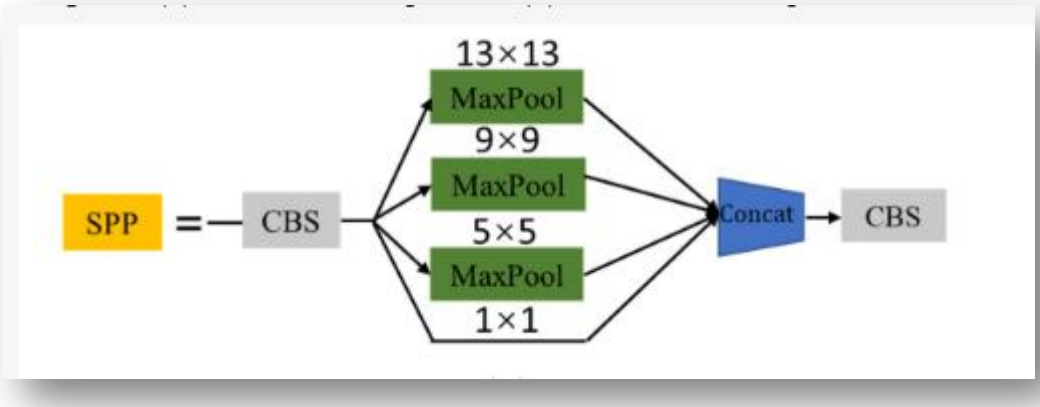


Figure2. 10: Spatial Pyramid Pooling [29]

2.3.1.2 Cou (Neck) :

Yolov5 utilise PANet comme Cou(Neck) pour agréger les fonctionnalités et il est basé sur Framework FPN, tout en améliorant la diffusion de l'information.

L'architecture FPN a mis en œuvre un chemin descendant pour transférer les caractéristiques sémantiques (de la couche de haut niveau), puis les concaténer en caractéristiques à grain fin (de la couche de bas niveau dans la dorsale) pour prédire les petits objets dans le détecteur à grande échelle . [24]

2.3.1.3 Tête (Head) :

La tête est principalement utilisée pour la partie détection finale. Il applique des boîtes d'ancrage sur des cartes d'entités et produit un vecteur de sortie final avec des probabilités de classe, des scores d'objets et des boîtes englobâtes. [24]

2.3.2 Avantages et Inconvénients de Yolo v5 : [30]

✓ **Les avantages :**

- Il est environ 88 % plus petit que YOLOv4 (27 Mo contre 244 Mo)
- Il est environ 180% plus rapide que YOLOv4 (140 FPS contre 50 FPS)
- Il est à peu près aussi précis que YOLOv4 sur la même tâche (0,895 mAP contre 0,892 mAP)

✓ **Les inconvénients :**

Mais le principal problème est que pour YOLOv5, aucun document officiel n'a été publié comme les autres versions de YOLO. De plus, YOLO v5 est toujours en cours de développement et nous recevons des mises à jour fréquentes d'ultralytics, les développeurs peuvent mettre à jour certains paramètres à l'avenir.

2.4 Conclusion :

Dans ce chapitre, nous avons examiné en détail les différentes approches de détection d'objets basées sur l'apprentissage en profondeur. Nous avons commencé par passer en revue les méthodes les plus populaires et largement utilisées, en explorant leurs architectures respectives. Ensuite, nous nous sommes concentrés sur la méthode YOLO, en expliquant en quoi elle consiste et comment elle fonctionne. Par la suite, nous nous sommes penchés plus spécifiquement sur YOLOv5, en fournissant une explication détaillée de ses caractéristiques et de ses avantages. Notre choix est donc tombé sur YOLOv5 pour implémenter notre système. Dans le chapitre suivant nous expliquerons comment l'utiliser pour réaliser notre conception.

chapitre3

conception

3.1 Introduction

Dans ce chapitre, nous aborderons en détail les différentes phases de conception du système, en mettant l'accent sur l'utilisation des diagrammes UML (Unified Modeling Language) pour modéliser et décrire l'architecture et le fonctionnement du système. Dans la première partie du chapitre, nous présenterons un schéma général du système, qui illustre les différentes composantes et leur interaction. Nous détaillerons ensuite la conception du système en utilisant les diagrammes UML, en mettant en évidence les différentes classes, les flux de données et les interactions entre les composantes principales. Le but de cette phase est de fournir une base solide pour la mise en œuvre du système, en définissant clairement les rôles et les responsabilités de chaque composante, ainsi que les échanges d'informations nécessaires pour assurer un fonctionnement fluide et efficace du système.

3.2 Conception du système :

Dans cette section, nous présentons schéma générale proposée de Notre systéme.

3.2.1 Le schéma générale :

Dans ce projet, nous exploiterons le sens auditif pour permettre à une personne malvoyante de "visualiser" les objets qui se trouvent devant elle grâce à la caméra. Nous mettrons en œuvre l'algorithme de pointe "You Only Look Once: Unified, Real-Time Object Detection" (YOLO) entraîné sur l'ensemble de données yolov5s pour identifier les objets présents devant la personne. Ensuite, nous associerons une étiquette à chaque objet identifié, puis nous la convertirons en une sortie audio en utilisant la technologie de synthèse vocale (Text to Speech, TTS). Cette sortie audio permettra à la personne de prendre conscience de son environnement en écoutant la description des objets détectés. Ainsi, notre système utilisera le sens de l'ouïe pour fournir une perception anticipée de l'environnement à la personne malvoyante.

Le schéma suivant représente les schéma global de notre conception, qui sera détaillée par la suite.

