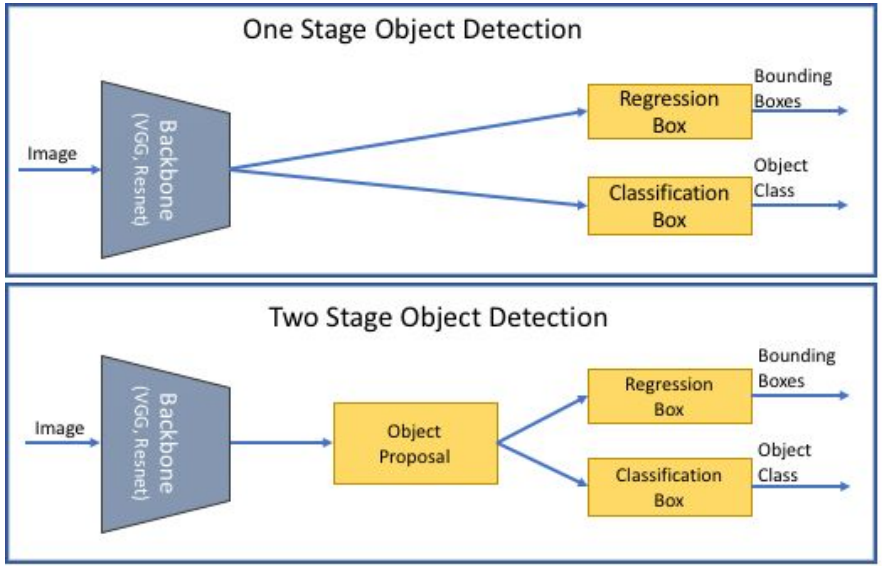
Adaptation YoloV1&2 From scratch

PascalVoc To PlantDoc

ARNOUDTS Kevin - BRZYCHCY Loïc - DARVILLE Killian

Pourquoi Yolo ?



Une phase :

- Traitement en temps réel (45-150FPS)
- Meilleur équilibre précision/vitesse

Deux phases :

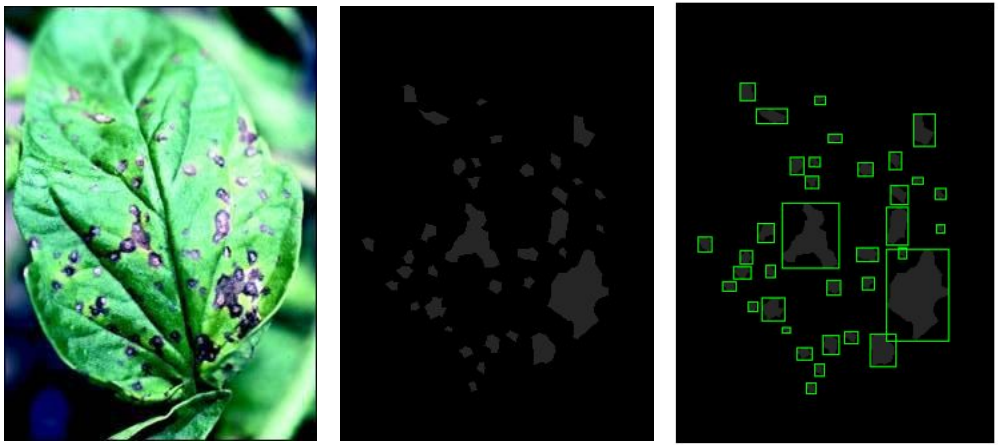
- Faible vitesse d'inférence (0-7FPS)
- Consommation excessive de ressources

Dataset

	Nombre d’images	Nombre de classes	Objets	Annotations
VOC 2007	9.963	20	Véhicules, animaux, objets et personnes	Images + Masks + Bounding Box (XML)
VOC 2012	11.530	20	Même catégories que 2007	Images + Masks + Bounding Box (XML)
PlantDoc	588	1	Maladies	Images + Masks

