

# Plan de Travail Prévisionnel

## Dataset retenu

Pour cette preuve de concept, j'ai choisi d'utiliser le jeu de données Cityscapes, une référence dans la segmentation sémantique urbaine. Il se compose d'environ 5 000 images haute résolution (2048×1024) capturées dans des villes européennes à partir de caméras embarquées sur des véhicules. Les annotations sont fournies sous forme de masques pixel à pixel.

## Modèle envisagé

Le modèle retenu est SegFormer B5, une architecture de segmentation sémantique introduite par NVIDIA en 2021.

Contrairement aux modèles traditionnels basés sur les CNN (comme U-Net), SegFormer repose sur une architecture Transformer hiérarchique (Mix Transformer, MiT), qui offre :

- Une excellente capacité à capturer les dépendances globales,
- Une meilleure efficacité mémoire (grâce à un encodeur hiérarchique et un décodeur simplifié),
- Des performances de pointe sur Cityscapes et d'autres benchmarks (mIoU, Dice, etc.).

Objectif du modèle :

Permettre une segmentation urbaine précise sur des systèmes embarqués, en conservant un compromis entre vitesse, précision et légèreté, ce qui le rend pertinent pour la vision embarquée dans des véhicules autonomes.

## Références bibliographiques

1. Article scientifique :

- <https://arxiv.org/abs/2105.15203>

2. Article de blog :

- <https://medium.com/geekculture/semantic-segmentation-with-segformer-2501543d2be4>

3. Benchmark & Application :

- <https://debuggercafe.com/segformer-for-semantic-segmentation>

## Explication de la démarche de test du nouvel algorithme (preuve de concept)

Pour tester ce nouvel algorithme, la démarche repose sur la comparaison équitable avec un modèle baseline :

- Modèle baseline : U-Net avec backbone EfficientNetB3 (modèle issu du projet précédent).
- Nouveau modèle : SegFormer B5 (modèle Transformer récent).

Conditions expérimentales strictement identiques :

- Dataset réduit et homogène
- Taille des images : 256x256
- Métriques : Accuracy, Dice, mIoU, Loss, Temps d'entraînement

Les performances seront mesurées via :

- Une interface graphique (Streamlit) comparant les masques réels, prédits et les scores.
- Une API FastAPI exposant le modèle SegFormer.
- Un dashboard interactif illustrant les résultats, respectant les critères WCAG.

L'objectif est de démontrer que SegFormer B5 apporte une amélioration significative en termes de vitesse, précision et capacité à généraliser, tout en restant exploitable dans un contexte de production embarqué.