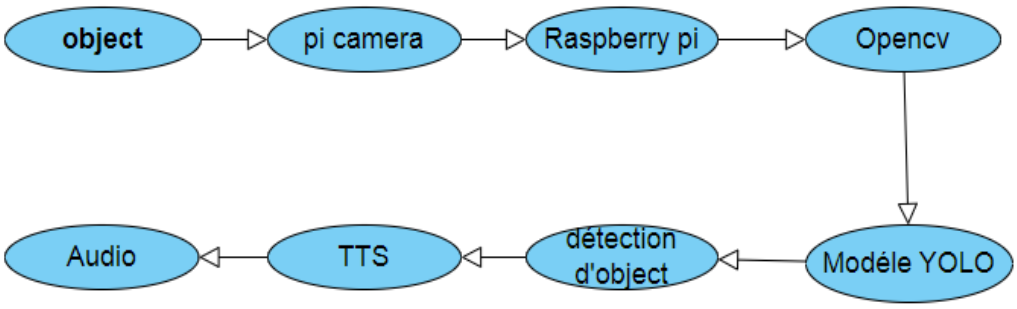


Figure3. 1 : Le schéma général

3.2.1.1 Opencv : [31]

Dans ce projet, nous utilisons la bibliothèque OpenCV, spécialisée dans la vision par ordinateur, pour développer un système d'assistance pour les personnes malvoyantes. OpenCV nous permet de traiter les images et les vidéos en temps réel, en détectant et en localisant les objets présents dans l'environnement.

En intégrant OpenCV avec d'autres bibliothèques telles que NumPy, nous pouvons effectuer des opérations avancées de traitement d'images, notamment la détection d'objets spécifiques, l'extraction de caractéristiques et l'analyse des données visuelles.

La première version d'Opencv était la 1.0. Opencv est publié sous une licence BSD elle est donc gratuite pour une utilisation académique et commerciale. Elle possède des interfaces C++, C, Python et Java et prend en charge Windows, Linux, Mac OS, iOS et Android. Lors de la conception d'Opencv, l'accent était mis sur les applications en temps réel pour l'efficacité des calculs. Tout est écrit en C/C++ optimisé pour tirer parti du traitement multicœur.

Figure3. 2 : opencv

3.2.1.2 Yolov5 :

Dans ce projet, nous utilisons des détecteurs d'objets tels que YOLOv5, qui est formé pour détecter divers objets. Le processus d'entraînement consiste à utiliser un ensemble d'images avec leurs annotations correspondantes pour ajuster le modèle et lui apprendre à reconnaître les objets. Une fois l'entraînement terminé, nous obtenons un fichier modèle qui contient les connaissances acquises par le modèle YOLOv5. Ce modèle pré-entraîné est capable de détecter différents types d'objets tels que des personnes, des voitures, des vélos, des chiens, des chats, des avions, des bateaux, et bien d'autres. [32]

Dans notre programme, nous effectuons quatre étapes simples :

- Chargement du modèle YOLOv5 : Nous importons le modèle pré-entraîné YOLOv5 dans notre programme.
- Exécution du modèle YOLOv5 sur le flux vidéo : Nous alimentons le flux vidéo capturé par la caméra dans le modèle YOLOv5 pour détecter les objets présents dans chaque image
- Traitement des sorties du modèle : Nous traitons les sorties du modèle pour extraire les classes et les coordonnées des boîtes englobantes pour chaque objet détecté.

- Utilisation des informations obtenues : Nous utilisons les informations extraites pour prendre des décisions ou générer des sorties appropriées, telles que des descriptions audio des objets détectés.

Figure3. 3 : yolov5

3.2.1.3 Object détection : [33]

La détection d'objets est l'une des tâches de vision par ordinateur les plus importantes. En un mot, étant donné une image, un détecteur d'objet trouvera : qui détecte les instances d'objets sémantiques dans les images/caméras/vidéos (en créant des cadres de délimitation autour d'eux dans notre cas).

Nous obtiendrons également les coordonnées de la boîte englobant de chaque objet détecté dans nos images, superposerons les boîtes sur les objets détectés.

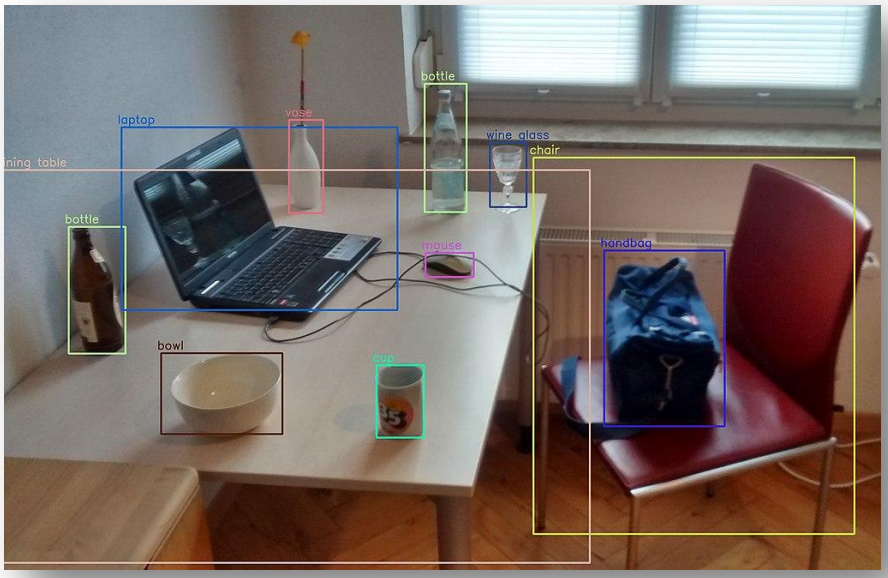


Figure3. 4 : Object détection

3.2.1.4 Conversion text en voix (text-to-speech(TTS)) : [34]

pyttsx3 est une bibliothèque de conversion texte-parole en Python. Contrairement aux bibliothèques alternatives, elle fonctionne hors ligne et elle est compatible avec Python 2 et 3. Une application appelle la fonction de fabrique pyttsx3.init() pour obtenir une référence à un pyttsx3. C'est un outil très facile à utiliser qui convertit le texte saisi en parole.

La prédiction de classe des objets détectés dans chaque trame sera une chaîne, par ex. "chat". Nous obtiendrons également les coordonnées (essentiellement la distance) des objets dans l'image. Nous pourrons ensuite envoyer le texte au package Text-to-Speech pyttsx3.

.

