

Modèles d'apprentissage profond à partir de représentations structurelles d'images

Environnement

Direction :	Michaël Clément (Bordeaux INP - LaBRI) michael.clement@labri.fr Rémi Giraud (Bordeaux INP - IMS) remi.giraud@ims-bordeaux.fr
Financement :	Bourse ministérielle (2300€ bruts /mois)
Laboratoire :	Laboratoire IMS, campus de Talence (proche Bordeaux) Entouré d'autres laboratoires de recherche (LaBRI, IMB)
Équipe :	Signal et Image. 20 permanents, 20 doctorants/stagiaires
Enseignements :	Possibilité d'enseigner durant la thèse
Mots-clés :	Apprentissage profond ; Vision par ordinateur ; Analyse d'images ; Segmentation ; Relations spatiales.

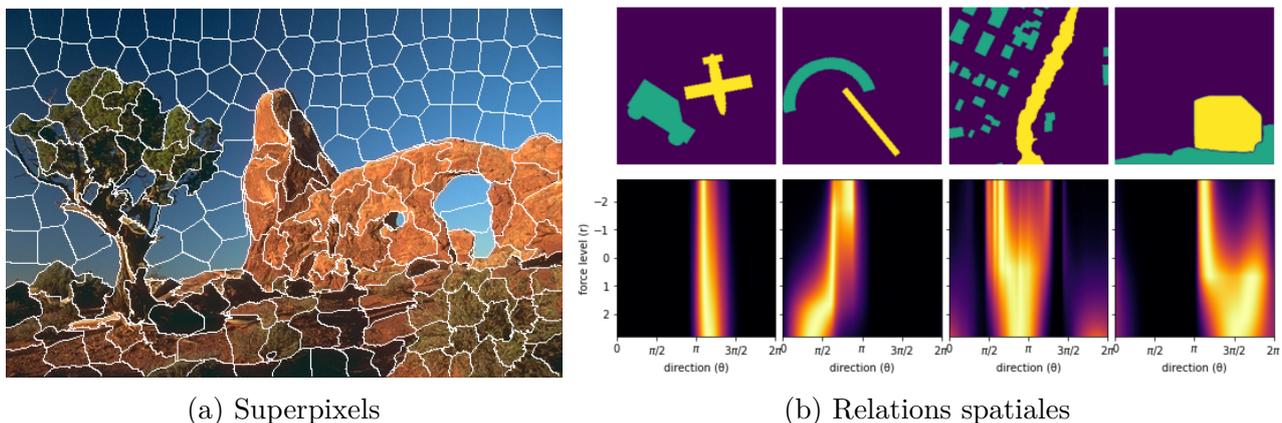


Figure 1 – Outils considérés durant cette thèse. (a) Exemple de segmentation en superpixels [1]. (b) Représentation 2D des relations spatiales entre objets d'une image [2].

Contexte scientifique

Les méthodes de l'état de l'art en apprentissage automatique reposent généralement sur des procédures d'entraînement de grands réseaux de neurones et bases de données annotées. Les modèles résultant sont souvent volumineux et complexes, et peuvent être difficiles à évaluer ou interpréter. En particulier, les réseaux populaires en vision par ordinateur utilisent pour beaucoup les blocs de convolution, qui sont des opérations à l'échelle pixellique appliquées sur des grilles régulières. Ces architectures présentent encore certaines limitations, notamment sur la taille des images qui doit souvent être limitée lors de la phase d'apprentissage, pour des contraintes de temps de calcul et de représentation mémoire. Dans ce contexte, il existe un vrai challenge pour intégrer des représentations structurelles ou symboliques de données afin de synthétiser efficacement l'information et aussi permettre de guider le processus d'apprentissage.

Une première piste de recherche consiste à utiliser une sous-représentation de l'image type segmentation en superpixels [1] (voir Figure 1(a)), réduisant le nombre d'éléments à traiter tout en respectant la structure et les contours des objets de l'image. Cependant, l'absence de structure régulière (forme/taille/voisinage) pose problème pour les utiliser dans un réseau de neurones convolutionnel. Des premiers travaux ont été initiés dans [3] pour considérer une segmentation hiérarchique en superpixels et dans [4] sur la définition d'opérateurs de convolution et de blocs d'attention adaptés aux superpixels. Dans cette thèse, nous proposons d'approfondir et de consolider ces travaux pour différentes applications d'analyse et de traitement d'images.

