

Plan Prévisionnel - Preuve de Concept

Segmentation Sémantique

Projet : Amélioration de la segmentation sémantique urbaine avec SegFormer

Dataset Retenu

Cityscapes : Dataset de référence pour la segmentation sémantique de scènes urbaines comprenant 2975 images haute résolution annotées pixel par pixel. Les images sont capturées dans **50 villes allemandes différentes** avec **34 classes principales** que nous regroupons en **8 catégories pertinentes** (flat, vehicle, human, construction, object, nature, sky, void). La résolution sera progressivement testée à 512×512 puis 768×768 pixels, contre 224×224 dans le projet précédent, pour exploiter pleinement les capacités des architectures modernes.

Modèle Envisagé

Algorithme : SegFormer

Justifications du choix :

SegFormer (Xie et al., 2021) combine l'efficacité des Vision Transformers avec un décodeur MLP simple, atteignant des performances état de l'art sur Cityscapes (**84.0% mIoU avec SegFormer-B5** contre 80.9% pour DeepLabv3+ ResNet-101). Cette architecture résout les limitations des CNN par son contexte global natif et évite les problèmes des Transformers classiques grâce à l'absence d'encodage positionnel fixe. Comparé à notre baseline actuelle FPN + EfficientNetB0 (74.6% mIoU @224x224), SegFormer présente un **potentiel d'amélioration de ~10 points mIoU**, particulièrement sur les petits objets critiques pour les classes Human et Object.

Note : les résultats de référence sur l'article utilisent parfois des techniques d'inférence sliding window, différentes de notre protocole à résolution fixe 768×768.

Objectif et contexte :

SegFormer vise la compréhension précise pixel-wise des scènes urbaines. Son architecture hiérarchique capture simultanément détails fins et contexte global, essentiel pour la segmentation

d'éléments critiques (piétons, véhicules, signalétique). Le modèle s'applique aux systèmes de perception embarqués nécessitant un équilibre performance/efficacité computationnelle.

Références Bibliographiques

1. Xie, E., Wang, W., Yu, Z., et al. (2021). *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. NeurIPS 2021. [Article de recherche principal]
2. **PyTorch Semantic Segmentation Implementation Guide** - Towards Data Science (2023). *Modern Semantic Segmentation Architectures Comparison*. [Guide pratique d'implémentation]
3. Cordts, M., Omran, M., Ramos, S., et al. (2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding*. CVPR 2016. [Dataset de référence]

Explication de la Démarche de Test

Méthode Baseline

FPN + EfficientNetB0/ResNet34 : Réentraînement des modèles précédents (initialement 74.6% mIoU en 224×224) aux nouvelles résolutions 512×512 et 768×768.

Migration framework obligatoire : passage de `segmentation_models` (TensorFlow/Keras) vers `segmentation_models_pytorch` pour accéder aux architectures récentes.

Méthode de Test SegFormer

Évaluation comparative progressive :

1. Test des variantes SegFormer-B0, B1 selon contraintes mémoire GPU
2. Comparaison métriques : Mean IoU, accuracy pixel-wise, analyse par classe
3. Mesure des temps d'inférence pour validation pratique
4. Sélection des **2 meilleurs modèles par famille** (FPN vs SegFormer)

Preuve de Concept

Interface Streamlit déployée sur serveur NAS personnel (CPU N100) permettant :

- Upload d'images urbaines avec **double inférence simultanée** (FPN + SegFormer)

- Comparaison visuelle des masques de segmentation
- Affichage temps d'inférence et métriques de performance
- Téléchargement des modèles optimaux pour usage pratique

Code réutilisé : Framework `segmentation_models_pytorch` pour les implémentations. Application originale par le dataset Cityscapes en hautes résolutions et la comparaison architecturale CNN vs Transformer sur notre cas d'usage spécifique.