Figure3. 5 : Texte pour parler pyttsx3

3.2.2 Digramme de cas d'utilisation :

Ce type de diagramme met en évidence les différentes fonctionnalités de notre application, offrant à l'utilisateur la possibilité de démarrer le système et d'accéder aux services présentés dans le diagramme de cas d'utilisation ci-dessous :

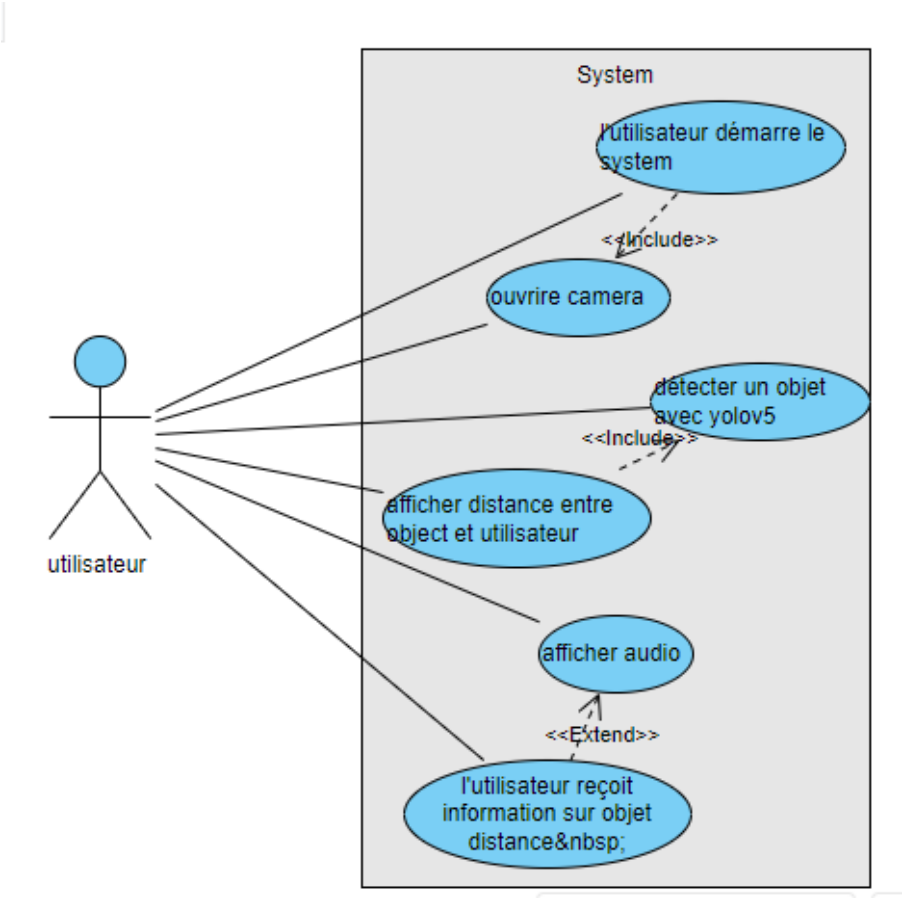


Figure3. 6 : Digramme de cas d'utilisation

3.2.3 Digramme de séquence :

Pour faciliter la compréhension de fonctionnement de notre application nous utilisant un scénario représenter dans le diagramme de séquence suivant.

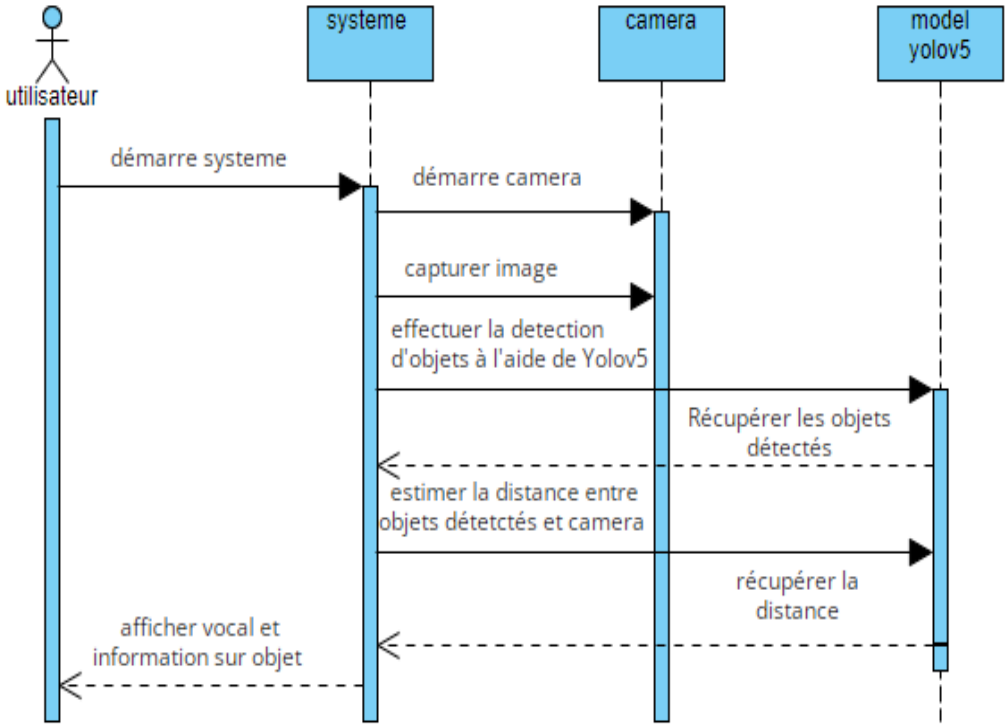


Figure3. 7 : Digramme de séquence

3.2.4 Digramme de class :

Le diagramme de classes de notre application représente les différentes composantes principales de notre système et décrit les relations entre elles. Il permet de visualiser la structure de notre application et les interactions entre les classes.

Dans ce diagramme, nous identifions les différentes classes qui composent notre système, ainsi que leurs attributs et méthodes. Nous utilisons des associations pour représenter les relations entre les classes, telles que l'agrégation, la composition, l'héritage, etc. Ces relations nous aident à comprendre comment les différentes classes interagissent les unes avec les autres.

