

You Only Look Once (YOLO)

BEX Roméo, RIVALDI Tristan, LAMURE Maxence

Université de Montpellier
Master 2 Statistiques et Sciences des Données

2024 - 2025



STATISTIQUE
SCIENCE DES DONNÉES MIND
UNIVERSITÉ DE MONTELLIER

Table des matières

Introduction

Méthodologie

Structure du réseau

Comparaison des Méthodes

Limites et Faiblesses

Conclusion

Introduction

La **classification d'images** consiste à associer des étiquettes à une image en fonction de son contenu visuel. C'est un problème clé de la *vision par ordinateur*, souvent résolu via l'apprentissage supervisé. Applications courantes :

- ▶ **Reconnaissance faciale**
- ▶ **Voitures autonomes**

Exemple d'image



Figure: Image d'entrée.

Image avec Boîtes Englobantes

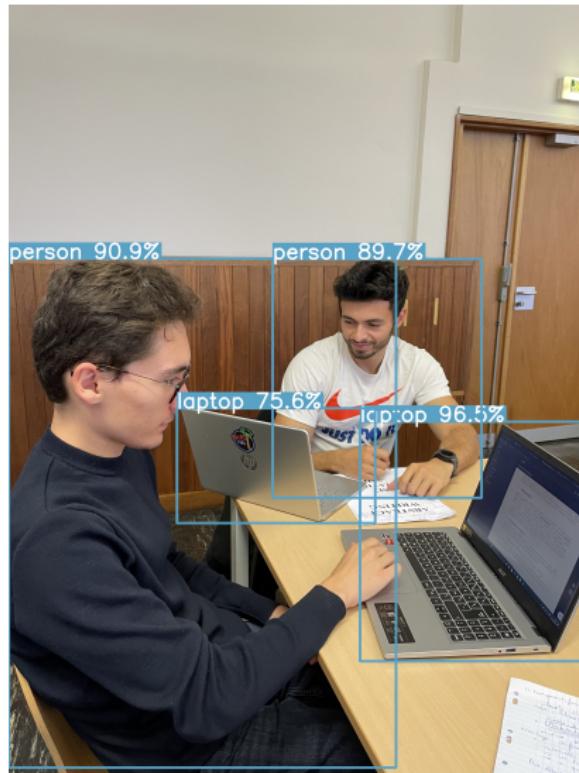


Figure: Image avec boîtes englobantes prédites par YOLOv8.

Formulation du problème

- ▶ YOLO redimensionne puis divise l'image en une grille $S \times S$.

Formulation du problème

- ▶ YOLO redimensionne puis divise l'image en une grille $S \times S$.
- ▶ Chaque cellule prédit des **boîtes englobantes** et des **probabilités de classes**.

Formulation du problème

- ▶ YOLO redimensionne puis divise l'image en une grille $S \times S$.
- ▶ Chaque cellule prédit des **boîtes englobantes** et des **probabilités de classes**.
- ▶ YOLO utilise une **régression directe**, prédisant en une seule étape les coordonnées des objets et leurs classes, ce qui permet un traitement rapide.

Structure du réseau YOLO

- ▶ Basé sur un réseau de convolution profond (24 couches + 2 couches entièrement connectées).

Structure du réseau YOLO

- ▶ Basé sur un réseau de convolution profond (24 couches + 2 couches entièrement connectées).
- ▶ Sortie : tenseur de taille $S \times S \times (B \times 5 + C)$. ($C = 80$ avec notre modèle)

Structure du réseau YOLO

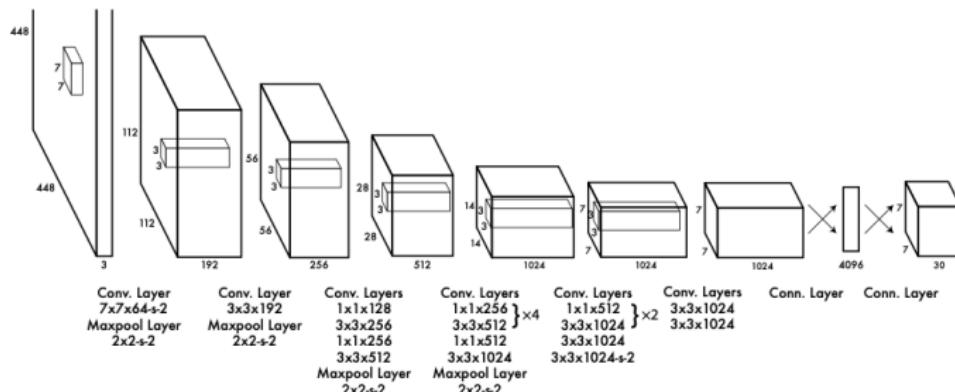
- ▶ Basé sur un réseau de convolution profond (24 couches + 2 couches entièrement connectées).
- ▶ Sortie : tenseur de taille $S \times S \times (B \times 5 + C)$. ($C = 80$ avec notre modèle)
- ▶ B : nombre de boîtes englobantes que chaque cellule de la grille prédit.

