

Apprentissage profond à partir de représentations structurelles d'images

Stage de Master 2 (5-6 mois)

Laboratoire : LaBRI, UMR CNRS 5800
351 cours de la Libération, 33400 Talence, France

Encadrants : Michaël Clément / LaBRI / Image et Son (michael.clement@labri.fr)
Rémi Giraud / IMS / Signal et Image (remi.giraud@ims-bordeaux.fr)

Mots-clés : *Apprentissage profond ; Représentations structurelles ; Relations spatiales*

Description du sujet

Contexte

Les récentes avancées liées à l'apprentissage profond (*deep learning*) et aux réseaux de neurones convolutifs représentent des perspectives prometteuses dans le domaine de la vision par ordinateur. Ces méthodes ont permis d'atteindre des performances remarquables pour diverses problématiques comme la classification d'images, la détection d'objets, ou encore la segmentation sémantique (voir la Figure 1). Cependant, ces stratégies d'apprentissage nécessitent toujours une très grande quantité de données annotées, et le comportement des modèles reste parfois difficilement interprétable.

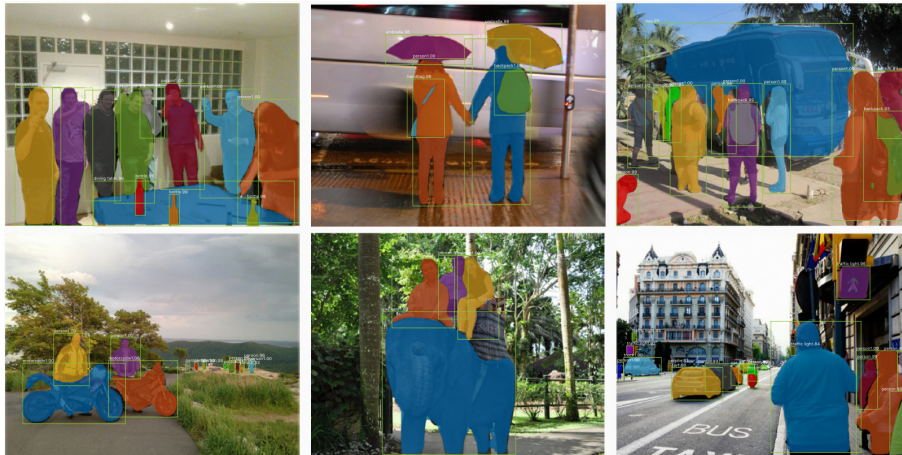


Figure 1 – Exemples de segmentations sémantiques obtenues par l'approche *Mask R-CNN* [1].

Parallèlement, les méthodes de partitionnement d'images en régions ont été progressivement introduites. Par exemple, les décompositions en superpixels [2] peuvent être utilisées pour limiter le nombre d'éléments à traiter dans l'image et ainsi diminuer les temps de calcul. De plus, ces méthodes permettent également d'interpréter le contenu des images d'un point de vue structurel, à partir un ensemble de régions de couleurs homogènes, respectant les contours des objets. L'image peut ainsi être représentée par un graphe relationnel de régions, potentiellement à différentes échelles, depuis lequel des caractéristiques sur la structure de l'image peuvent être extraites. Ces caractéristiques peuvent alors constituer une information complémentaire par rapport aux traitements à l'échelle pixellique des réseaux de neurones, dans le but d'améliorer la qualité et l'interprétabilité des résultats.

