Projet 7 : Preuve de concept avec YOLOv7

Ali-higo Ebo Adou

November 4, 2022

Outline

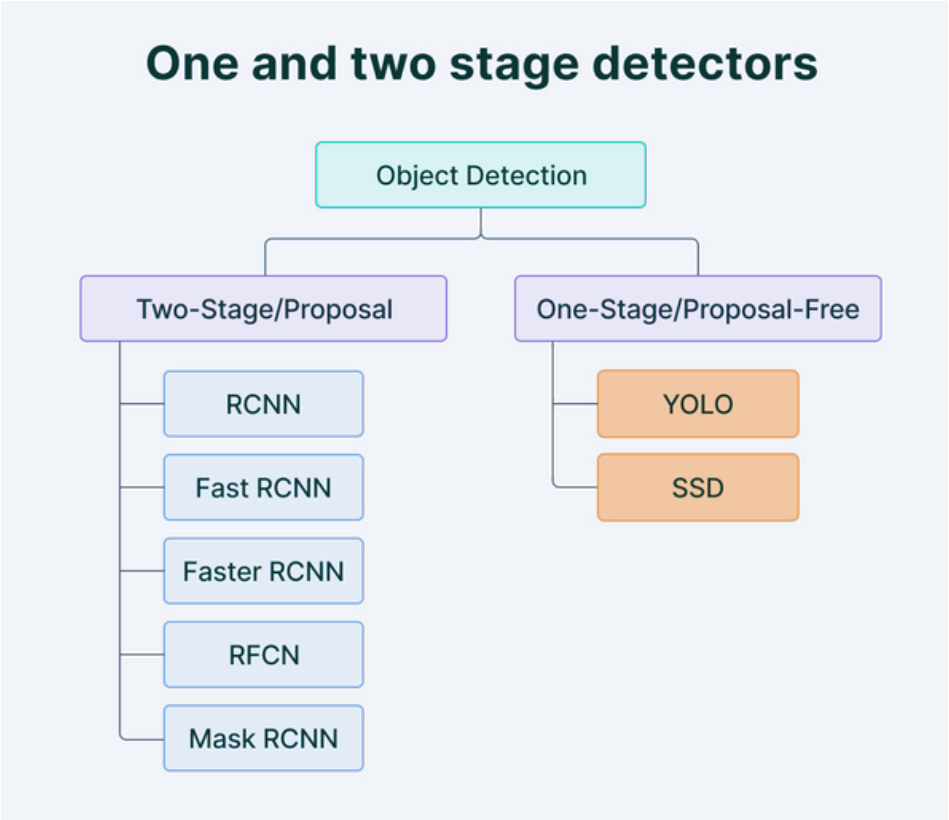
1. Introduction
2. Le jeu de données
3. Modèles de l'état de l'art à "deux passes"
4. La famille des modèles YOLO
5. Résultats du transfert d'apprentissage
6. Conclusion

Outline

1. Introduction
2. Le jeu de données
3. Modèles de l'état de l'art à "deux passes"
4. La famille des modèles YOLO
5. Résultats du transfert d'apprentissage
6. Conclusion

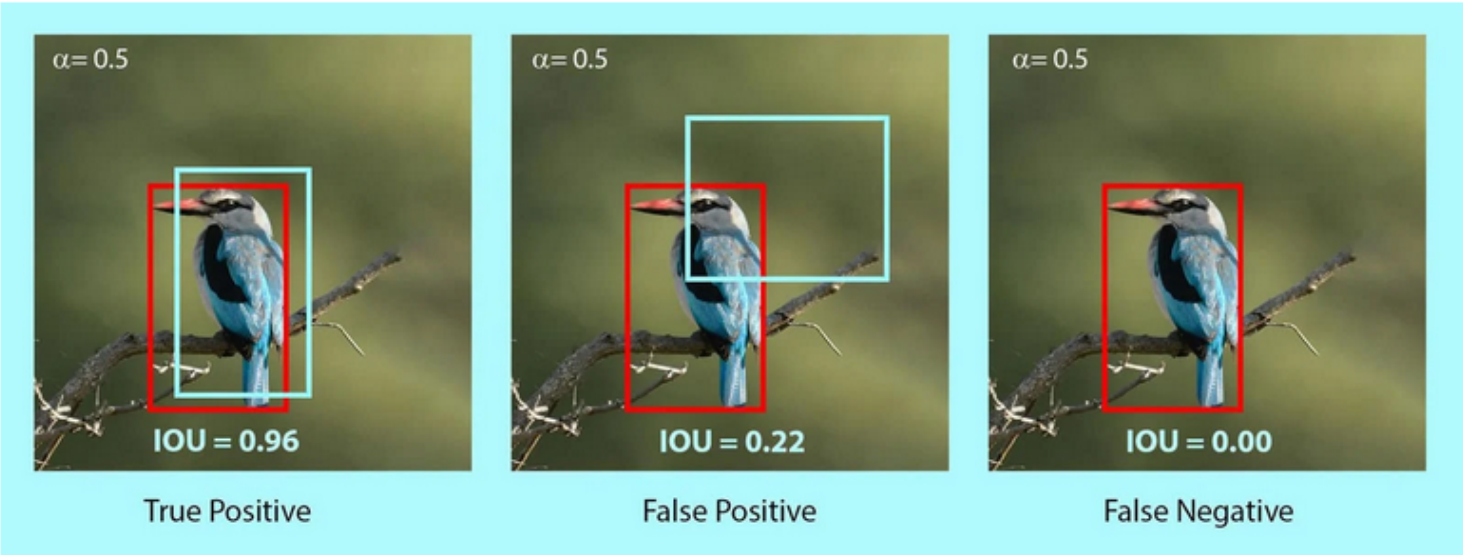
Détection d'objets en *Computer Vision* (CV)