Structure du réseau YOLO

- ▶ Basé sur un réseau de convolution profond (24 couches + 2 couches entièrement connectées).
- ▶ Sortie : tenseur de taille $S \times S \times (B \times 5 + C)$. ($C = 80$ avec notre modèle)
- ▶ B : nombre de boîtes englobantes que chaque cellule de la grille prédit.
- ▶ 5 : Chaque boîte englobante est décrite par 5 paramètres : $(x, y, w, h, \text{score de confiance})$.

Structure du réseau YOLO

- ▶ Basé sur un réseau de convolution profond (24 couches + 2 couches entièrement connectées).
- ▶ Sortie : tenseur de taille $S \times S \times (B \times 5 + C)$. ($C = 80$ avec notre modèle)
- ▶ B : nombre de boîtes englobantes que chaque cellule de la grille prédit.
- ▶ 5 : Chaque boîte englobante est décrite par 5 paramètres : $(x, y, w, h, \text{score de confiance})$.



Fonction de perte

La fonction de perte de YOLO prend en compte les erreurs de localisation et de classification. Elle utilise deux hyperparamètres λ_{coord} et λ_{noobj} pour pondérer les différentes parties de la perte :

Fonction de perte

La fonction de perte de YOLO prend en compte les erreurs de localisation et de classification. Elle utilise deux hyperparamètres λ_{coord} et λ_{noobj} pour pondérer les différentes parties de la perte :

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}_{\text{Erreur de localisation}} \quad (1)$$

Fonction de perte

La fonction de perte de YOLO prend en compte les erreurs de localisation et de classification. Elle utilise deux hyperparamètres λ_{coord} et λ_{noobj} pour pondérer les différentes parties de la perte :

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}_{\text{Erreur de localisation}} \quad (1)$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{\left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]}_{\text{Erreur de taille (largeur/hauteur des boîtes)}} \quad (2)$$

Fonction de perte

La fonction de perte de YOLO prend en compte les erreurs de localisation et de classification. Elle utilise deux hyperparamètres λ_{coord} et λ_{noobj} pour pondérer les différentes parties de la perte :

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}_{\text{Erreur de localisation}} \quad (1)$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{\left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]}_{\text{Erreur de taille (largeur/hauteur des boîtes)}} \quad (2)$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{(C_i - \hat{C}_i)^2}_{\text{Erreur avec objet}} + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{noobj}} \underbrace{(C_i - \hat{C}_i)^2}_{\text{Erreur sans objet}} \quad (3)$$

Fonction de perte

La fonction de perte de YOLO prend en compte les erreurs de localisation et de classification. Elle utilise deux hyperparamètres λ_{coord} et λ_{noobj} pour pondérer les différentes parties de la perte :

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]}_{\text{Erreur de localisation}} \quad (1)$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{\left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]}_{\text{Erreur de taille (largeur/hauteur des boîtes)}} \quad (2)$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \underbrace{(C_i - \hat{C}_i)^2}_{\text{Erreur avec objet}} + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{noobj}} \underbrace{(C_i - \hat{C}_i)^2}_{\text{Erreur sans objet}} \quad (3)$$

Suite

$$+ \sum_{i=0}^{S^2} \mathbf{1}_i^{\text{obj}} \sum_{c \in \text{classes}} \underbrace{(p_i(c) - \hat{p}_i(c))^2}_{\text{Erreur de classification}} \quad (4)$$

