



Développer une preuve de concept

Mestapha Oumouni

P7 OC parcours IML
Mentor: Samir Tanfous
16/10/2022

Sommaire

Contexte du projet

Segmentation sémantique

Modèle Unet et SegFormer

Unet simulation

SegFormer fine-tuning

Conclusion

Contexte du projet

- ❖ Réaliser une veille thématique sur un problème de Data Science
- ❖ Mission: Mise en oeuvre d'une approche récente pour un problème de DS et une comparaison avec une méthode baseline
- **Problématique** : Segmentation des images sous-marine
- **Objectif** : Réaliser la tâche de la segmentation avec un Modèle récent pour améliorer la méthode utilisée en production chez un client
Tester le modèle sur des données avec des artefact de la lumière



Introduction

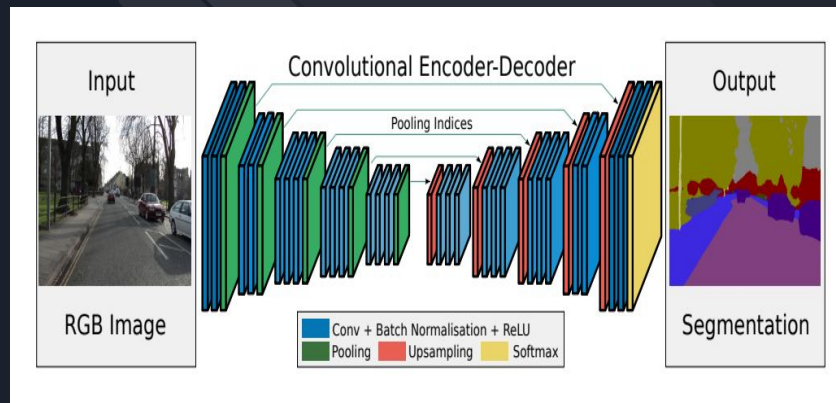
- ★ La vision par ordinateur est de plus en expansion avec beaucoup d'applications
- ★ La segmentation d'images: sous-domaine clé de la vision par ordinateur:
 - Analyse d'images médicales (ex: extraction des limites tumorales)
 - Véhicules autonomes : repérer la surface navigable et piétonnière
 - La vidéosurveillance: comprendre les scènes
 - la réalité augmentée
- ★ Les réseaux de neurones convolutifs ne cessent pas de s'améliorer pour le domaine de la vision

Resumé d'Etat d'art

- ❑ FCN: réseau d'encodeurs suivi d'un réseau de décodeurs
- ❑ U-Net et V-Net: inspiré de FCN

Introduction des transformateurs pour la vision: Modèle inspiré de ViT initialisé pour la classification

- Extension de ViT avec des structures pyramidales (PVT) pour les tâches de la segmentation
- SegFormer: Encodeur de transformateur hiérarchique et un décodeur All-MLP léger



Dataset et Modèles choisis

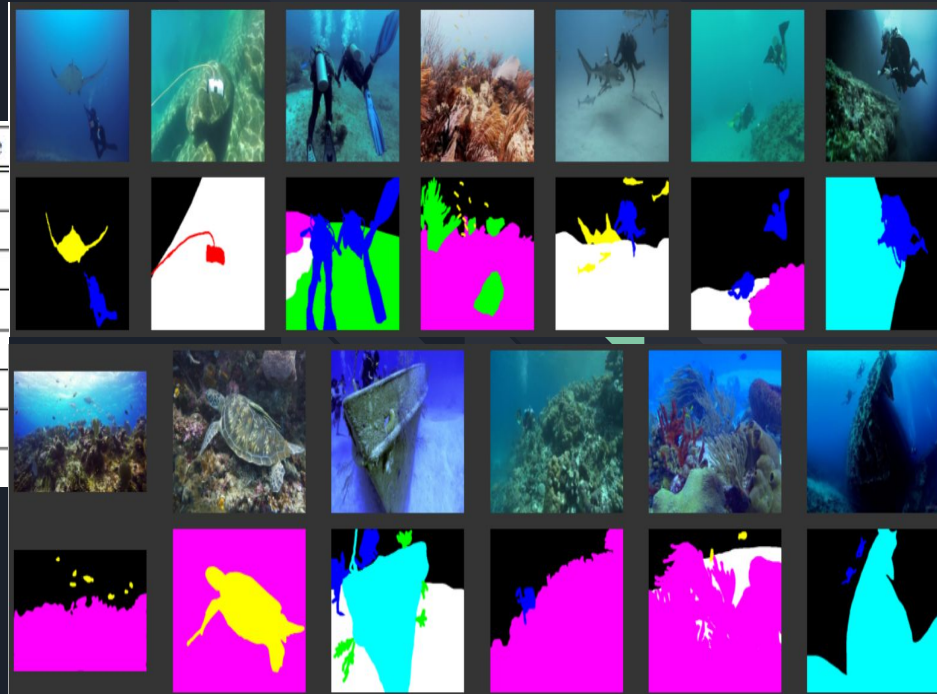
1525 (image,masque) train/validation set