- Classification
- Localisation



Métriques pour la localisation

- La localisation est correcte si IoU ≥ 0.5



$\Longrightarrow$ ***mAP*** : Moyenne des précisions moyennes.

Outline

1 Introduction

2 Le jeu de données

3 Modèles de l'état de l'art à "deux passes"

4 La famille des modèles YOLO

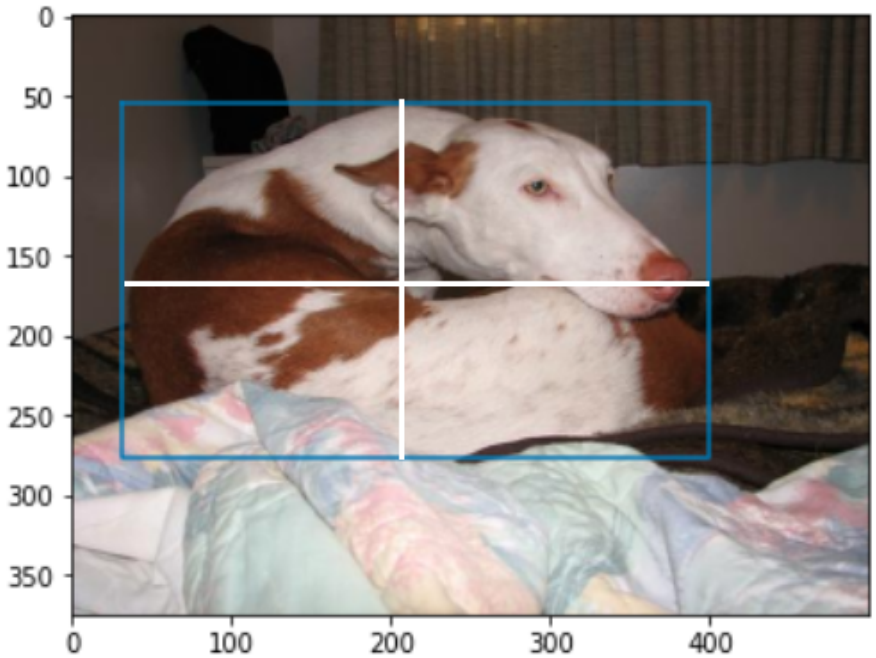
5 Résultats du transfert d'apprentissage

6 Conclusion

Stanford Dogs Dataset

- 120 races de chiens
- plus de 20 000 images (plus de 150 images par race)