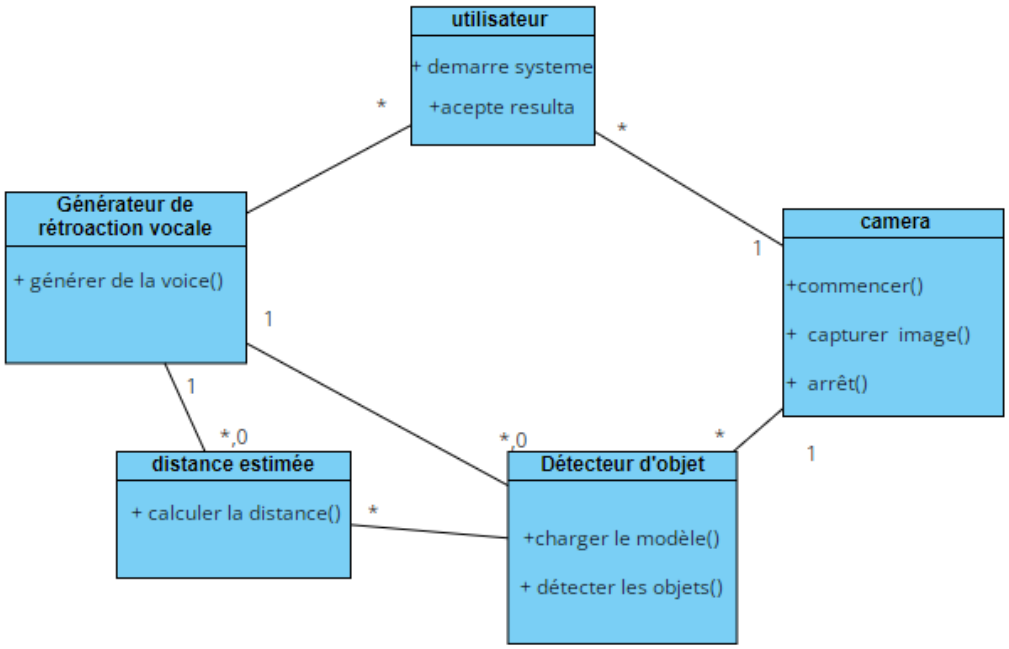


Figure3. 8 : Digramme de class

3.3 Conclusion :

En conclusion, Les diagrammes UML ont permis de visualiser de manière claire et concise la structure, les fonctionnalités et les interactions du système.

Le diagramme de cas d'utilisation a joué un rôle essentiel en identifiant les acteurs, les actions et les services du système. Il nous a aidés à comprendre les besoins des utilisateurs et à définir les fonctionnalités clés à mettre en œuvre. Grâce à ce diagramme, nous avons pu établir une base solide pour la conception de l'interface utilisateur et l'expérience utilisateur globale.

Le diagramme de classes a fourni une représentation visuelle de l'architecture interne du système. Il a détaillé les classes, les attributs, les méthodes et les relations entre les différentes entités.

En utilisant ces diagrammes, nous avons pu prendre des décisions éclairées tout au long du processus de conception. Ils ont servi de référence pour la mise en œuvre du système, en assurant la cohérence et la simplicité.

chapitre4

Implémentation et Résultat

4.1 Introduction :

Guidés par les diagrammes UML, ce chapitre se concentrera sur l'implémentation de notre système de détection d'objets basé sur l'apprentissage en profondeur. Nous suivrons différentes étapes pour concrétiser notre approche. Python, un langage de programmation polyvalent, sera notre choix pour mettre en œuvre les modèles d'apprentissage en profondeur. Nous tirerons parti de l'environnement de développement PyCharm, qui offre des fonctionnalités avancées pour faciliter le développement et le débogage de notre solution. De plus, nous examinerons la possibilité d'intégrer la caméra Raspberry Pi (PiCam) dans notre système, ce qui nous permettra de capturer des images en temps réel pour la détection d'objets. Finalement, nous explorerons l'option d'inclure des instructions audio pour une interaction plus intuitive avec les utilisateurs.

4.2 Plateformes et outils de programmation utilisés:

Dans cette partie, nous présenterons les matériels et logiciels utilisés de notre application.

a. Logiciel:

Dans cette partie nous allons définir l'environnement de travail et le langage de programmation.

- l'environnement de travail:

PyCharm: PyCharm est un environnement de développement intégré utilisé pour programmer en Python. Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django. Développé par l'entreprise tchèque JetBrains, c'est un logiciel multi-plateforme qui fonctionne sous Windows, Mac OS X et GNU/Linux. Il est décliné en édition professionnelle, diffusé sous licence propriétaire, et en édition communautaire diffusé sous licence Apache. [35]

Figure4. 1 : PyCharm

➢ le langage de programmation:

- **Python:** est un langage de programmation de haut niveau, polyvalent et très populaire. Le langage de programmation Python (le dernier Python 3) est utilisé dans le développement Web, les applications d'apprentissage automatique, ainsi que dans toutes les technologies de pointe de l'industrie logicielle. Le langage de programmation Python convient très bien aux débutants, ainsi qu'aux programmeurs expérimentés dans d'autres langages de programmation comme C++ et Java. [36]

Figure4. 2 : Python

- **Installation des librairies:**

Les étapes d'implémentation sont réalisé à l'aide de diverses bibliographies telles que :

Opencv,torch, pyttx3

- **Pytorch :** PyTorch est une bibliothèque d'apprentissage automatique open source utilisée pour développer et former des modèles d'apprentissage en profondeur basés sur des réseaux de neurones. Il est principalement développé par le groupe de recherche sur l'IA de Facebook. PyTorch peut être utilisé avec Python ainsi qu'avec C++. Naturellement, l'interface Python est plus soignée. Pytorch (soutenu par des biggies comme Facebook, Microsoft, SalesForce, Uber) est immensément populaire dans les laboratoires de recherche. [37]

Figure4. 3: pytorch

b. Matériel:

Dans cette partie nous allons définir le Raspberry pi et le pi caméra et les écouteurs utilisés.

