



Architecture de Détection d'Objets YOLO

Présenté par :

Ibrahima Ndende SOW

Moustapha MANGANE

Diaramouna Bakar SOUMARE

Mbathio Ba DIALLO

PLAN

Introduction

I. Caractéristiques principales et fonctionnement

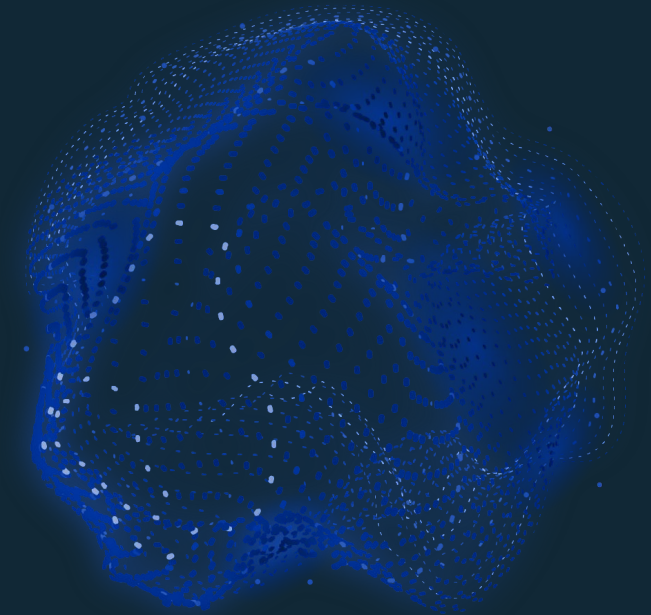
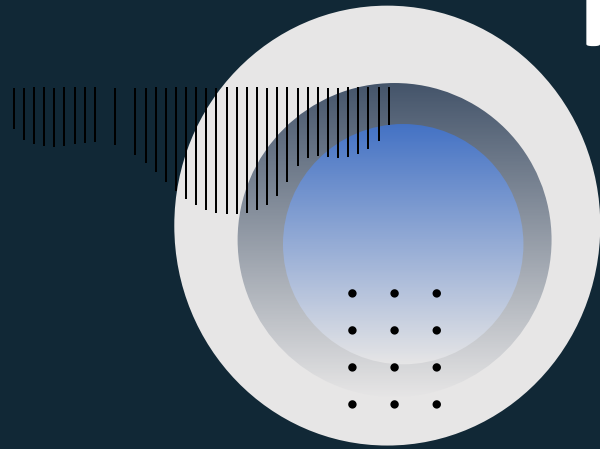
II. Avantages

III. Limites

IV. Input et Output du Modèle

V. Implémentation

Conclusion



INTRODUCTION



La détection d'objets est une tâche fondamentale en vision par ordinateur. Elle consiste à localiser et classifier plusieurs objets dans une image ou une séquence d'images. Parmi les architectures les plus performantes, la famille des méthodes YOLO (You Only Look Once) s'est imposée comme une solution de référence grâce à sa rapidité et son efficacité en temps réel

I. Caractéristiques principales et fondamental du concept de YOLO

Introduction en 2016

Présenté par Joseph Redmon et ses collègues dans l'article "You Only Look Once: Unified, Real-Time Object Detection" à CVPR 2016.

1

Différence Fondamentale

Contrairement aux méthodes R-CNN qui utilisent plusieurs étapes, YOLO effectue toutes les opérations en une seule passe à travers un réseau de neurones convolutionnel.

2

Approche Révolutionnaire

Traite la détection d'objets comme un problème de régression unique, générant directement les coordonnées des boîtes et les classes des objets.

3

Fonctionnement Global de YOLO



Prétraitement

L'image d'entrée est redimensionnée à une taille fixe (416x416 ou 640x640) puis normalisée.



Division en grille

L'image est divisée en une grille SxS. Chaque cellule est responsable de prédire les objets dont le centre tombe dans cette cellule.