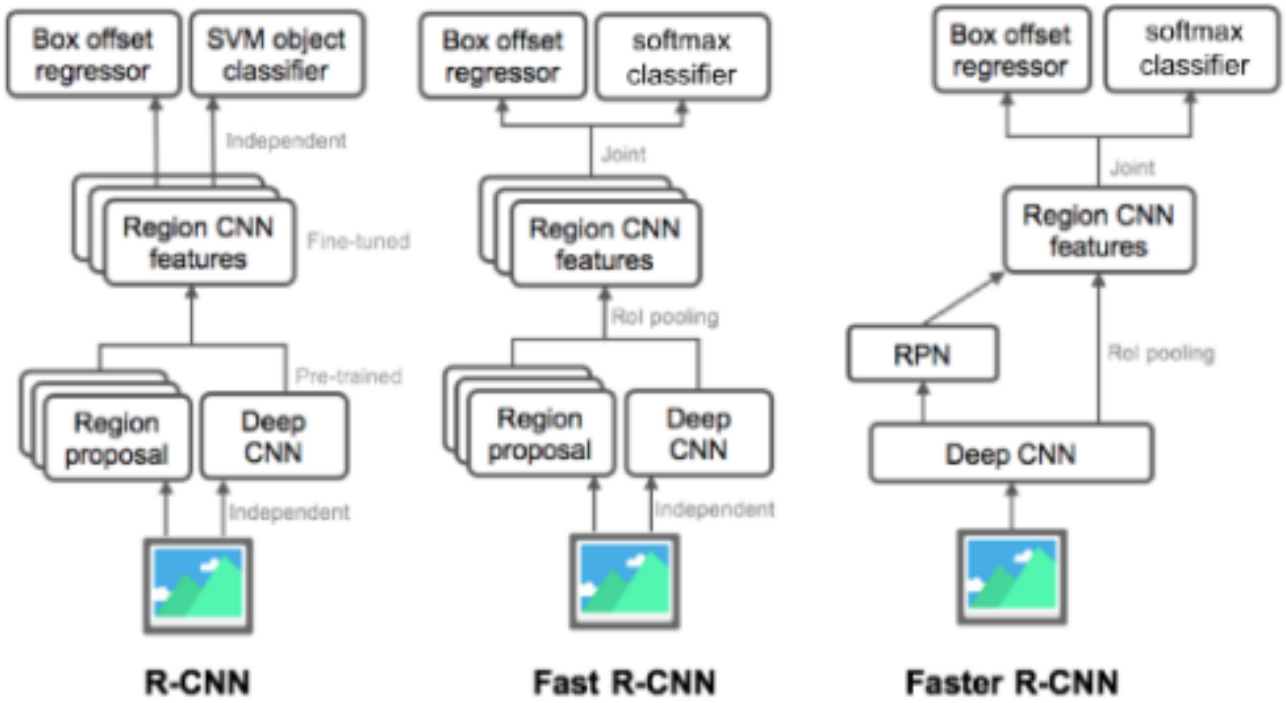
  $\longrightarrow$ fichier d'annotations pour la localisation