- **Raspberry pi:**

Le Raspberry Pi est un nano-ordinateur monocarte à processeur ARM de la taille d'une carte de crédit conçu par des professeurs du département informatique de l'université de Cambridge dans le cadre de la fondation Raspberry Pi3. [38]

Le Raspberry Pi est un ordinateur peu coûteux de la taille d'une carte de crédit qui se branche sur un écran d'ordinateur ou un téléviseur et utilise un clavier et une souris standard. Il s'agit d'un petit appareil capable qui permet aux personnes de tous âges

d'explorer l'informatique et d'apprendre à programmer dans des langages comme Scratch et Python en vidéo haute définition, à faire des feuilles de calcul, du traitement de texte et à jouer à des jeux. [39]

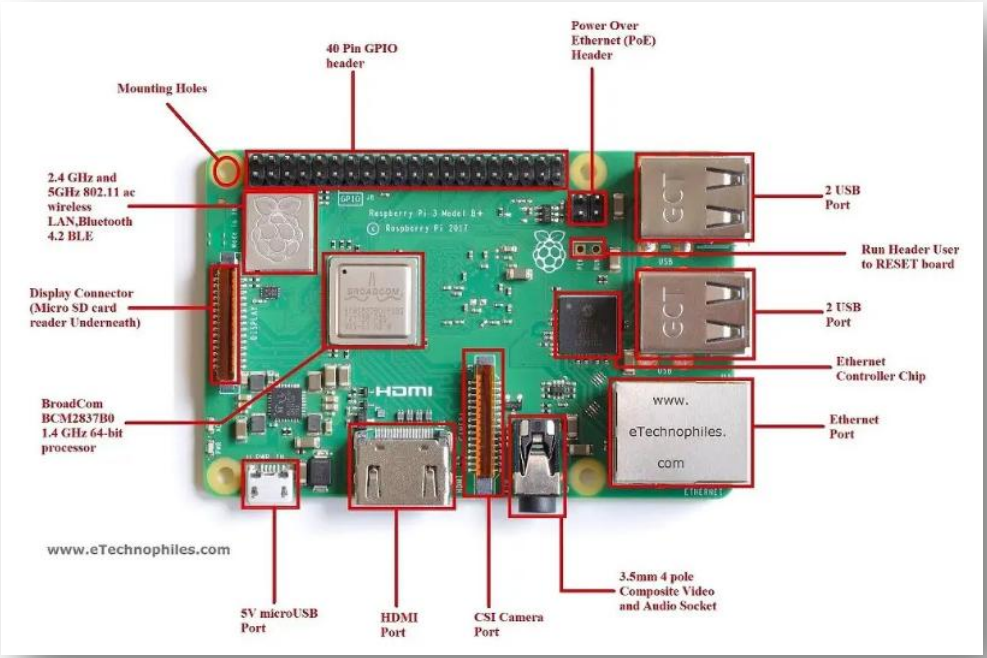


Figure4. 4 : Le Raspberry Pi

- **Pi caméra:** [40]

Le module caméra pi est une caméra qui peut être utilisée pour prendre des photos et des vidéos haute définition.

La carte Raspberry Pi possède une interface CSI (Camera Serial Interface) à laquelle nous pouvons attacher directement le module PiCamera.

Ce module de caméra Pi peut se connecter au port CSI du Raspberry Pi à l'aide d'un câble ruban à 15 broches.

Figure4. 5 : Pi caméra

- **les écouteurs :**

Un écouteur est un dispositif qui se place dans une oreille et qui permet de restituer des contenus sonores. Il transforme des signaux électriques en sons perceptibles par l'oreille, on les accroche ou les met sur celle-ci pour l'écoute de sons. Il est également appelé oreillette ou casque par analogie de fonction avec le casque audio. [41]

Figure4. 6 : les écouteurs

4.3 Résultats obtenus :

La code Python détection de et la reconnaissance d'objets et le convertir en son. se base sur un modèle yolov5 Pour ce cas, a été entraîné pour reconnaître une liste de 80 objets tels qu'une bouteille, un chien, un chat, une personne…etc.

Tout d'abord, nous importons les packages nécessaires pour que notre modèle puisse fonctionner.

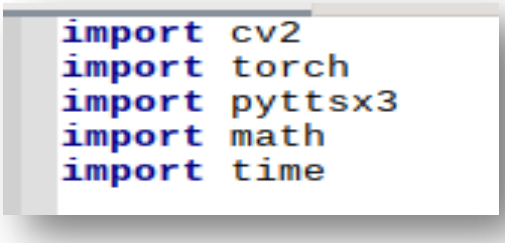


Figure4. 7: les packages

Ensuite, nous devons charger notre modèle. Pour ce faire, on utilise la fonction : torch

Figure4. 8:code de charge de model

Puis,Initialiser le moteur de synthèse vocale.

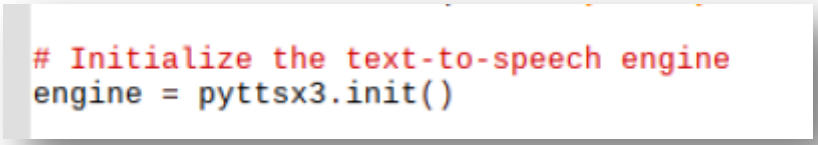


Figure4. 9 : code de lire synthèse vocale.