Adaptation PlantDoc

1. Génération de BoundingBox à partir des masks et enregistrement dans un fichier CSV

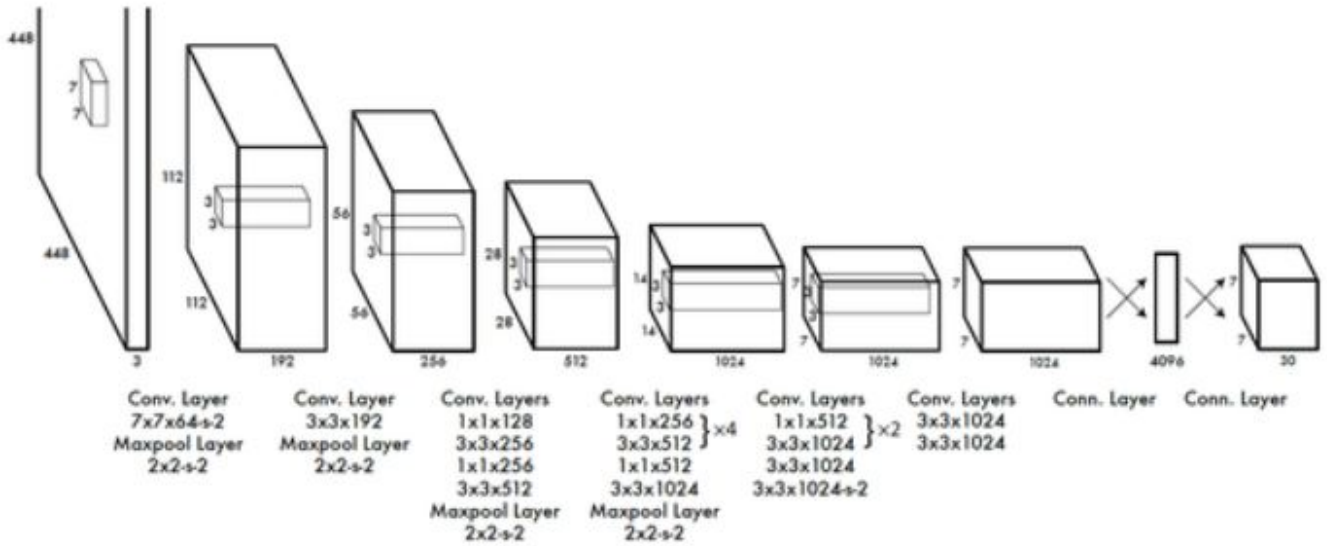


```
image_id,width,height,bbox,source
00004.jpg,2160,1440,"[1917.0, 1360.0, 159.0, 77.0]",disease
00004.jpg,2160,1440,"[1658.0, 939.0, 204.0, 215.0]",disease
00004.jpg,2160,1440,"[1947.0, 888.0, 81.0, 85.0]",disease
```

2. Transformation CSV en XML


```
▼ object {1}
  ▼ annotation {5}
      folder : images
      filename : 00004.png
    ▼ size {3}
        width : 2160
        height : 1440
        depth : 3
      segmented : 0
    ▼ object [18]
      ▼ 0 {6}
          name : disease
          pose : Unspecified
          truncated : 0
          occluded : 0
          difficult : 0
        ▼ bndbox {4}
            xmin : 1917
            ymin : 1360
            xmax : 2076
            ymax : 1437
      ▶ 1 {6}
      ▶ 2 {6}
      ▶ 3 {6}
      ▶ 4 {6}
```

YoloV1 - Architecture



YoloV1 - Modification couche sortie

YoloV1 génère une sortie sous forme d'un tenseur (S,S,(B*5)+C):

- S,S représente la taille de la grille, étant de 7x7.
- B représente le nombre d'ancrage par cellule.
- C représente le nombre de classes.

Ainsi, sur PascalVOC, le tensor était de forme (7,7,30).

Nous avons dû le modifier en (7,7,6).

YoloV1 - Modification fonction de perte.

Fonction de perte présente dans le modèle: Cross entropy (utilisé pour PascalVOC afin de gérer ses 20 classes).