110 (image,masque) test set

Catégorie d'objet	RGB couleur	Code (couleur)	# d'image/catégorie
Fond (plan d'eau)	000	BW(noir)	1288
Plongeurs (humain)	001	HD(bleu)	405
Plantes, herbiers marins	010	PF(vert)	239
Épaves ou ruines	011	WR (ciel)	275
Robots (instruments)	100	RO (rouge)	101
Récifs, invertébrés	101	RI(rose)	1028
Poissons, vertébrés	110	FV(jaune)	1030
plancher, rochers	111	SR(bleu)	635

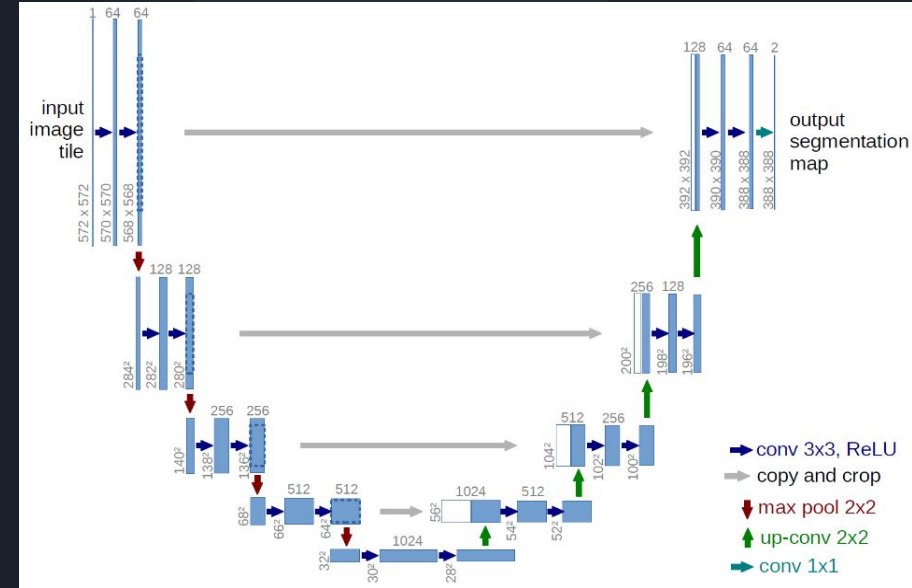
Modèle (baseline): U-net

Modèle (récent): SegFormer



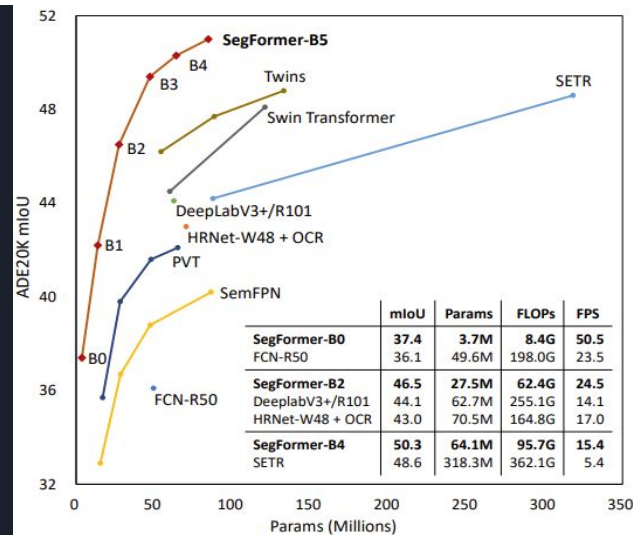
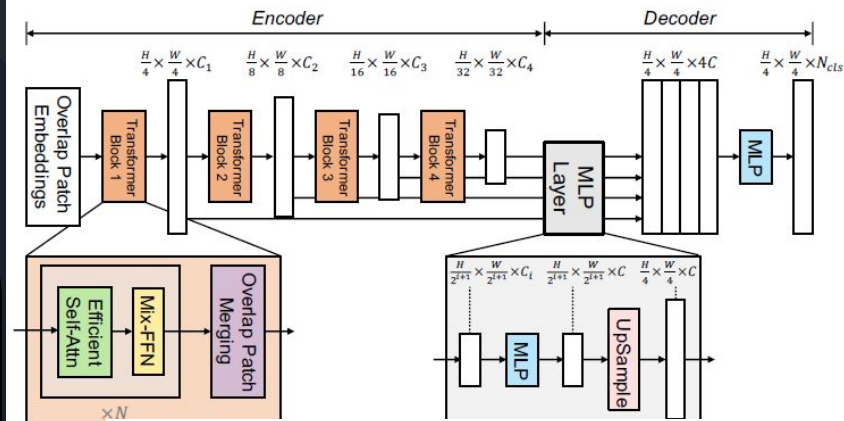
U-net architecture