Après avoir initialisé la connexion avec la caméra et capturé un cadre, nous appliquons le modèle YOLOv5 au cadre. Ensuite, nous extrayons les boîtes englobantes et les étiquettes du résultat obtenu. Enfin, nous dessinons les cadres de délimitation et les étiquettes correspondantes sur le cadre capturé

```python
# Open a connection to the camera
cap = cv2.VideoCapture(0)

while True:
    # Capture a frame from the camera
    ret, frame = cap.read()

    if not ret:
        print("Failed to capture frame.")
        break
df = result.pandas().xyxy[0]
print(df)

# Check if any objects have been detected
if len(df) > 0:
    object_detected = True
else:
    object_detected = False

# Draw the bounding boxes and labels on the frame
for ind in df.index:
    x1, y1 = int(df['xmin'][ind]), int(df['ymin'][ind])
    x2, y2 = int(df['xmax'][ind]), int(df['ymax'][ind])
    label = df['name'][ind]
    cv2.rectangle(frame, (x1, y1), (x2, y2), (255, 255, 0), 2)
    cv2.putText(frame, label, (x1, y1 - 5), cv2.FONT_HERSHEY_PLAIN, 2, (2
```

Figure4. 10 :le code de détection Object

Ensuite, Calculez la distance entre l'objet en fonction de sa largeur en pixels.

```python
# Calculate the distance to the object based on its width in pixels
object_width_px = x2 - x1
distance = actual_width * focal_length / object_width_px
distance = round(distance, 2)
dist_text = 'Distance: ' + str(distance) + ' meter'
cv2.putText(frame, dist_text, (x1, y2 + 20), cv2.FONT_HERSHEY_PLAIN, 2, (255,
```

Figure4. 11 : le code de distance

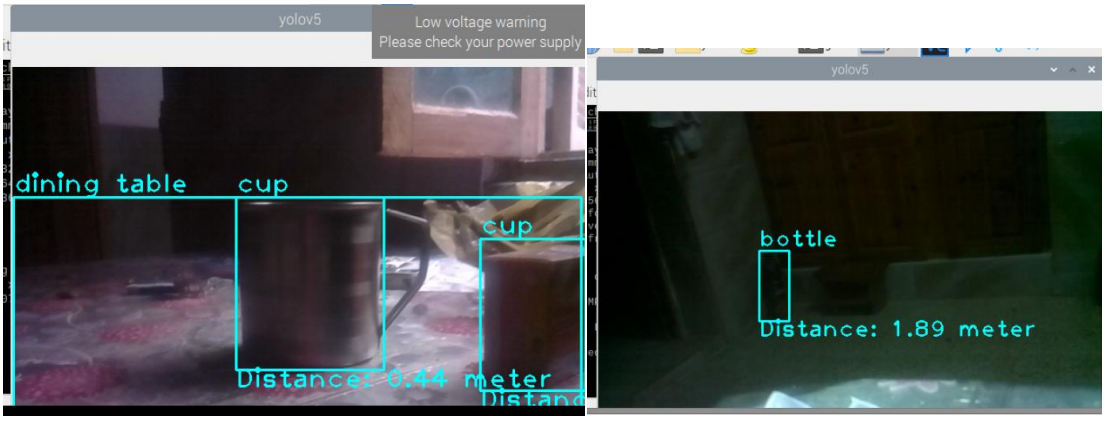
Enfin, prononcez l'étiquette et la distance à l'aide de la synthèse vocale

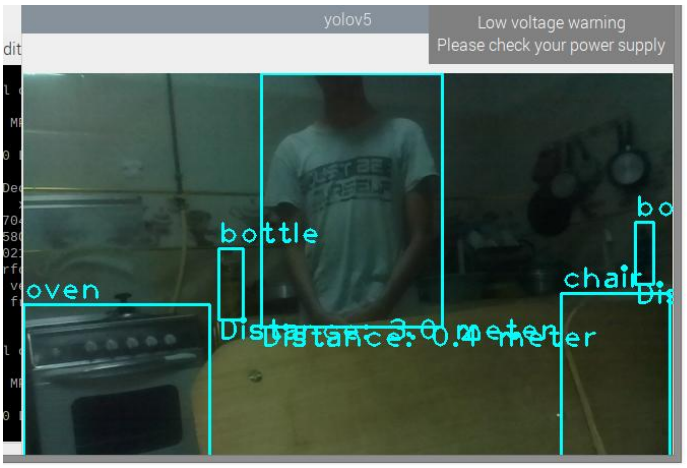
```python
# Speak the label and distance using text-to-speech
engine.say(label)
engine.say('and distance of')
engine.say(str(distance))
engine.say('meter')
engine.runAndWait()
```

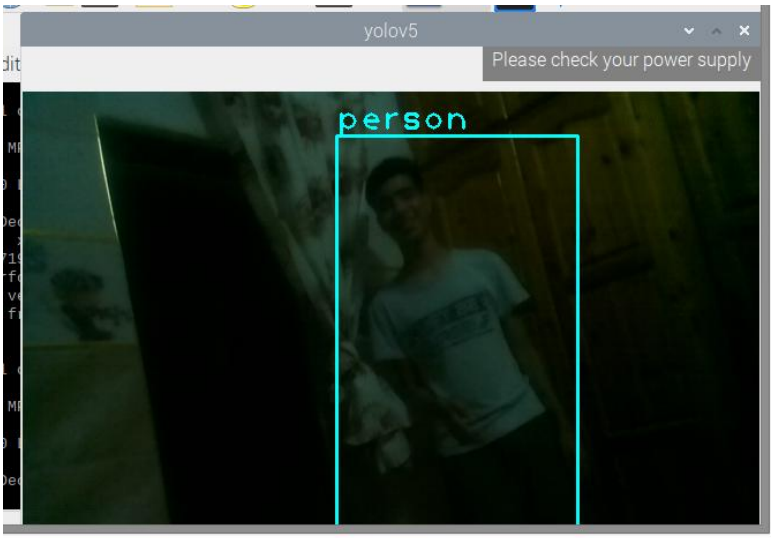
Figure4. 12 :le code de la synthèse vocale

- ✓ **Résultat de exécution :**

Dans cette partie nous allons tester/valider les performances du d'système.