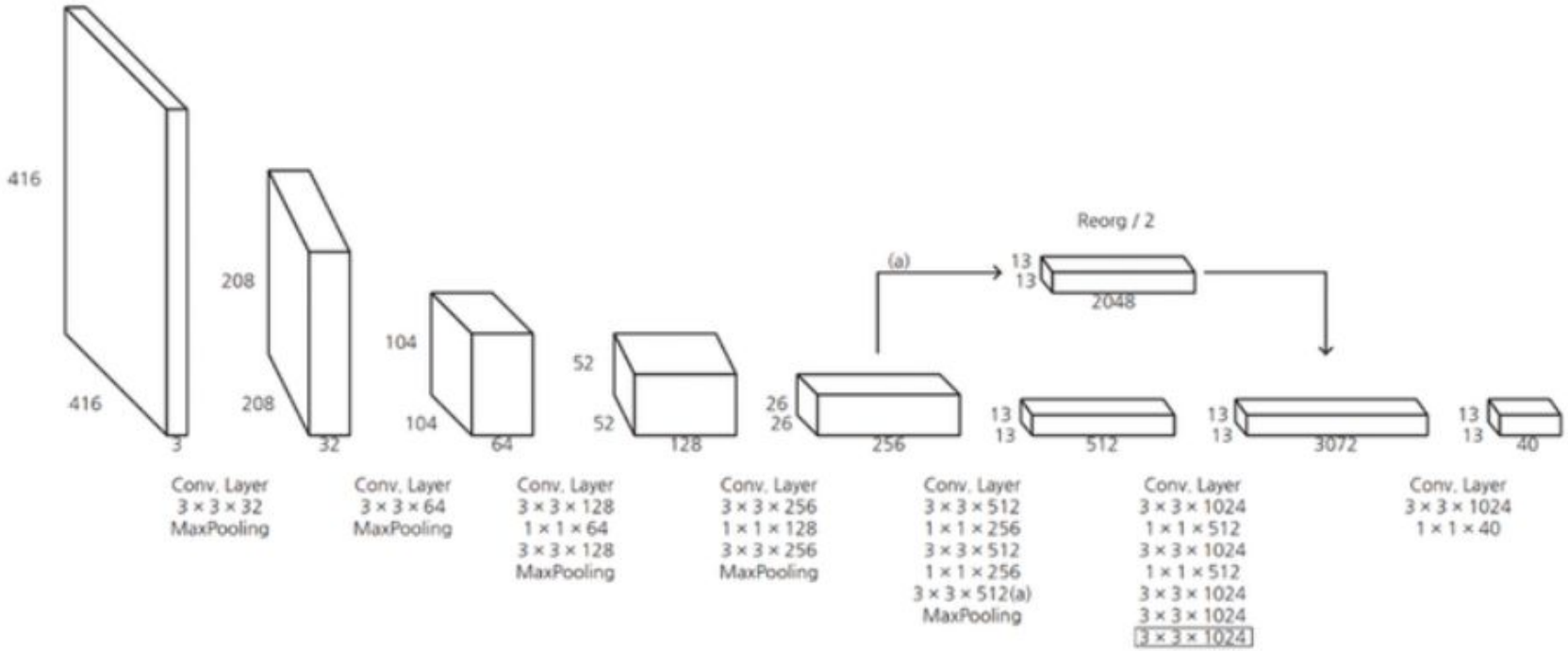
PlantDoc n’ayant qu’une classe, une fonction de perte binaire serait l’idéal.

- Problème avec l’ancienne version tensorflow.
- Cross entropy de taille 1 pour simuler du binaire.

YoloV1 - Autres modifications

- Réduction du learning rate car plus petit dataset.
  - Après 40 epochs seulement, grâce au learning rate scheduler.
- Vérification taille des images (448x448).

YoloV2 - Architecture



YoloV2 - Adaptation

Identique à Yolo1 sauf, pour une exception, qui est la nouveauté de YoloV2:

Les Anchors Boxes.

Nécessite un pré-calcul de la distribution de la taille des boites du jeu d'entraînement.

Réaliser avec K-Means. Permettant de trouver un certain nombre de groupes représentatifs des tailles des boîtes.

La distance utilisé est l’Intersection over Union.