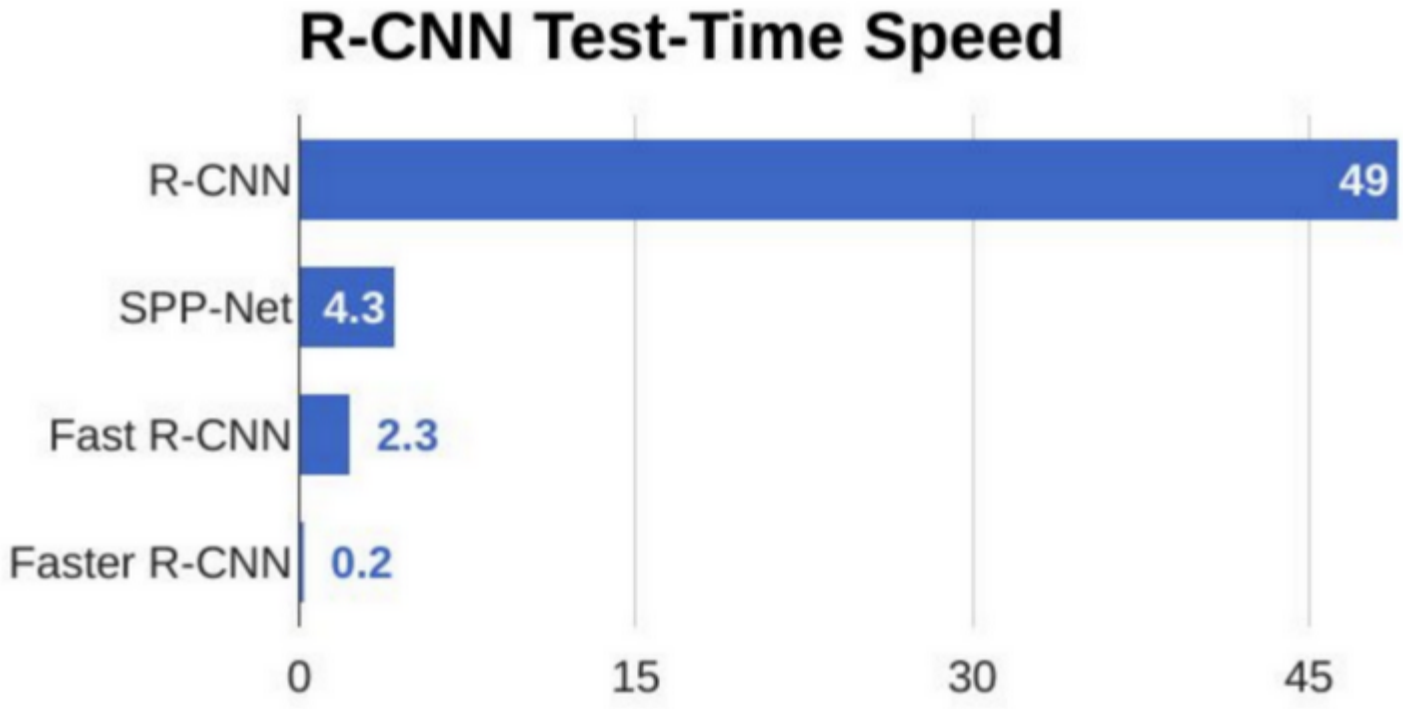
Outline

1. Introduction
2. Le jeu de données
3. Modèles de l'état de l'art à "deux passes"
4. La famille des modèles YOLO
5. Résultats du transfert d'apprentissage
6. Conclusion

- **R-CNN** a été le premier algorithme à appliquer le deep learning à la tâche de détection d'objets.
- Pour proposer des régions d'intérêts, les deux premiers modèles utilisent l'algorithme *Selective Search*.
- **Faster R-CNN** s'affranchit de ce dernier qui est chronophage.