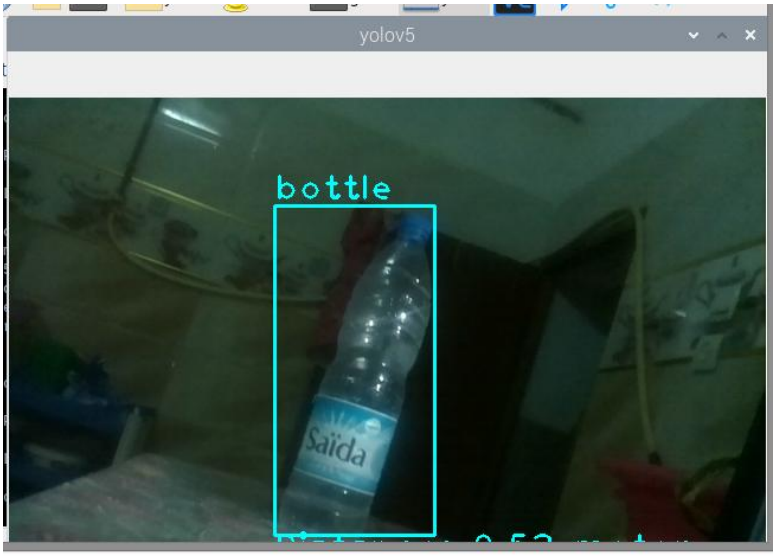


Figure4. 13: Résultats obtenus

4.4 Conclusion :

En conclusion, ce chapitre a présenté l'implémentation de notre système de détection d'objets basé sur l'apprentissage en profondeur. Nous avons suivi les diagrammes UML pour guider notre processus de développement. En utilisant Python et l'environnement de développement PyCharm, nous avons pu mettre en œuvre notre modèle d'apprentissage en profondeur de manière efficace. De plus, l'intégration de la caméra Raspberry Pi (PiCam) nous a permis de capturer des images en temps réel pour la détection d'objets. Par l'ajout d'instructions audio, notre système offre une interaction plus intuitive avec l'environnement pour les personnes malvoyantes.

Conclusion général

Notre mémoire a exploré en détail les techniques de détection d'objets basées sur l'apprentissage en profondeur. Nous avons commencé par une revue des méthodes les plus courantes, telles que Faster R-CNN, YOLO et SSD, en mettant en évidence leurs avantages et leurs limites. Ensuite, nous nous sommes concentrés sur YOLOv5, une des versions les plus stables de la famille YOLO, en expliquant son fonctionnement et ses améliorations par rapport aux versions antérieures.

Nous avons également abordé l'implémentation pratique de ces méthodes, en utilisant Python comme langage de programmation polyvalent et en exploitant les outils tels que PyCharm pour faciliter le développement. L'intégration de la caméra Raspberry Pi (PiCam) et la possibilité d'ajouter des instructions audio ont enrichi notre système de détection d'objets, le rendant plus interactif et convivial pour les personnes malvoyantes.

Ce mémoire met en évidence l'importance croissante de l'apprentissage en profondeur dans le domaine de la vision par ordinateur et offre un aperçu détaillé des techniques et des pratiques d'implémentation. Les connaissances acquises et les résultats obtenus dans ce domaine sont une base solide pour des applications futures dans des domaines tels que la sécurité, la surveillance, l'automatisation industrielle, et bien d'autres.

L’aspect technique de notre projet nous a donné la possibilité de traiter des détails qu’on n’a pas forcément rencontré dans notre parcours académique, ce que nous a enrichit nos connaissances et nous a donné idée sur le monde professionnel.

Nous visons à refaire l’expérience avec un Raspberry Pi 4, ce qui va améliorer les performances de notre système. Nous pensons à ajouter d’autres fonctionnalités, tel que la détection des portes. L’interaction personne malvoyante – Raspberry via un mic, pour paramétrer l’application sera aussi une idée à explorer.

En conclusion, ce mémoire témoigne de notre compréhension approfondie des techniques de détection d'objets basées sur l'apprentissage en profondeur et de notre capacité à les mettre en œuvre dans des scénarios réels. Il ouvre également la voie à

de nouvelles recherches et à des avancées continues dans ce domaine passionnant et en constante évolution.

Bibliographie

[1] M. F. Zohra, «CONCEPTION ET REALISATION D'UNE CANNE INTELLIGENTE,» MASTER EN ELECTRONIQUE, Université Abdelhamid Ibn Badis Mostaganem, 02/07/2020.

[2] S. S. eddine, «Etude et réalisation d'une canne intelligente pour les non-voyants,» MEMOIRE DE FIN D'ETUDES, Université Larbi Ben M'hidi - Oum El Bouaghi, juillet 2019.

[3] «https://www.mieuxvivresamalvoyance.com/c-est-quoi-etre-malvoyant/,» [En ligne]. Available: consult 22:30. [Accès le 15 2 2023].

[4] B. M. N. /. B. A. D. F. Zohra, «DETECTION ET RECONNAISSANCE DE VISAGE DANS UNE IMAGE,» Projet de Fin d'Etudes, Université –Ain Temouchent- Belhadj Bouchaib , 2020/2021.

[5] B. O. Akram, «La proposition d'une nouvelle approche basée Deep Learning pour la prédiction du cancer d u sain,» Mémoire de fin d'étude, Université L'arbi Ben M'hidi Oum El Bouaghi, 2019-2020.

[6] «https://commentouvrir.com/definitions/les-reseaux-adversariens-generatifs-expliques,» , [En ligne]. Available: consult 22:55. [Accès le 5 6 2023].

[7] A. F. Heba Hakim*, «Survey: Convolution Neural networks in Object Detection,» *Journal of Physics: Conference Series,* Vols. %1 sur %2doi:10.1088/1742-6596/1804/1/012095, p. 2, 2021 .

[8] «Réseaux neuronaux convolutifs (CNN), Deep Learning et vision par ordinateur,» 23:30, [En ligne]. Available: https://www.intel.fr/content/www/fr/fr/internet-of-things/computer-

vision/convolutional-neural-networks.html. [Accès le 14 2 2023].

[9] É. B. J. QUEIROS, «Deep Learning sur des données 3D,» Nantes Université Polytech Nantes, Janvier-Mai 2020.

[10] M. Fethia, «Détection d'objets par Deep Neural Network à l'aide du modèle YOLO en temps réel.,» Mémoire de Fin d'Etudes Master, Université 8 Mai 1945 –Guelma-, 2020/2021.

[11] «Mieux comprendre le deep learning appliqué à la reconnaissance d'images,» 21:15, [En ligne]. Available: https://france.devoteam.com/paroles-dexperts/mieux-comprendre-le-deep-learning-applique-a-la-reconnaissance-dimages/. [Accès le 5 4 2023].