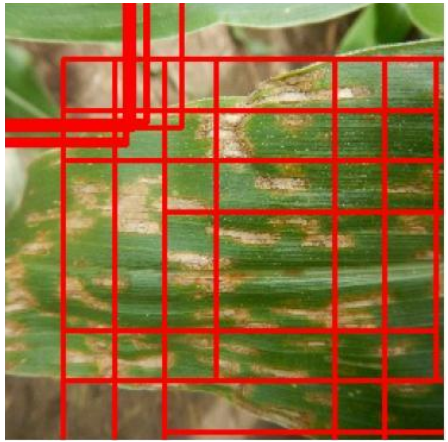
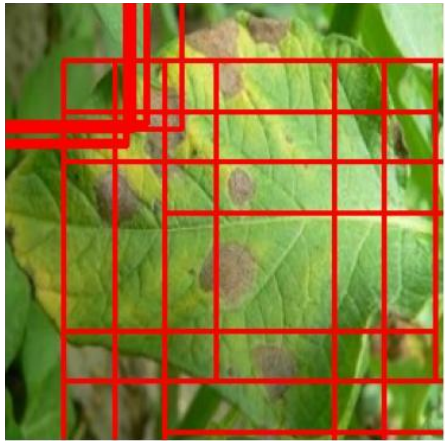
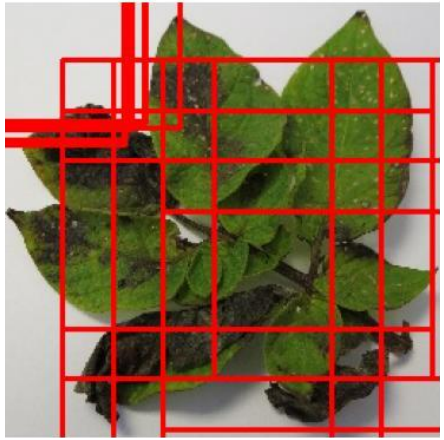
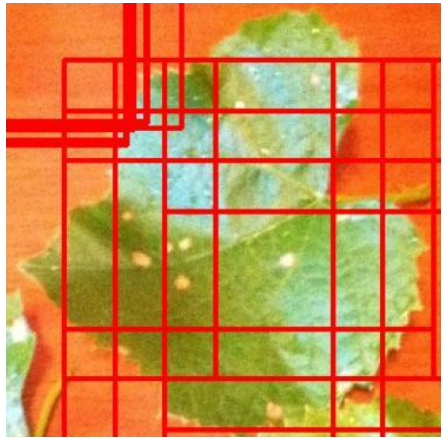
Comparatif YOLOv1 vs YOLOv2

	Yolov1	Yolov2
Architecture	24 couches convolutives + 2 FC (inspiré de GoogLeNet)	19 couches convolutives + 5 max-pooling (inspiré de Darknet-19)
Taille d'entrée	448 × 448 pixels	Multi-scale (varie entre 320 et 608 px)
Méthode d’ancrage	Grille 7 x 7	Anchor boxes (K-means IoU)
Détection en multi-classes	Une seule classe par cellule	Plusieurs classes possibles pour une même cellule
Gestion des petits objets	Mauvaise	Meilleure (multi-scale + anchor boxes)

Résultats

- Problème de loss, pas d’adaptation fournissant de bons résultats.
- Sortie toujours identique.
- Box prédites “aléatoires”.

Exemples de prédictions

Résultats théoriques

Detection Frameworks	Train	mAP	FPS
Fast R-CNN [5]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[15]	2007+2012	73.2	7
Faster R-CNN ResNet[6]	2007+2012	76.4	5
YOLO [14]	2007+2012	63.4	45
SSD300 [11]	2007+2012	74.3	46
SSD500 [11]	2007+2012	76.8	19
YOLOv2 288×288	2007+2012	69.0	91
YOLOv2 352×352	2007+2012	73.7	81
YOLOv2 416×416	2007+2012	76.8	67
YOLOv2 480×480	2007+2012	77.8	59
YOLOv2 544×544	2007+2012	**78.6**	40

- Nette amélioration des performances de YOLOv2
- Polyvalence : Différentes résolution sans réentraînement

Des questions ?