- Modèle développé de base pour segmenter les images médicales
- Deux blocs:
 - Encodeurs
 - suite de conv 3x3 + ReLu +Maxpooling
 - capture les feature de l'image
 - réduire la dimension spatiale
 - Décodeur
 - suite de conv 3x3 + ReLu suivi d'up_con 2x2
 - conv 1x1 + activation sigmoïde pour conclure la carte des label
- Pont reliant les deux bloc avec deux conv 3x3 +ReLU



SegFormer architecture

- Inspirer de ViT pour la tâche de la segmentation
- Encodeur Transformer hiérarchique génère des features multi-échelles de type CNN
- décodeur MLP simple et légère
- SegFormer développé sous forme d'une série d'encodeurs Mix Transformer (MiT)-B0 à B5
- Un modèle pré entraîné sur des données ADE20k peut être téléchargé sur HuggingFace



Comparaison de méthodes

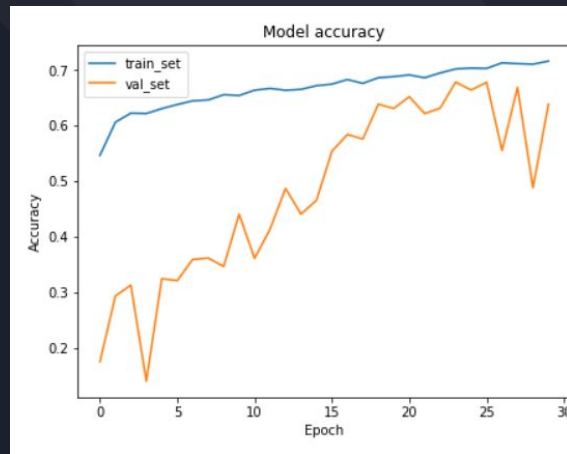
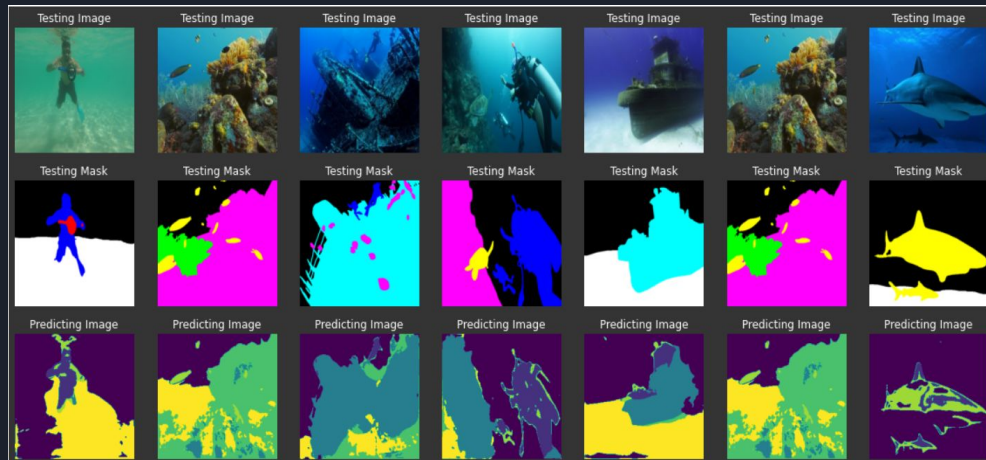
Keras :