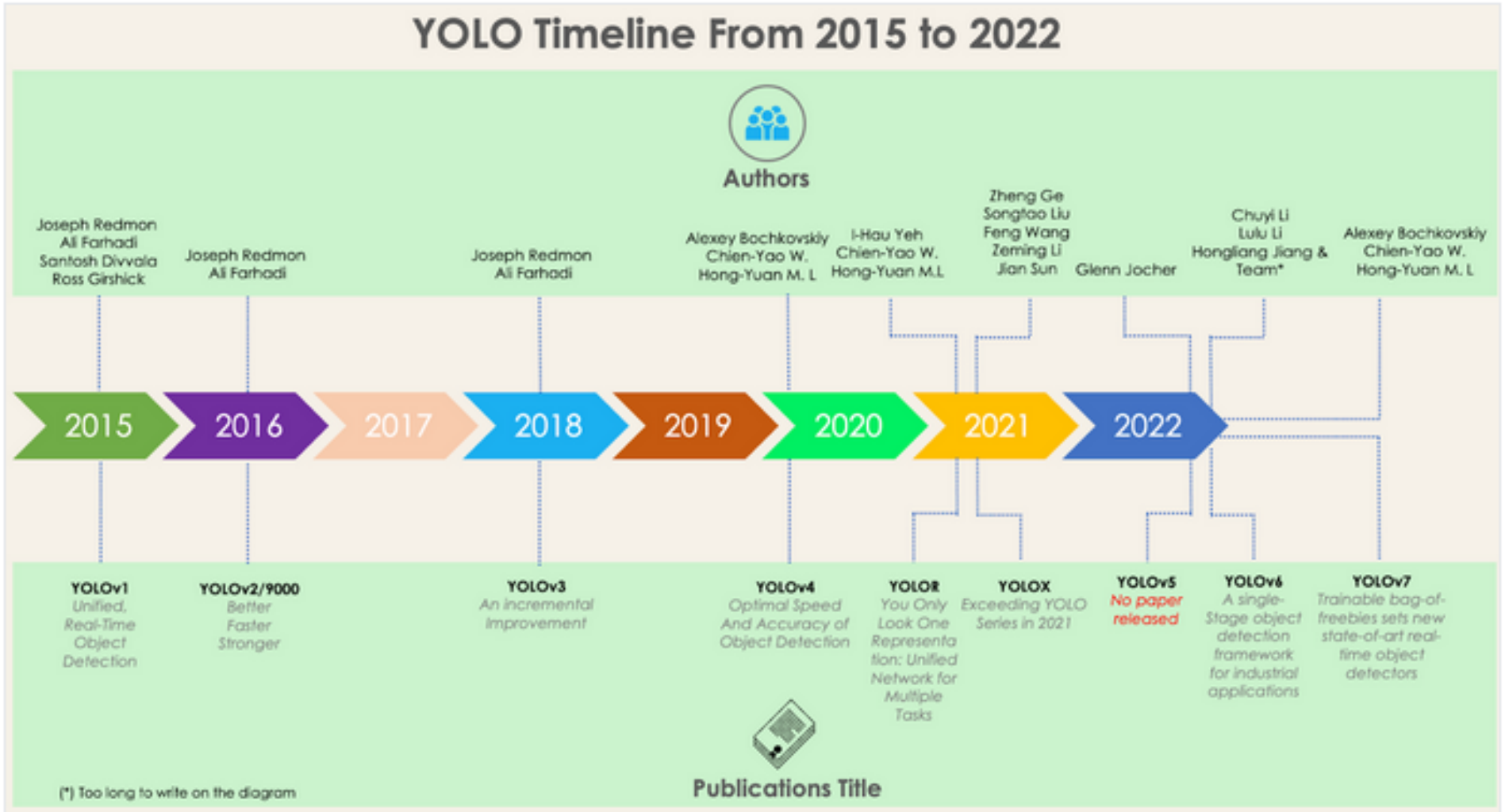


Runtime à l'inférence (en secondes)



Outline

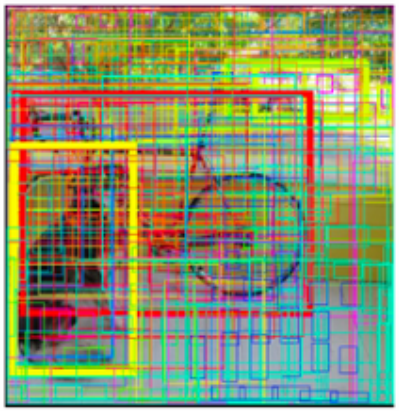
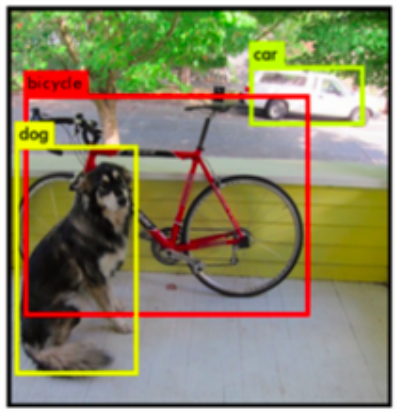
1. Introduction
2. Le jeu de données
3. Modèles de l'état de l'art à "deux passes"
4. La famille des modèles YOLO
5. Résultats du transfert d'apprentissage
6. Conclusion


YOLO Timeline From 2015 to 2022
Authors
Joseph Redmon
Ali Farhadi
Santosh Divvala
Ross Girshick
Joseph Redmon
Ali Farhadi
Joseph Redmon
Ali Farhadi
Alexey Bochkovskiy
Chien-Yao W.
Hong-Yuan M. L
I-Hau Yeh
Chien-Yao W.
Hong-Yuan M.L
Zheng Ge
Songtao Liu
Feng Wang
Zeming Li
Jian Sun
Glenn Jocher
Chuyi Li
Lulu Li
Hongliang Jiang &
Team*
Alexey Bochkovskiy
Chien-Yao W.
Hong-Yuan M. L
2015
2016
2017
2018
2019
2020
2021
2022
YOLOv1
Unified, Real-Time Object Detection
YOLOv2/9000
Better Faster Stronger
YOLOv3
An incremental improvement
YOLOv4
Optimal Speed And Accuracy of Object Detection
YOLOR
You Only Look One Representation: Unified Network for Multiple Tasks
YOLOX
Exceeding YOLO Series in 2021
YOLOv5
No paper released
YOLOv6
A single-Stage object detection framework for industrial applications
YOLOv7
Trainable bag-of-freebies sets new state-of-art real-time object detectors
(*) Too long to write on the diagram
Publications Title

