

Stage de fin d'études – Oculométrie Assistée par la Segmentation d'Images

L'oculométrie, ou suivi du regard (*eye-tracking*), est un outil couramment utilisé dans les sciences humaines et sociales, qui permet une analyse fine et objective des processus cognitifs et comportementaux. En capturant les mouvements oculaires et en identifiant avec précision les points d'attention visuelle, cette technologie offre des données empiriques objectives sur les mécanismes de perception, d'attention, et de prise de décision. L'exploitation des données d'oculométrie présente plusieurs défis, notamment en raison de la complexité et de la densité des informations recueillies. Tout d'abord, les mouvements oculaires, tels que les saccades, les fixations et les clignements, sont extrêmement rapides et se produisent à des fréquences élevées, générant ainsi des volumes massifs de données. Ces données brutes nécessitent un prétraitement pour isoler des séquences pertinentes, corriger les artefacts et synchroniser les signaux avec d'autres variables expérimentales. Ensuite, l'interprétation des résultats est délicate, car les schémas oculaires ne traduisent pas directement les processus cognitifs : une fixation prolongée, par exemple, peut refléter une grande attention ou au contraire une confusion.

Des progrès techniques récents ont permis l'émergence de dispositifs d'oculométrie adaptés à des acquisitions en extérieur (hors du laboratoire). Ceci rend l'analyse des résultats encore plus délicate car les environnements observés par le sujet sont inconnus et imprévisibles (contrairement à des expériences en laboratoire pendant lesquelles les sujets sont confrontés à des environnements contrôlés).

La segmentation d'image est un processus fondamental en vision par ordinateur qui consiste à diviser une image en plusieurs régions ou segments, chaque région correspondant à des parties distinctes de l'image ayant des caractéristiques communes, comme la couleur, la texture ou l'intensité. L'objectif est de faciliter l'analyse et la compréhension de l'image en isolant les objets ou les zones d'intérêt, comme des visages, des bâtiments ou des cellules biologiques, en fonction de leur signification ou de leur fonction. Contrairement à la classification d'image qui assigne une seule étiquette à une image entière, la segmentation permet de classer chaque pixel, offrant une compréhension plus détaillée et précise.



Figure 1 - Exemple de segmentation (manuelle) d'une scène urbaine

Les progrès récents en apprentissage profond ont révolutionné la segmentation d'image, notamment grâce à l'émergence des réseaux de neurones convolutifs (*Convolutional Neural Networks*, CNN) et des architectures spécifiques comme U-Net [2], Mask R-CNN [3] ou encore DeepLab [4]. Ces modèles exploitent de vastes ensembles de données annotées et des techniques de régularisation pour apprendre des représentations hiérarchiques, permettant une segmentation automatique efficace. Les méthodes basées sur l'apprentissage profond surpassent largement les approches traditionnelles, en particulier pour les tâches complexes comme la segmentation sémantique (où chaque pixel est associé à une catégorie) et la segmentation d'instances (qui différencie les occurrences d'un même objet).

Des travaux récents [1] ont prototypé un système croisant la collecte de données oculométriques avec la segmentation d'images. L'idée, simple, consiste à identifier grâce à la segmentation d'image quel objet est actuellement fixé par un sujet, en croisant le point de fixation obtenu par oculométrie avec la carte de segmentation prédite par le réseau de neurones. Ces travaux se sont attachés à montrer la forte corrélation entre les résultats obtenus automatiquement et les résultats obtenus manuellement par un annotateur humain. Bien qu'intéressants, ces travaux ne permettent pas réellement à des chercheurs en SHS de se saisir de ces nouveaux outils car les auteurs utilisent un modèle de segmentation générique (entraîné sur MS-COCO, soit 90 classes d'objet de la vie quotidienne), et ne couvrant pas suffisamment d'objets pour présenter un intérêt dans de nombreux cas.

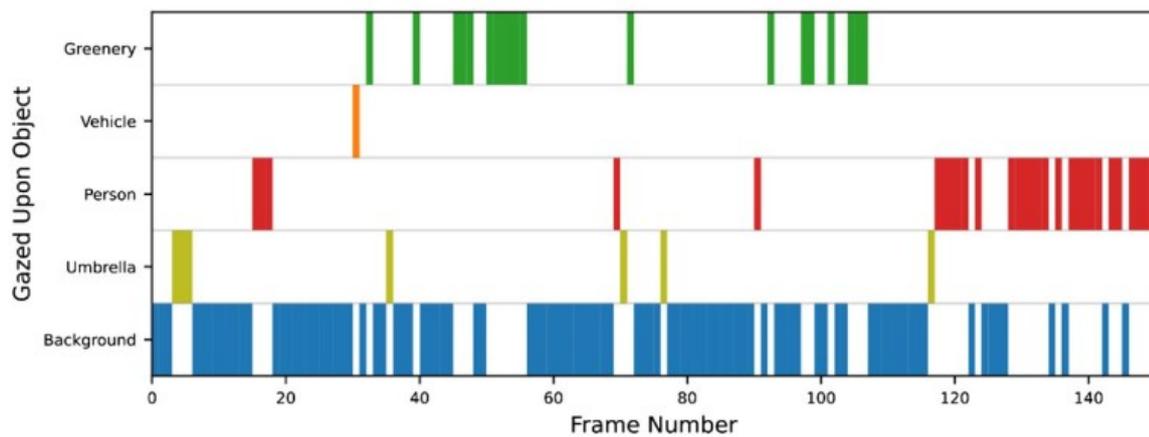


Figure 2 - Exemple de restitution d'une analyse automatisée de données oculométriques [1]

Notre ambition pour ce stage est de nous concentrer sur deux cas précis pour lesquels un modèle de segmentation spécifique doit être entraîné. Le premier cas est d'intérêt pour le laboratoire CLLE (Cognition, Langues, Langage, Ergonomie) et concerne la conduite en milieu urbain. Ceci nécessite donc l'entraînement d'un modèle de segmentation de scènes urbaines.

Le second cas est d'intérêt dans le cadre du projet RECONNECT (financement TIRIS), et nécessite de reconnaître les différents éléments d'un paysage naturel : flore (éventuellement différents types de végétation comme des prairies, des arbustes, arbres, etc.), faune, eau (rivières, lacs), montagne, etc. Ce second cas présente une difficulté supplémentaire car il existe peu de bases de données de segmentation présentant des annotations pour ces classes.

Les objectifs du stage sont donc :

- l'entraînement d'un modèle de segmentation sémantique de scènes urbaines
- la mise au point d'une base de données (par agrégation de bases existantes) pour la segmentation sémantique de scènes naturelles
- la restitution visuelle des données oculométriques à la manière de la Figure 2 ci-dessus

Le stage sera encadré par Axel Carlier (Axel.Carlier@toulouse-inp.fr) et aura lieu au laboratoire IRIT, au bâtiment F de l'ENSEEIH. La gratification mensuelle sera d'environ 620€/mois.

[1] Deane, Oliver, Eszter Toth, and Sang-Hoon Yeo. "Deep-SAGA: a deep-learning-based system for automatic gaze annotation from eye-tracking data." *Behavior Research Methods* 55.3 (2023): 1372-1391.

[2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015.

[3] He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.

[4] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017): 834-848.