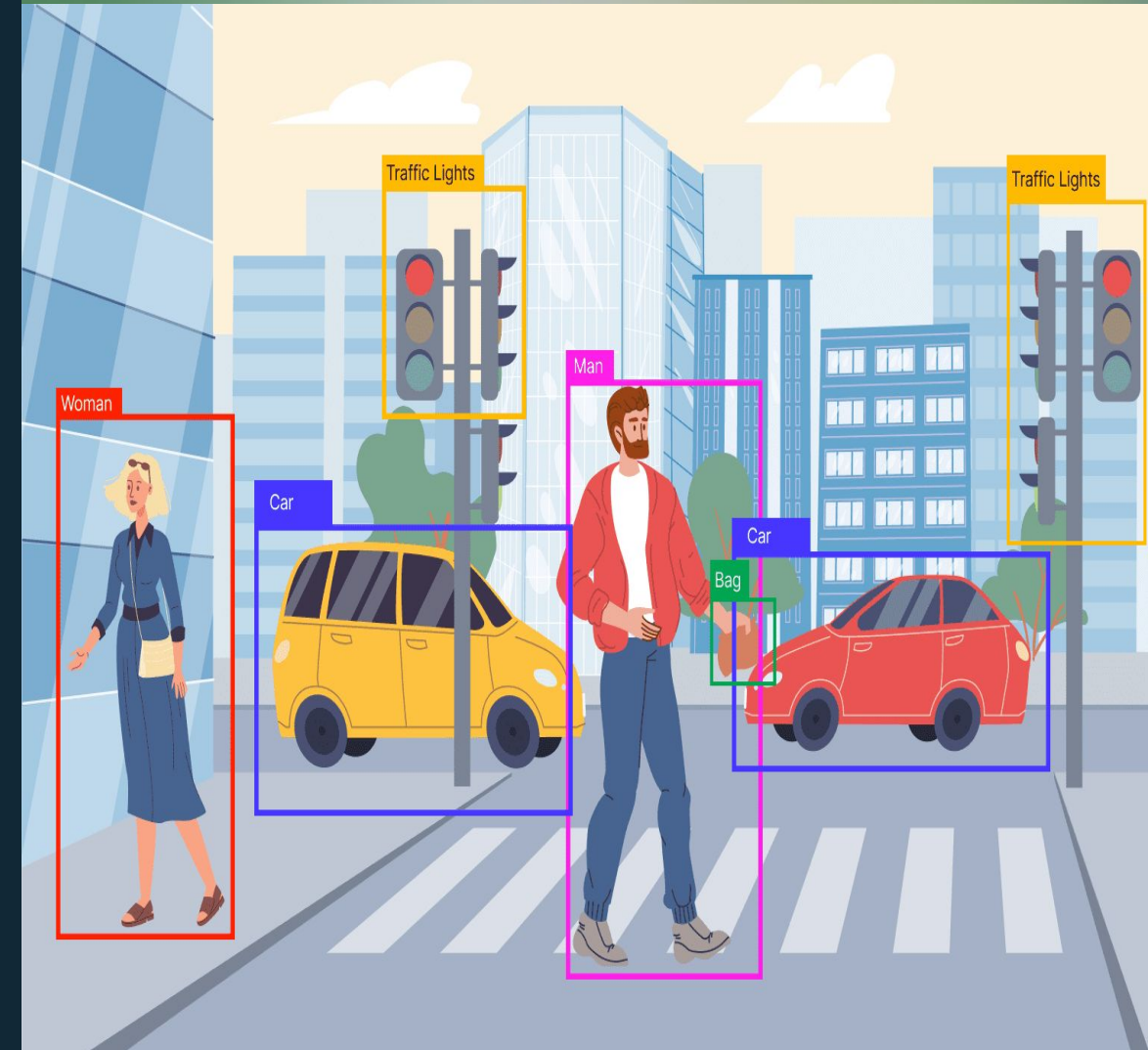
Prédiction par cellule

Pour chaque cellule, le réseau prédit B boîtes de délimitation (x, y, w, h, confiance) et une distribution de classes.



Post-traitement

Filtrage par seuil de confiance, Non-Maximum Suppression pour éliminer les doublons, génération du résultat final.



Évolution de l'Architecture YOLO

YOLOv1

Première version utilisant une architecture allégée de GoogLeNet. Établit le concept de base mais avec des limitations sur les petits objets.

YOLOv2 et v3

Introduction de Darknet-19 et Darknet-53, des architectures convolutionnelles profondes améliorant la précision tout en maintenant la rapidité.

YOLOv4

Combine CSPDarknet53 avec PANet et SPP pour améliorer significativement la précision sans sacrifier la vitesse de traitement.