Localisation

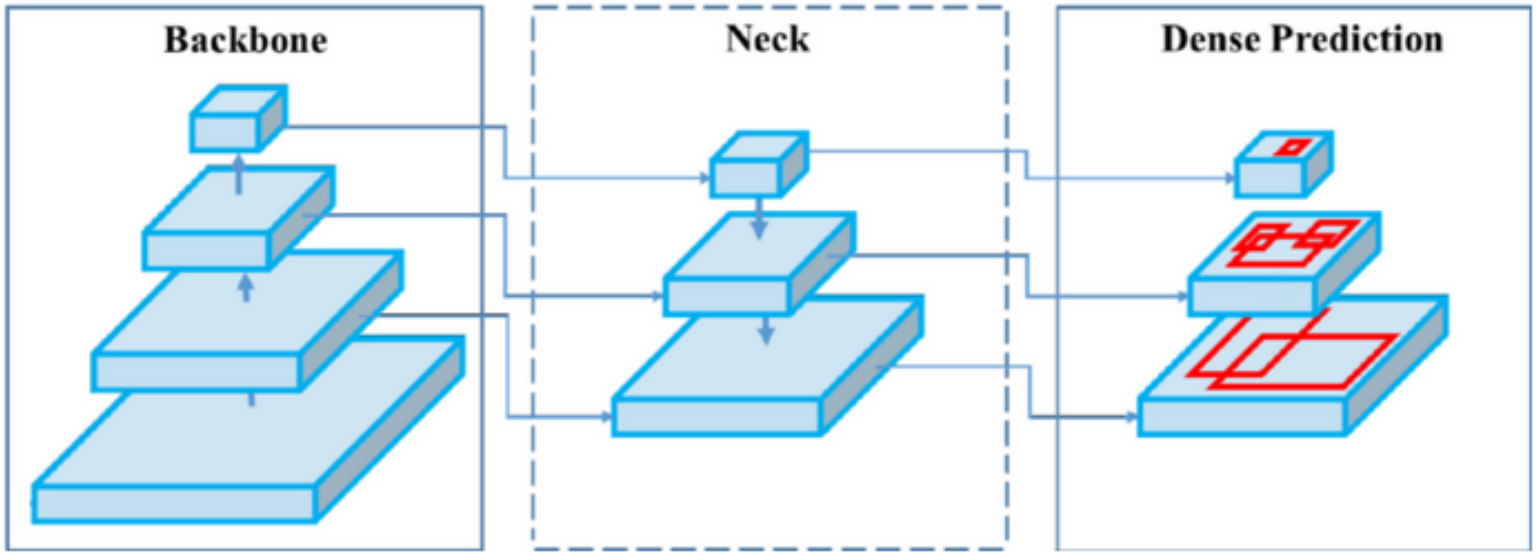
- L'image est découpée en une grille de taille $(S \times S)$
- Chaque cellule permet de générer plusieurs cadrages
- L'IoU est calculée pour regrouper les cadrages qui détectent le même objet
- *Non-max suppression* pour cadrage final