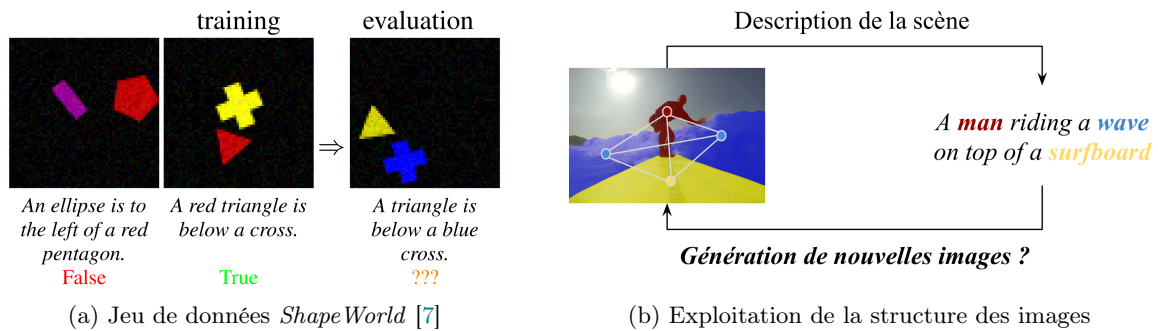


Figure 2 – L’apprentissage de caractéristiques de relations spatiales entre régions pourrait permettre de nouvelles applications telles que la génération d’images garantissant une certaine structure spatiale.

Objectifs du stage

Ce stage de recherche a pour objectif d’étudier un nouveau cadre méthodologique permettant d’exploiter des caractéristiques structurales à partir de réseaux de neurones profonds. Certains travaux récents ont déjà pu illustrer l’intérêt des approches par superpixels en apprentissage profond, par exemple pour assurer la cohérence de résultats de segmentation [3], ou encore pour la détection d’objets saillants [4]. Cependant, à l’heure actuelle aucune méthode ne prend en compte explicitement les relations spatiales qui existent entre les régions, qui consistent pourtant une information pertinente, se rapprochant de la sémantique de l’image.

En particulier, nous proposons d’étendre nos récents travaux sur l’apprentissage de relations spatiales entre régions [5], notamment afin que les descripteurs structurels puissent être intégrés au sein d’architectures de réseaux de neurones convolutifs. Une piste consisterait notamment à mettre en place un système inspiré des réseaux antagonistes génératifs [6] qui serait capable de générer des configurations spatiales entre différents objets à partir de propositions en langage naturel, et inversement (voir la Figure 2). Selon l’avancement du stage, ces propositions méthodologiques pourront être employées pour des tâches de segmentation et de classification à partir de données issues d’images naturelles, de données géospatiales, ou bien encore de l’imagerie médicale.

Profil recherché

Étudiant-e en Master 2 ou dernière année d’école d’ingénieurs, spécialisé-e en informatique, image et/ou intelligence artificielle. Des connaissances et expériences solides en traitement d’images, apprentissage profond et programmation (Python, TensorFlow/PyTorch, C/C++) sont demandées. La maîtrise de l’anglais scientifique et des qualités rédactionnelles sont également très importantes.

Candidature : Pour candidater, envoyer un dossier avec CV, lettre de motivation, relevés de notes, ainsi que toute pièce susceptible de renforcer la candidature (lettre de recommandation, etc.). Pour l’envoi des pièces demandées, ou pour toute question sur le sujet : michael.clement@labri.fr

Références

- [1] He, K., Gkioxari, G., Dollár, P., and Girshick, R. “Mask R-CNN”. *International Conference on Computer Vision (ICCV)*. 2017, pp. 2961–2969.
- [2] Stutz, D., Hermans, A., and Leibe, B. “Superpixels: An evaluation of the state-of-the-art”. *Computer Vision and Image Understanding* 166 (2018), pp. 1–27.
- [3] Farabet, C., Couprie, C., Najman, L., and LeCun, Y. “Learning hierarchical features for scene labeling”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), pp. 1915–1929.
- [4] He, S., Lau, R. W., Liu, W., Huang, Z., and Yang, Q. “SuperCNN: A superpixelwise convolutional neural network for salient object detection”. *International Journal of Computer Vision* 115.3 (2015), pp. 330–344.
- [5] Clément, M., Kurtz, C., and Wendling, L. “Learning spatial relations and shapes for structural object description and scene recognition”. *Pattern Recognition* 84 (2018), pp. 197–210.
- [6] Goodfellow, I. et al. “Generative adversarial nets”. *Advances in Neural Information Processing Systems (NIPS)*. 2014, pp. 2672–2680.
- [7] Kuhnle, A. and Copestake, A. *ShapeWorld - A new test methodology for multimodal language understanding*. 2017. arXiv: [1704.04517](https://arxiv.org/abs/1704.04517).