



Plan Prévisionnel

Projet 10

Développez une preuve de concept

Introduction

Dans ce projet, une veille technologique sera réalisée afin d'identifier une méthode récente visant à améliorer les performances du modèle. Parmi les trois options A, B et C pour le choix du jeu de données et du modèle, nous retenons l'**Option A**, qui autorise de s'appuyer sur l'un des projets du parcours OpenClassrooms. Notre sélection porte sur le projet 8: « Participation à la conception d'un système de voiture autonome – segmentation d'images ».

1. Dataset retenu

Pour cette preuve de concept, nous utiliserons un jeu de données d'images urbaines annotées, calqué sur la structure du dataset **Cityscapes**, référence dans le domaine de la perception pour véhicules autonomes.

Contenu: Images de scènes de rue (routes, véhicules, piétons, bâtiments) capturées du point de vue conducteur.

Traitement : Afin d'assurer la robustesse du modèle et de répondre aux contraintes temps réel, les 34 classes originales seront simplifiées en **8 catégories fonctionnelles** (Flat, Human, Vehicle, Construction, Object, Nature, Sky, Void).

Préparation : Les données subiront des augmentations via la librairie `Albumentations` (redimensionnement 256x256, normalisation, et technique de régularisation **CutMix**) pour améliorer la généralisation du modèle.

2. Modèle envisagé

Algorithme sélectionné : YOLO (You Only Look Once) - Segmentation

Nous avons choisi de mettre en œuvre les architectures récentes de **YOLO (versions v8, v9, v11)** adaptées à la segmentation.

Justification de la performance :

Le choix de YOLO repose sur la nécessité de résoudre le compromis **Précision / Latence** critique pour les systèmes embarqués :

Les architectures classiques de segmentation sémantique (comme U-Net) offrent une excellente précision mais souffrent d'une lourdeur computationnelle entraînant une latence élevée.

YOLO, historiquement conçu pour la détection d'objets ultra-rapide, a récemment intégré des têtes de segmentation. L'hypothèse est que YOLO peut offrir une précision (mIoU) compétitive par rapport à U-Net tout en divisant drastiquement le temps d'inférence, permettant une réactivité de l'ordre de la milliseconde nécessaire à la sécurité routière.

Objectif et Contexte :

L'objectif consiste à permettre à un véhicule autonome de « voir » et de réaliser une sémantisation pixel-par-pixel de son environnement en temps réel. Ce module de perception est fondamental pour les décisions opérationnelles (freinage, évitement) dans un contexte de conduite réelle. Une comparaison entre plusieurs approches sera menée, allant d'une segmentation classique (**Mini U-Net** avec encodeur **VGG16**¹) à des architectures de détection/segmentation en temps réel de pointe (familles **YOLOv8**, **YOLOv9** et **YOLOv11**).

3. Références bibliographiques

Notre étude s'appuie sur l'état de l'art suivant :

1. U-Net: Convolutional Networks for Biomedical Image Segmentation (Ronneberger et al., 2015)
Rôle : Définit l'architecture "encoder-decoder" qui sert de baseline académique pour notre comparaison.
[https://arxiv.org/pdf/1505.04597](https://arxiv.org/pdf/1505.04597.pdf)
2. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information (Wang et al., 2024) & Ultralytics YOLO Docs
Rôle : Référence pour les architectures "One-Stage" les plus récentes. YOLOv9 introduit des concepts comme le PGI pour améliorer la précision, tandis que la documentation Ultralytics détaille les implémentations de segmentation pour v8 et v11 utilisées dans ce projet.
[https://arxiv.org/pdf/2402.13616](https://arxiv.org/pdf/2402.13616.pdf) (YOLOv9) / Lien GitHub Ultralytics
3. Yaseen, M. (2024). What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector. arXiv preprint arXiv: [https://arxiv.org/pdf/2409.07813](https://arxiv.org/pdf/2409.07813.pdf)
4. Jahan, M. K., Bhuiyan, F. I., Amin, A., Mridha, M. F., Safran, M., Alfarhood, S., & Che, D. (2025). Enhancing the YOLOv8 model for realtime object detection to ensure online platform safety. *Scientific Reports*, 15(1), 21167. <https://www.nature.com/articles/s41598-025-08413-4>
5. The Cityscapes Dataset for Semantic Urban Scene Understanding (Cordts et al., 2016)
Rôle Référence pour la structure des données et les métriques d'évaluation en environnement urbain. [https://arxiv.org/pdf/1604.01685](https://arxiv.org/pdf/1604.01685.pdf)

¹ Réseau de neurones convolutionnel profond (16 couches apprenantes) développé par VGG.

Architecture avec blocs convolutifs 3x3 et max-pooling, suivis de couches denses et d'une sortie softmax (ImageNet 1000 classes).

4. Explication de la démarche de test (Preuve de Concept)

Notre démarche consiste à mettre en compétition une approche classique et une approche temps réel via un pipeline automatisé.

Méthode Baseline :

Nous entraînerons un modèle **Mini-Unet** avec un encodeur **VGG16** (pré-entraîné ImageNet). Ce modèle servira de référence pour la qualité de segmentation (haute précision attendue, mais vitesse lente).

Méthode Cible (Challenger) :

Nous entraînerons plusieurs variantes de YOLO (Nano, Small, Compact) sur le même jeu de données.

5. Protocole d'évaluation :

Les modèles seront évalués sur un jeu de validation (20% du dataset) selon deux axes principaux :

1. **Métriques Quantitatives** : Comparaison du score *Mean IoU* (qualité du masque) et du **Temps d'inférence** (ms/image sur GPU).
2. **Analyse Qualitative** : Génération visuelle des masques de segmentation et analyse des matrices de confusion pour valider la pertinence opérationnelle du "meilleur modèle" retenu.