

# You Only Look Once (Yolo)

BEX Roméo, RIVALDI Tristan, LAMURE Maxence

24 Septembre 2024

## Contents

<b>0</b>	<b>Introduction</b>	<b>2</b>
<b>1</b>	<b>Système de détection Yolo</b>	<b>2</b>

## 0 Introduction

La détection d'objets est une tâche essentielle en vision par ordinateur, utilisée dans des domaines aussi variés que la conduite autonome, la surveillance, la reconnaissance d'images et les systèmes de sécurité. Ces applications nécessitent non seulement de localiser des objets dans une image, mais aussi de les classer correctement, et cela, avec un temps de traitement minimal pour pouvoir fonctionner en temps réel.

*You Only Look Once (YOLO)* est une approche originale introduite pour résoudre ce problème de détection d'objets. Contrairement aux méthodes traditionnelles qui analysent une image en faisant de la classification, YOLO propose de le faire en utilisant un modèle de régression. Il traite la détection comme un simple problème de régression qui va de l'image entière aux coordonnées des boîtes englobantes (Bounding Boxes) et aux probabilités des classes d'objets, tout cela en une seule fois à travers le réseau de neurones.

Cette approche permet à YOLO d'être rapide tout en conservant une bonne précision. YOLO est capable de traiter des images à des vitesses allant jusqu'à 45 images par seconde, ce qui en fait un très bon candidat pour des applications en temps réel. Cependant, comme tout modèle, YOLO présente également des limites et des compromis, notamment en ce qui concerne la précision de la localisation des petits objets.

Dans ce rapport, nous explorerons l'architecture de YOLO, les principes mathématiques sous-jacents, et nous examinerons ses performances par rapport aux autres approches de détection d'objets. Nous analyserons également ses avantages, ses limites et ses applications dans le monde réel.

## 1 Système de détection Yolo

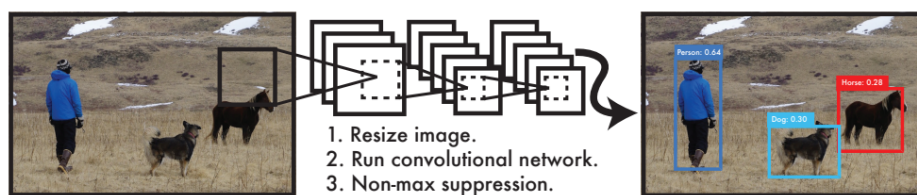


Figure 1: Système de détection Yolo

Comme on peut voir ci-dessus, la reconnaissance d'image est divisée en 3 parties qui sont : - La première étape consiste à faire un **redimensionnement** de l'image d'entrée à une taille fixe de  $448 \times 448$  pixels. Cela permet

d'uniformiser la taille des images, peu importe leurs dimensions d'origine, ce qui facilite le traitement par le réseau de neurones.

- Ensuite, l'image redimensionnée est passée dans un **réseau de neurones convolutionnel (CNN)**. Le CNN traite l'image pour extraire des caractéristiques et prédire plusieurs boîtes englobantes (bounding boxes) et les classes d'objets correspondantes. Le réseau prédit simultanément plusieurs objets dans l'image.

-Après que le modèle a généré les prédictions, la dernière étape est **la suppression non-maximale**. Cette méthode élimine les boîtes englobantes qui se chevauchent et qui représentent le même objet, en ne gardant que la boîte avec le score de confiance le plus élevé. Cela permet d'éviter les multiples détections d'un même objet.