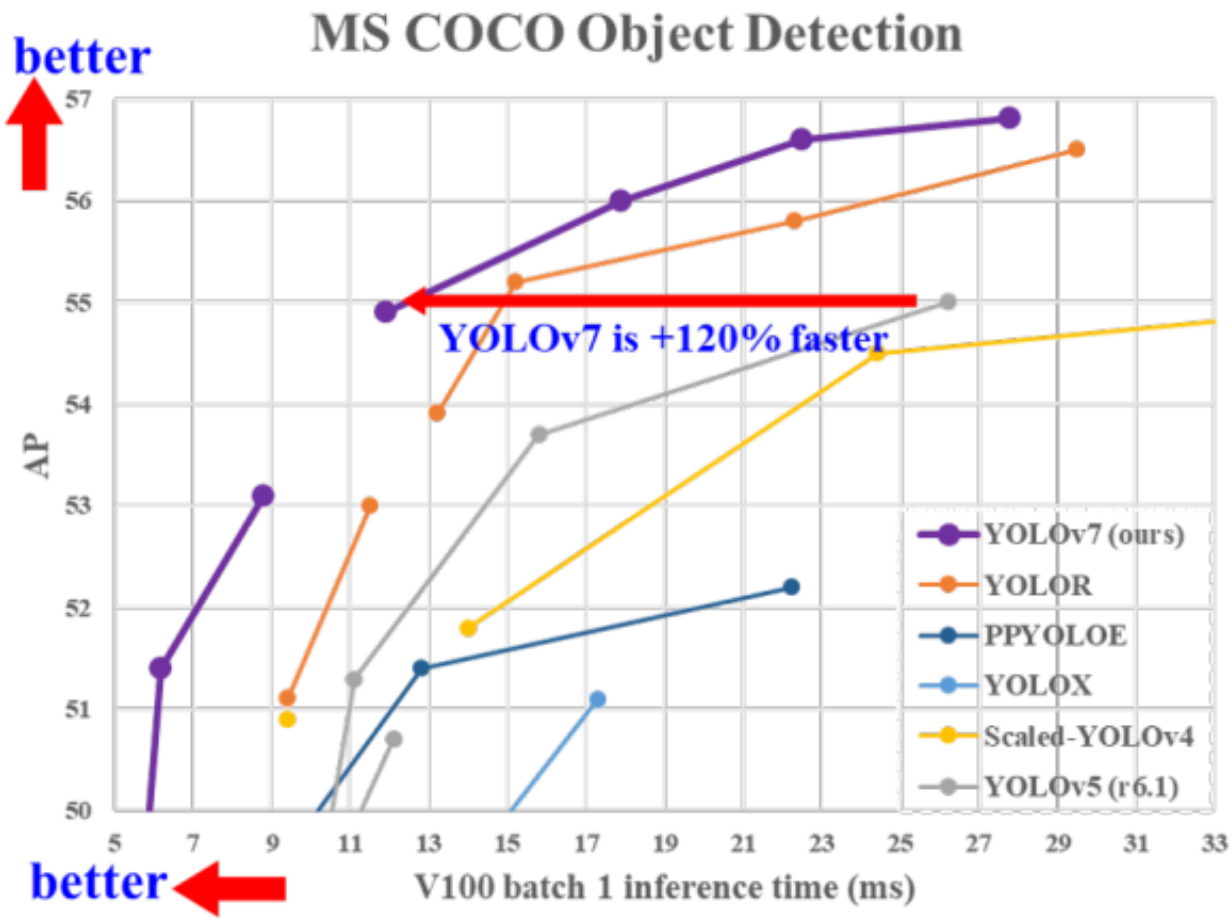



Architecture d'un réseau "une passe"

- L'image en entrée est fournie à la première couche de la **Backbone**, et la partie **Head** ("*Dense Prediction*") retourne les détections sous forme de rectangle d'encrage.
- Le **Neck** a pour rôle d'extraire et combiner des features de différentes résolutions et complexités utiles à notre tâche de détection.



Performances à l'état de l'art



Outline

1 Introduction

2 Le jeu de données

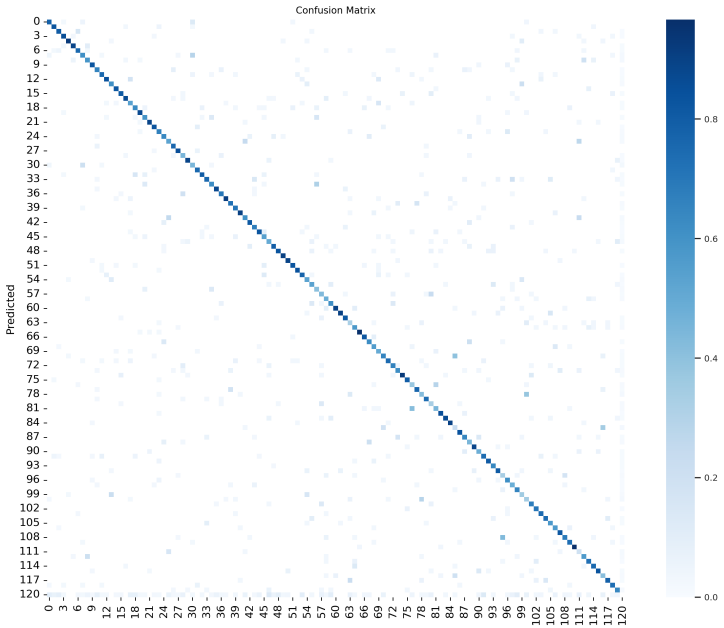
3 Modèles de l'état de l'art à "deux passes"

