

Segmentation d'images adaptative guidée par modèles de langage

Stage M1 de 3 à 4 mois

Encadrement : Michaël Clément (Bordeaux INP - LaBRI) michael.clement@labri.fr
Rémi Giraud (Bordeaux INP - IMS) remi.giraud@ims-bordeaux.fr
Laboratoire : Laboratoire LaBRI, équipe Traitement et Analyse de Données
Mots-clés : Apprentissage profond ; Segmentation d'images ; Modèles de langage.

Contexte scientifique

La segmentation d'images a considérablement progressé grâce aux réseaux de type CNN comme U-Net et, plus récemment, aux Transformers comme le ViT. Les récents modèles de fondation, tels que Segment Anything Model (SAM) [1], offrent une segmentation performante et interactive via des *prompts*, ouvrant la voie à des applications personnalisables. Cependant cette personnalisation n'est souvent possible que par des requêtes sémantiques, ou des clics dans l'image indiquant le centre des objets.

Dans de nombreux domaines, un contrôle précis des segmentations est essentiel. En imagerie médicale, par exemple, il faut parfois imposer des contraintes anatomiques ou topologiques [2]. Plus largement, l'entraînement des modèles à intégrer ces exigences sous forme de requêtes utilisateur permettrait de mieux gérer la forme des objets, le bruit ou des structures spécifiques, rendant ainsi les modèles plus adaptés aux applications complexes.

De nombreuses approches utilisent des fonctions de perte personnalisées pour des segmentations complexes, mais elles restent difficiles à paramétrer et peu généralisables. Une alternative prometteuse consiste à automatiser l'adaptation de la segmentation via l'apprentissage de représentations. Les méthodes d'alignement multimodal, comme CLIP [3], pourraient ainsi intégrer des contraintes exprimées en langage naturel. Récemment, une approche a même permis d'orienter la segmentation par des requêtes textuelles sémantiques [4] (voir Figure 1).

Objectifs du stage

Ce stage vise à étudier la faisabilité et poser les bases d'une méthode adaptative de segmentation d'images guidées par des grands modèles de langage (*large language models*, LLM) permettant de formuler des contraintes sémantiques et structurelles formulées en langage naturel. Le travail de recherche sera articulé autour de deux axes :

1. Modèles de segmentation adaptatifs :
 - Concevoir des modèles interactifs intégrant des contraintes spécifiques.
 - Formuler ces contraintes sous forme de fonctions différentiables.
 - Évaluer la segmentation avec des métriques externes adaptées.
2. Données d'apprentissage :
 - Générer des jeux de données synthétiques avec des formes, des configurations précises et des relations spatiales.
 - Appliquer ces approches à des cas complexes, notamment en imagerie médicale.

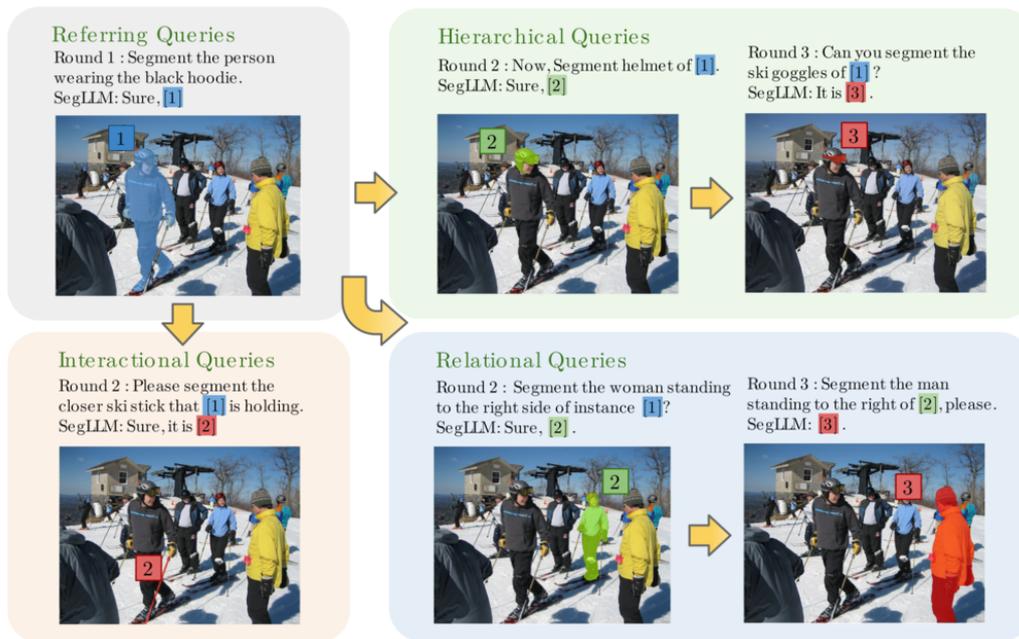


Figure 1 – Illustration du processus de contrôle de segmentation par requête sémantique [4]

Profil recherché

Étudiant·e en Master 1 ou en 2A à l’Université ou en école d’ingénieurs, avec une spécialisation en informatique. Des bases techniques solides en programmation sont demandées (Python, C, C++) ainsi que des connaissances en traitement d’images et apprentissage profond (PyTorch, TensorFlow). La maîtrise de l’anglais scientifique et des qualités rédactionnelles sont également importantes.

Pour candidater : envoyer CV, lettre de motivation et derniers relevés de notes aux encadrants du stage.

Références

- [1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo et al. “Segment anything”. *ICCV*. 2023, p. 4015-4026.
- [2] P. Rougé, O. Merveille et N. Passat. “ccDice: A topology-aware Dice score based on connected components”. *International Workshop on Topology-and Graph-Informed Imaging Informatics at MICCAI*. 2024.
- [3] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark et al. “Learning transferable visual models from natural language supervision”. *ICLR*. 2021.
- [4] X. Wang, S. Zhang, S. Li, K. Kallidromitis, K. Li, Y. Kato, K. Kozuka et T. Darrell. “SegLLM: Multi-round Reasoning Segmentation”. *arXiv preprint arXiv:2410.18923* (2024).