- Bibliothèque open source
- High-level API: Entraîner les modèles de Deep Learning avec une syntaxe simple
- Plus une « boîte noire », facile pour commencer et facile à déployer

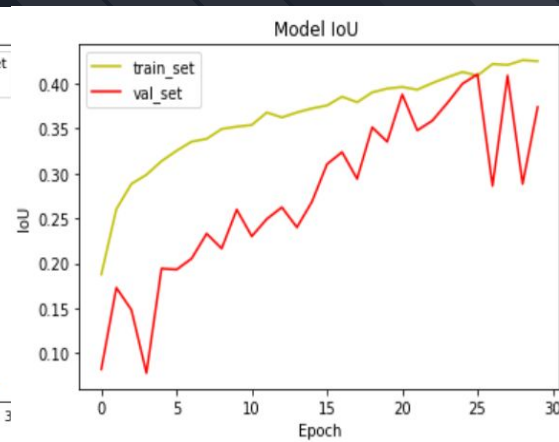
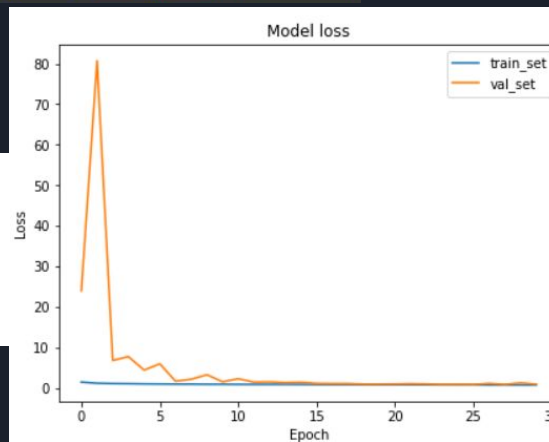
PyTorch :

- Bibliothèque open source développée par Facebook
- Low-level API, la programmation est basée sur des tensors
- Flexibilité, possibilité de personnaliser les fonctions et les modèles facilement
- Utilisé en milieu académique, les nouveaux algorithmes sont souvent en PyTorch

Unet



temps d'entraînement (min)	43.35
mIoU (test)	0.24
Accuracy (test)	0.55

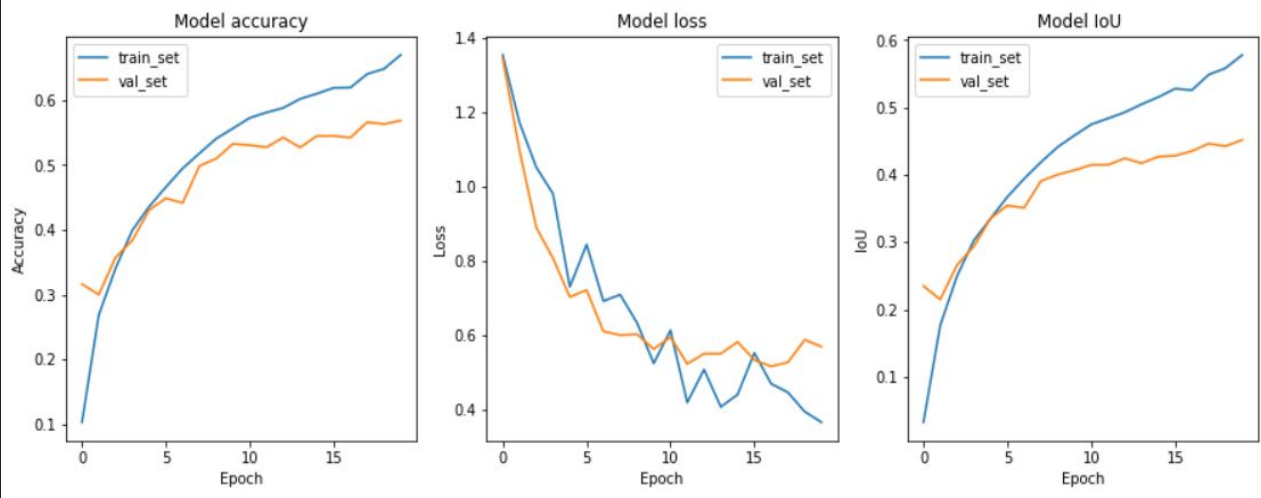


SegFormer

Model pre-entraîné sur des données ADE20k charger depuis HuggingFace
et fine-tune sur nos images

```
model = SegformerForSemanticSegmentation.from_pretrained("nvidia/mit-b0",  
                                                         num_labels=8,  
                                                         id2label=id_to_label,  
                                                         label2id=label_to_id,)
```