Comparaison avec d'autres méthodes

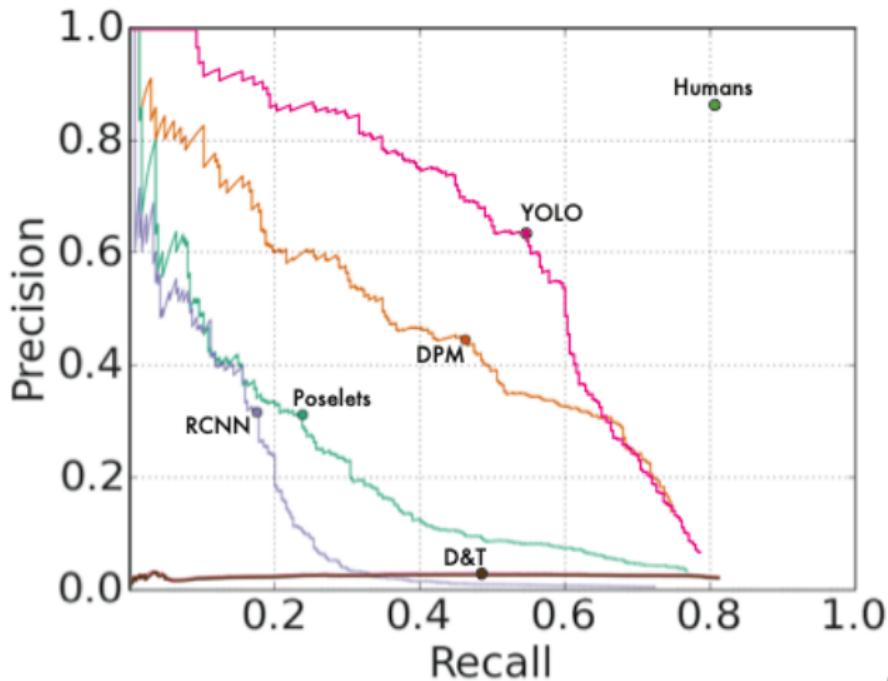


Figure: Courbes précision-rappel comparant YOLO avec d'autres méthodes.

Limites et Faiblesses

- ▶ Problèmes de localisation pour les petits objets.

Limites et Faiblesses

- ▶ Problèmes de localisation pour les petits objets.
- ▶ Sensibilité aux petites boîtes, impactant l'IoU (précision des boîtes).

Limites et Faiblesses

- ▶ Problèmes de localisation pour les petits objets.
- ▶ Sensibilité aux petites boîtes, impactant l'IoU (précision des boîtes).



Figure: Limites de YOLOv8 sur une scène complexe.

Limites et Faiblesses - Exemple Supplémentaire

- ▶ Difficulté à généraliser à des objets aux proportions inhabituelles.

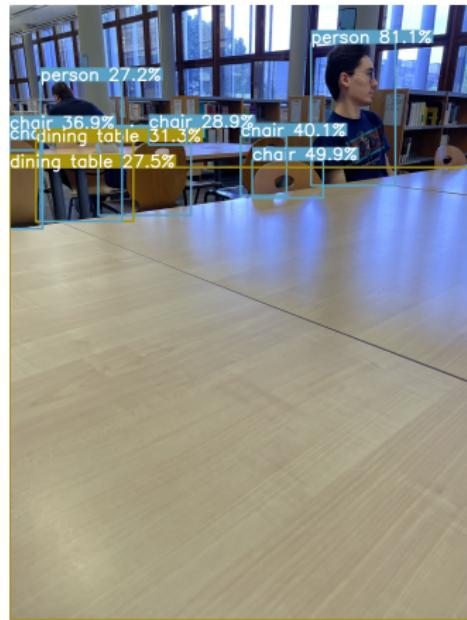
Limites et Faiblesses - Exemple Supplémentaire

- ▶ Difficulté à généraliser à des objets aux proportions inhabituelles.



Limites et Faiblesses - Exemple Supplémentaire

- ▶ Difficulté à généraliser à des objets aux proportions inhabituelles.



Conclusion

- ▶ YOLO simplifie la détection d'objets en intégrant toutes les étapes en une seule.
- ▶ Avantages : rapidité, simplicité d'utilisation.
- ▶ Limites : localisation des petits objets, boîtes noires
- ▶ Applications : robotique, surveillance,...

Questions ?

Merci pour votre attention !
Avez-vous des questions ?