

Étude et combinaison de critères de sélection de régions pour la segmentation faiblement supervisée

Stage M2/Ingénieur en Informatique

Informations générales

- Mots-clés : vision par ordinateur, deep learning, segmentation, fusion
- Durée du stage : 6 mois (gratification standard)
- Structure d'accueil : Université de Paris, Laboratoire d'Informatique Paris Descartes (LIPADE), équipe [Systèmes Intelligents de Perception](#)
- Adresse : 45 rue des Saints-Pères, 75006 Paris
- Encadrement : Camille Kurtz, Laurent Wendling, Robin Deléarde - (prenom.nom@u-paris.fr)

Sujet du stage

Contexte

La segmentation d'objet est un problème majeur en vision par ordinateur, qui a largement bénéficié des avancées apportées par l'apprentissage profond, avec des modèles comme FCN, R-CNN, Mask R-CNN ou encore plus récemment Mask Scoring R-CNN (voir la Figure 1). Cependant ces solutions requièrent un volume important de données annotées au niveau pixellique, y compris lorsque l'on utilise des méthodes de transfert de l'apprentissage d'un dataset à un autre. A l'inverse, annoter les images avec uniquement un rectangle englobant (*bounding box*) est beaucoup plus rapide (15 fois plus rapide selon certaines études). C'est ainsi que s'est développée dans le même temps la tâche de segmentation faiblement supervisée, dont le but est d'entraîner un modèle uniquement à partir d'annotations faibles (*bounding box* et/ou labels).

Concernant les *bounding box*, la plupart des approches se décomposent en 3 étapes pour chaque image de la base d'apprentissage : 1. proposition de régions cohérentes (avec un algorithme type MCG ou Selective Search), 2. sélection de la meilleure région à partir de la *bounding box* et éventuellement affinage de la sortie, 3. entraînement du modèle de façon complètement supervisée en utilisant les régions produites comme annotations niveau pixels. On pourra par exemple se référer à la solution BoxSup [1] proposée par Microsoft en 2015, ou [2] qui est la seule méthode à proposer de la segmentation par instance dans un cadre faiblement supervisé à notre connaissance. Dans BoxSup [1], la sélection de la meilleure région se fait grâce à un critère de couverture de la *bounding box* par la région, critère qui est intégré dans la fonction de coût lors de l'apprentissage.

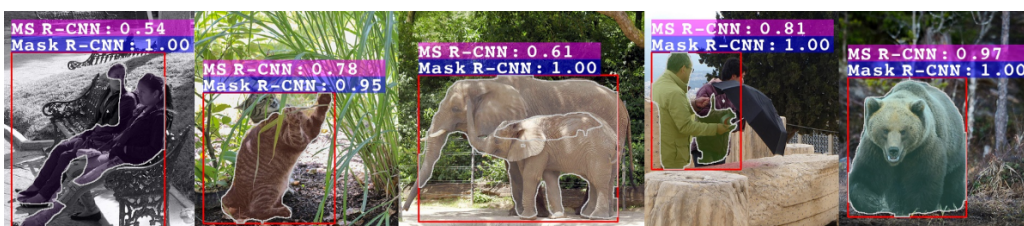


Figure 1 – Illustration d'un résultat d'instance segmentation obtenu via Mask Scoring R-CNN.

Travail a réaliser

Dans ce stage, nous proposons d'utiliser plusieurs critères géométriques intelligents pour réaliser la sélection de région, et de les combiner de façon à optimiser le score de segmentation (mIoU typiquement). Ils seront à intégrer dans un modèle de segmentation faiblement supervisée type BoxSup, afin de pouvoir évaluer une telle solution sur un dataset usuel comme Pascal-VOC ou COCO et ainsi se comparer à l'état-de-l'art.

Plusieurs critères ont déjà été étudiés par l'équipe dans un contexte similaire, utilisant notamment comme a priori le fait que la bounding box contient entièrement l'objet et pas plus (*tightness prior*), comme dans [3], ou encore [2] de façon indirecte (voir la Figure 2). Ces critères pourront être repris ici, et d'autres pourront être ajoutés en fonction des données à traiter. Le choix de la meilleure façon de les agréger reste à explorer et nous étudierons ici, entre autres, l'intégrale de Choquet qui généralise de nombreux opérateurs comme les statistiques d'ordre n , la somme pondérée, etc. À partir d'une approche robuste d'apprentissage fondée sur une descente de gradient par contraintes (qui requiert peu d'échantillons) il est possible ensuite d'estimer l'importance de chaque critère et leurs interactions [4]. Notre équipe a proposé plusieurs modèles robustes de sélection pour extraire des sous-ensembles pertinents de critères [5]. Ces modèles permettent d'améliorer sensiblement les taux de reconnaissance dans de nombreuses applications tout en garantissant une interprétabilité des résultats.

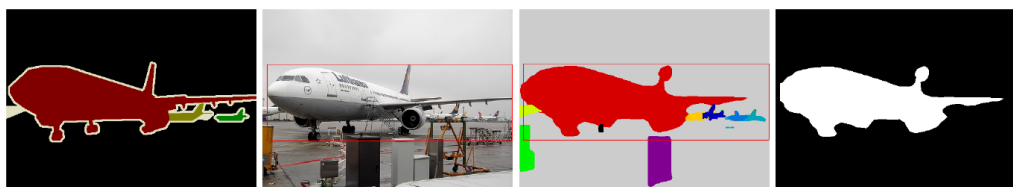


Figure 2 – Illustration du processus de sélection de la meilleure région à partir d'une bounding box.

Selon les résultats, cette méthode pourra faire l'objet d'une publication et être adaptée à d'autres algorithmes de segmentation faiblement supervisée, utilisant aussi les labels par exemple, ou pour adapter un modèle comme Mask Scoring R-CNN [6] au cas faiblement supervisé.

Profil recherché pour le/la candidat(e)

Nous recherchons un(e) étudiant(e) en Master 2 (ou équivalent Ingénieur) informatique, ayant des compétences en analyse d'images, en apprentissage automatique et en programmation Python. Une expérience avec un *framework* d'apprentissage profond est un plus.

Sources bibliographiques

- [1] Jifeng DAI, Kaiming HE et Jian SUN. "Boxsup : Exploiting bounding boxes to supervise convolutional networks for semantic segmentation". In : *ICCV*. 2015, p. 1635-1643.
- [2] Anna KHOREVA et al. "Simple does it : Weakly supervised instance and semantic segmentation". In : *CVPR*. 2017, p. 876-885.
- [3] Cheng-Chun HSU et al. "Weakly supervised instance segmentation using the bounding box tightness prior". In : *NIPS*. 2019, p. 6586-6597.
- [4] M. GRABISCH. "The application of fuzzyintegral in multicriteria decision making". In : *Euro. J. Operat. Res.* 89 (1995), p. 445-456.
- [5] J. RENDEK et L. WENDLING. "On Determining Suitable Subsets of Decision Rules Using Choquet Integral". In : *Int. J. Pattern Recognit. Artif. Intell.* 22.2 (2008), p. 207-232.
- [6] Zhaojin HUANG et al. "Mask scoring R-CNN". In : *CVPR*. 2019, p. 6409-6418.