[12] «Convolutional Neural Network : Tout ce qu'il y a à savoir,» 00:12, [En ligne]. Available: https://datascientest.com/convolutional-neural-network. [Accès le 11 6 2023].

[13] B. M. BOUHARKET, «Système d'Identification de Personnes via la Plaque d'Immatriculation de leurs Véhicules,» MEMOIRE, UNIVERSITE IBN KHALDOUN - TIARET, 18/09/2022.

[14] T. H. Eddine, «La détection d'objet avec OpenCV et deep learning,» MÉMOIRE DE MASTER2, Université de Biskra, 2019/2020.

[15] «https://www.sublimeo.com/deep-learning/,» [En ligne]. Available: consult 23:20. [Accès le 25 3 2023].

[16] A. N. /. M. Zouleikha, «Suivi d'objets à l'aide de l'algorithme,» Mémoire de Fin d'Etude, U n i v ersité Larbi Ben M'Hidi d'Oum El Bouaghi, 2021-2022.

[17] «https://ledatascientist.com/detection-recu-rcnn/,» [En ligne]. Available: consult 02:30. [Accès le 23 4 2023].

[18] «Single Shot Detector (SSD) + Architecture of SSD,» 19:35, [En ligne]. Available: https://iq.opengenus.org/single-shot-detector. [Accès le 20 3 2023].

[19] K. D. Souhila, «Classification des obstacles par la technique d'apprentissage profond,» Université Aboubakr Belkaïd –Tlemcen –, 20/09/2020.

[20] «guide-to-object-detection-using-yolo,» / by jantakarn / medium, [En ligne]. Available: https://medium.com/@aumjantakarn/guide-to-object-detection-using-yolo-33d74d7091d9. [Accès le 12 3 2023].

[21] «selecting-the-right-bounding-box-using-non-max-suppression-with-implementation,» 20:45, [En ligne]. Available: https://www.analyticsvidhya.com/blog/2020/08/selecting-the-right-bounding-box-using-non-max-suppression-with-implementation/. [Accès le 30 3 2023].

[22] M. D. F. E. Tahar., «Réalisation d'un système intelligent pour la collecte,» UNIVERSITE YAHIA FARES DE MEDEA, 2020-2021.

[23] «Object Detection with YOLOv5 and PyTorch,» 12:12, [En ligne]. Available: https://www.section.io/engineering-education/object-detection-with-yolov5-and-pytorch/. [Accès le 5 3 2023].

[24] M. ABDELOUAHEB, «Utilisation de méthodes de Deep learning pour l'extraction de texte dans les images,» Mémoire de master2, Université Mohamed Khider – BISKRA, 2020-2021.

[25] I. Katsamenis, «TraCon: A novel dataset for real-time trafficcones detection using deep learning,» p. 4, 24 5 2022.

[26] «CSPNet: A New Backbone that can Enhance Learning Capability of CNN,» p. 2, 25 4 2023.

[27] M. Y. D. Z. a. Y. G. Xuan Zhang1*, «Marine ship detection and classification based on YOLOv5,» *Journal of Physics: Conference Series*, p. 3, 2022.

[28] «https://huggingface.co/docs/timm/models/csp-darknet,» [En ligne]. Available: consult 23:20. [Accès le 18 3 2023].

[29] «ASFF-YOLOv5: Multielement Detection Method for Road Traffic in UAV

Images Based on Multiscale Feature Fusion,» 20 4 2023.

[30] «How to Use Yolo v5 Object Detection Algorithm for Custom Object Detection 10:15,» [En ligne]. Available: https://www.analyticsvidhya.com/blog/2021/12/how-to-use-yolo-v5-object-detection-algorithm-for-custom-object-detection-an-example-use-case/. [Accès le 23 4 2023].

[31] «opencv-overview,» 14:40, [En ligne]. Available: https://www.geeksforgeeks.org/opencv-overview/. [Accès le 23 4 2023].

[32] «Lightweight object detection algorithm based on YOLOv5 for unmanned surface vehicles,» 14:03, [En ligne]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2022.1058401/full. [Accès le 15 5 2023].

[33] «Object Detection in 2023: The Definitive Guide,» 15:30, [En ligne]. Available: https://viso.ai/deep-learning/object-detection/. [Accès le 23 5 2023].

[34] «Object Detection with Voice Feedback — YOLO v3 + gTTS,» 21:19. [En ligne]. Available: https://towardsdatascience.com/object-detection-with-voice-feedback-yolo-v3-gtts-6ec732dca91. [Accès le 8 6 2023].

[35] «https://fr.wikipedia.org/wiki/PyCharm,» [En ligne]. Available: consult 01:00. [Accès le 2 6 2023].

[36] «https://www.comment-devenir-developpeur.com/cours-de-python-gratuit,» [En ligne]. Available: consult 02:00. [Accès le 30 5 2023].

[37] «https://towardsdatascience.com/introduction-to-py-torch-13189fb30cb3,» [En ligne]. Available: consult 13:05. [Accès le 13 6 2023].

[38] «Raspberry_Pi,» 00:30, [En ligne]. Available: https://fr.wikipedia.org/wiki/Raspberry_Pi. [Accès le 25 5 2023].

[39] «what-is-a-raspberry-pi,» 01:00, [En ligne]. Available:

https://www.raspberrypi.org/help/what-is-a-raspberry-pi/. [Accès le 28 5 2023].

[40] «pi-camera-module-interface-with-raspberry-pi-using-python,» 20:30, [En ligne]. Available: https://www.electronicwings.com/raspberry-pi/pi-camera-module-interface-with-raspberry-pi-using-python. [Accès le 20 5 2023].

[41] «https://fr.wikipedia.org/wiki/Ecouteur,» [En ligne]. Available: consult 21:30. [Accès le 25 5 2023].

[42] «what is a power bank,» 22:30, [En ligne]. Available: https://www.qualitylogoproducts.com/blog/what-is-a-power-bank/. [Accès le 12 6 2023].