Une deuxième piste de recherche vise à intégrer des connaissances sémantiques sur les relations spatiales entre objets dans les réseaux de neurones. En effet, la structure spatiale des objets dans une image constitue une information pertinente à exploiter pour interpréter leur contenu [2] (voir Figure 1(b)). Pour cela, la thèse s'appuiera sur nos travaux récents utilisant des représentations structurelles en apprentissage profond avec par exemple [5] (représentations basées graphes des structures cérébrales) ainsi que sur notre expertise de long terme sur l'apprentissage des caractéristiques des relations spatiales [6].

Objectifs de la thèse

L'objectif principal de cette thèse est donc de réussir à lever les verrous méthodologiques qui empêchent à l'heure actuelle l'utilisation explicite et naturelle de représentations structurelles d'images au sein des modèles d'apprentissage profond. Le travail de recherche sera articulé autour des deux axes méthodologiques mentionnés ci-avant :

1. Conception de nouvelles architectures de réseaux de neurones permettant d'intégrer des représentations structurelles (graphes, hiérarchies de régions) ;
2. Intégration aux modèles de l'information de relations spatiales entre les structures afin de guider l'apprentissage sémantiquement.

Cette thèse pourra se dérouler suivant plusieurs étapes et objectifs intermédiaires :

- Réalisation d'un état de l'art complet sur les représentations structurelles en apprentissage profond : méthodes de segmentation hiérarchiques, de description des relations spatiales, réseaux de neurones sur graphes, modèles génératifs avec espaces latents, etc. ;
- Conception et prototypage de nouvelles méthodes intégrant directement des représentations structurelles dans des architectures de réseaux de neurones ;
- Validations expérimentales de ces méthodes sur différents contextes applicatifs : classification et segmentation d'images, colorisation d'images avec interaction utilisateur, super-résolution, etc. ;
- Valorisation des travaux par la publication d'articles scientifiques dans des conférences et revues de premier plan.

Selon les résultats obtenus, et les motivations du doctorant ou de la doctorante, d'autres applications reposant sur les contributions théoriques proposées pourront être considérées (génération d'images, segmentation de séquences d'images, etc.).

Des collaborations avec des partenaires académiques (Bordeaux Sciences Agro, Université Paris Cité, University of the Basque Country, University of Valencia) et industriels (Thales Mérignac) travaillant sur les mêmes thématiques pourront aussi être envisagées.

Profil recherché

Diplômé.e d'un Master 2 ou d'école d'ingénieurs, avec une spécialisation en informatique. Des bases techniques solides en programmation sont demandées (Python, 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 très importantes.

Pour candidater : envoyer CV, lettre de motivation et relevés de notes aux directeurs.

Références

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua et S. Süsstrunk. “SLIC superpixels compared to state-of-the-art superpixel methods”. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* (2012).
- [2] R. Deléarde, C. Kurtz et L. Wendling. “Description and recognition of complex spatial configurations of object pairs with Force Banner 2D features”. *Pattern Recognition* (2022).
- [3] R. Giraud, M. Boyer et M. Clément. “Multi-scale superpatch matching using dual superpixel descriptors”. *Pattern Recognition Letters, Special Issue on Hierarchical Representations* (2019).
- [4] H. Carrillo, M. Clément et A. Bugeau. “Super-attention for exemplar-based image colorization”. *Asian Conference on Computer Vision (ACCV)*. 2022.
- [5] H.-D. Nguyen, M. Clément, B. Mansencal et P. Coupé. “Towards better interpretable and generalizable AD detection using collective artificial intelligence”. *Computerized Medical Imaging and Graphics* (2023).
- [6] M. Clément, C. Kurtz et L. Wendling. “Learning spatial relations and shapes for structural object description and scene recognition”. *Pattern Recognition* (2018).