YOLOv5 et v8

Versions optimisées facilitant le déploiement avec différentes tailles de modèles (s, m, l, x) et support direct pour la segmentation et les points clés.

The YOLO timeline

TheAiEdge.io





II. Avantages de YOLO



Vitesse de traitement exceptionnelle

Capable de détecter des objets en temps réel, atteignant souvent plus de 45 images par seconde (jusqu'à 155 FPS pour YOLOv5 nano), idéal pour la vidéosurveillance, les drones et les véhicules autonomes.



Flexibilité et évolutivité

Structure modulaire permettant l'entraînement pour de multiples classes et tâches comme la segmentation sémantique, avec un déploiement facile sur des périphériques embarqués.



Détection globale et cohérente

Traite l'image entière en une seule fois, permettant de mieux capturer les dépendances contextuelles et d'avoir une meilleure cohérence spatiale que les méthodes à base de "region proposals".



Open Source et communauté active

YOLO est largement supporté Par une communauté open source très active.

III. Limites de

Difficultés avec les petits objets

Les premières versions de YOLO avaient des difficultés à détecter des petits objets proches les uns des autres, car la grille de prédiction imposait une limite stricte à la capacité de détection fine.

Précision inférieure aux méthodes two-stage

YOLO, particulièrement dans ses premières versions, obtenait une précision inférieure à Faster R-CNN ou Mask R-CNN sur les benchmarks standards comme COCO ou PASCAL VOC, sacrifiant une partie de la précision au profit de la vitesse.

Erreurs de localisation

Les prédictions globales rapides peuvent entraîner des erreurs de localisation, surtout lorsque plusieurs objets se chevauchent. La répartition fixe des boîtes limite également la capacité à distinguer des objets très proches.

Répartition fixe des boîtes

Chaque cellule ne peut prédire qu'un nombre limité de boîtes. Dans le cas où plusieurs objets sont très proches, cela limite la capacité du modèle à les distinguer.



IV. Input et Output du Modèle

Données en entrée

- Images statiques de taille fixe (par exemple 416x416 ou 640x640 pixels) pour détecter les objets dans des scènes figées
- Flux video qui peuvent être traités pour réaliser une détection d'objets en mouvement
- Webcam pour détecter des objets en temps réel

Sortie du modèle

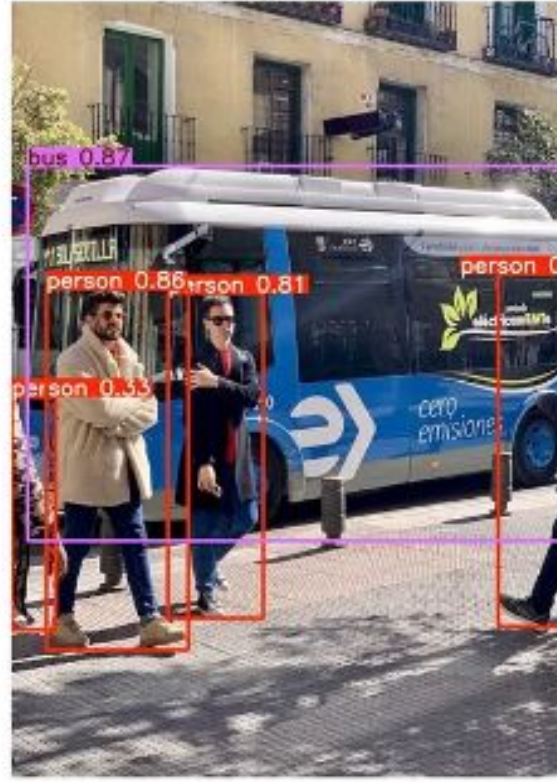
- Détection d'objets avec une "boîte englobante" délimitant l'objet à travers ses coordonnées
- Score de confiance pour chaque détection indiquant la fiabilité du modèle
- Classe prédite pour chaque objet avec une étiquette représentant sa classe

V.

IMPLEMENTATION

Image originale

Détections



Analyse d'image statique

Dans ce premier test, une image fixe est analysée par le modèle YOLO. Les objets présents sont identifiés avec une grande rapidité, et des boîtes englobantes sont tracées autour de chaque élément détecté.