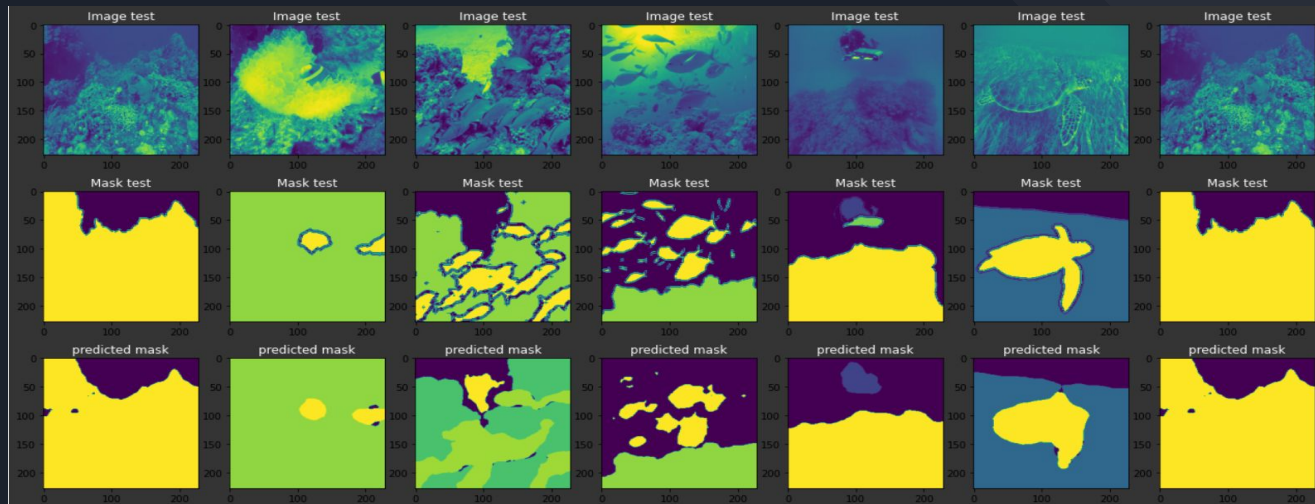
temps d'ajustement (min)	126.98
mIoU test	0.60
Accuracy test	0.71



per-category metrics:

	IoU
water	0.760330
person	0.518212
plants-grass	0.507251
wrecks-ruins	0.768283
box	0.093362
reefs-invertebrates	0.824704
animal	0.625970
sea-floor-rocks	0.746131

SegFormer



- SegFormer finetuned améliore les scores sur observé sur ADE20k
- Meilleurs score sur les données d'entraînement/Unet
- Meilleurs scores sur les données de test/Unet
- Un temps de calcul important/Unet

Bibliographie

[1]: E. Xie, W. Wang, Z. Yu, A. Anandkumar, Jose M. Alvarez, P. Luo. “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers”. arXiv:2105.15203.

[2] O. Ronneberger, P. Fischer, and T. Brox., “U-net: Convolutional networks for biomedical image segmentation”. in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015,

S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos “Image Segmentation Using Deep Learning: A Survey”; arXiv:2001.05566.

<https://huggingface.co/nvidia/segformer-b0-finetuned-ade-512-512>

<https://kaggle.com/datasets/ashish2001/semantic-segmentation-of-underwater-imagery-suim>

Conclusion

- Compétences acquises: segmentation d'images
 - Devenez plus familier avec les algo CNN
 - Découvrir l'Api pytorch
-
- ❖ Segmentation d'images sous-marines avec SegFormer et Unet
 - ❖ SegFormer prouve une sorte d'efficacité / Unet avec un plus en temps de calcul
 - ❖ Inférence d'un modèle pré-entraîné



Perspectives:

- ❑ Data augmentation
- ❑ acquisition d'autre images synthétiques
- ❑ Preprocessing adapté au images sous-marines (artéfacts de la lumière, turbidité)



Merci de votre attention