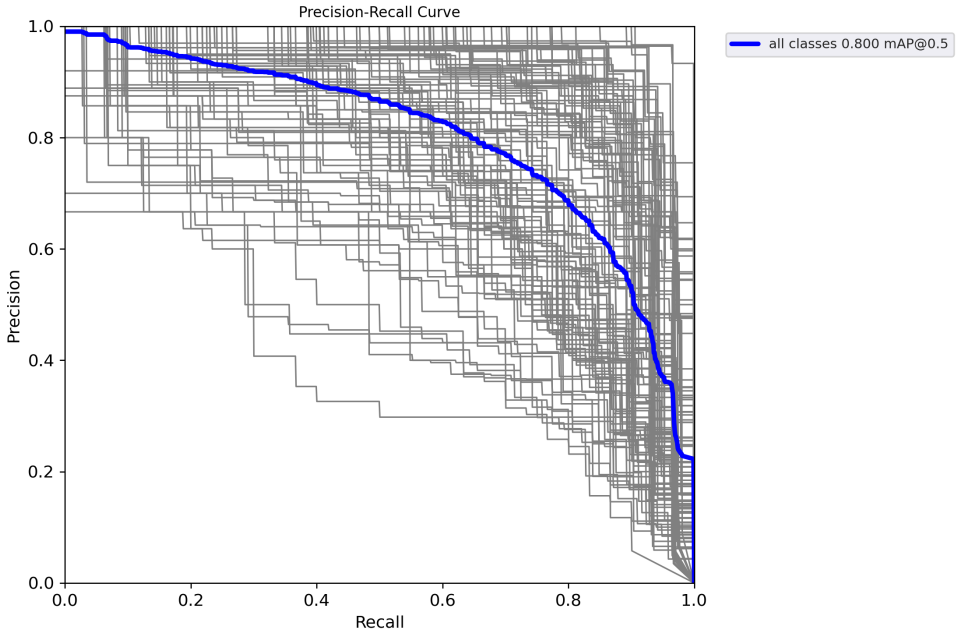
4 La famille des modèles YOLO

5 Résultats du transfert d'apprentissage

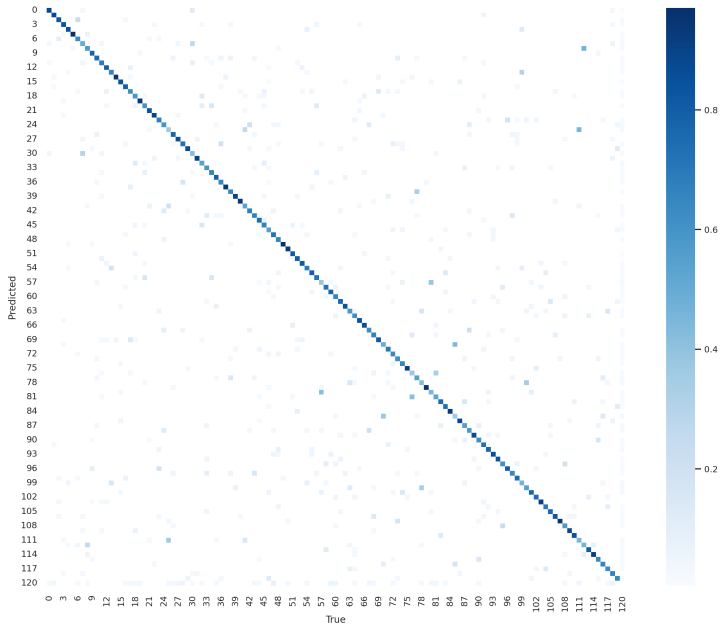
6 Conclusion

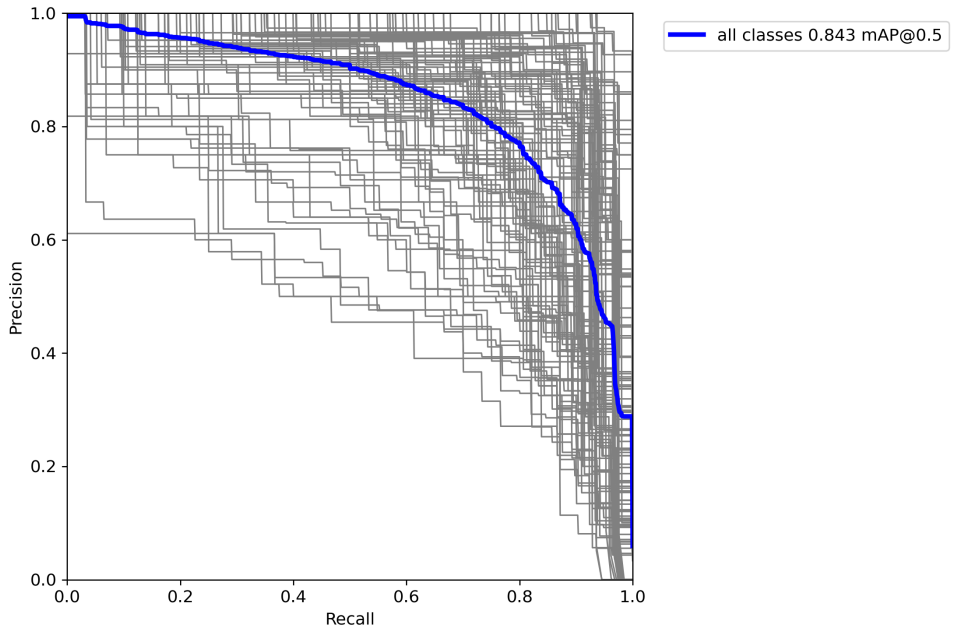
Résultat : YOLOv5





Résultat : YOLOv7





Outline

1 Introduction

2 Le jeu de données

3 Modèles de l'état de l'art à "deux passes"

4 La famille des modèles YOLO

5 Résultats du transfert d'apprentissage

6 Conclusion

Conclusion

Résultats

- YOLO aussi performant que VGG16 ou Xception pour les tâches de classification sur les 120 races.
- Temps de calcul comparable entre les deux versions testées de YOLO : environ 6,5h pour 70 epochs.
- Application réussie sur vidéos.

Merci de votre attention.