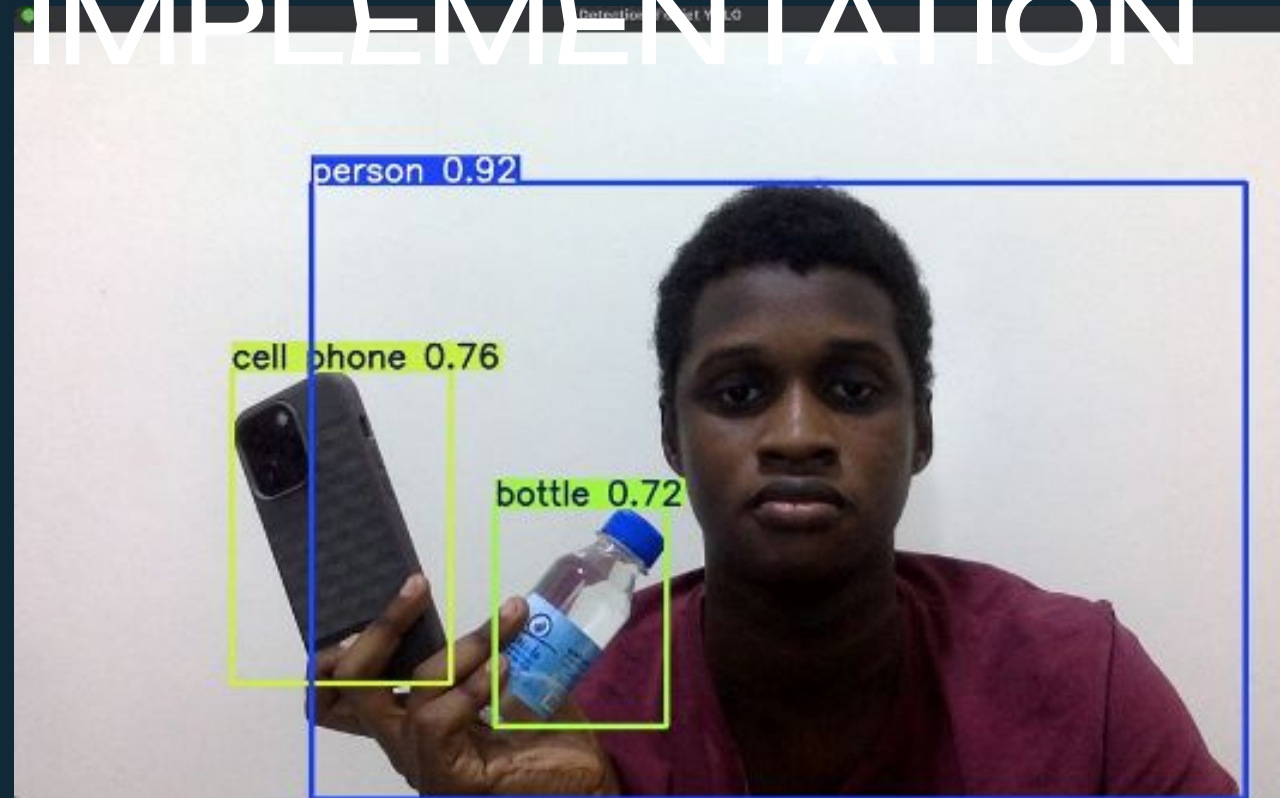


Traitement vidéo

YOLO est appliqué sur une vidéo préenregistrée. Chaque frame est traitée individuellement, permettant une détection continue des objets au fil du temps, idéale pour l'analyse de séquences de surveillance.

V.

IMPLEMENTATION



Détection en temps réel

Le modèle YOLO utilisé avec un flux en direct provenant de la webcam permet de capturer et détecter les objets dans l'environnement immédiat en temps réel, parfait pour la vidéosurveillance et l'interaction homme-machine.



CONCLUSION

YOLO a profondément transformé la manière dont les systèmes perçoivent et analysent les images en temps réel. Sa capacité à combiner vitesse, simplicité et précision en fait un outil incontournable pour toute application de vision par ordinateur nécessitant de la détection d'objets. YOLO a profondément transformé la manière dont les systèmes perçoivent et analysent les images et les vidéos en temps réel. Sa capacité à fonctionner avec des flux en direct en fait un outil incontournable pour de nombreuses applications pratiques, allant de la surveillance à la robotique interactive. Son efficacité sur les vidéos et les flux webcam renforce encore davantage sa position de leader parmi les architectures de détection d'objets.