



Apprentissage actif pour la détection et la segmentation d'objets

Présentation du laboratoire d'accueil

Basé à Paris-Saclay, le CEA List est l'un des quatre instituts de recherche technologique de CEA Tech, direction de la recherche technologique du CEA. Dédié aux systèmes numériques intelligents, il contribue au développement de la compétitivité des entreprises par le développement et le transfert de technologies.

L'expertise et les compétences développées par les 800 ingénieurs-chercheurs et techniciens du CEA List permettent à l'Institut d'accompagner chaque année plus de 200 entreprises françaises et étrangères sur des projets de recherche appliquée s'appuyant sur 4 programmes et 9 plateformes technologiques. 21 start-ups ont été créées depuis 2003. Labellisé Institut Carnot depuis 2006, le CEA List est aujourd'hui l'institut Carnot Technologies Numériques.

Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur et de l'Intelligence Artificielle selon quatre axes principaux :

- La reconnaissance visuelle (détection, reconnaissance fine et segmentation d'objets, segmentation de scène, détection d'anomalies)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, détection de comportements anormaux ou spécifiques)
- L'annotation intelligente à grande échelle de données visuelles
- Les modèles de perception et décision.

Description du stage

Le récent succès des méthodes d'apprentissage profond (*deep learning*) en vision par ordinateur repose en grande partie sur la disponibilité de grandes quantités de données labellisées pour l'entraînement de réseaux de neurones profonds. Cependant, l'annotation de données pour des tâches différentes de celles des benchmarks publics est très coûteuse. En effet, l'étiquetage manuel de nombreuses données demande beaucoup de temps humain, est sujet aux erreurs et requiert l'avis d'experts lorsque l'interprétation des données n'est pas évidente pour le non-spécialiste (ex : structures d'intérêt dans les images biomédicales, modèles précis de véhicules aériens, objets vus sur des images satellitaires, ...). L'apprentissage actif (*active learning*) a pour but de rendre plus efficace la labellisation de données en sélectionnant des échantillons qui, une fois labellisés, amélioreront le plus le modèle. Ainsi, le modèle peut être entraîné de façon faiblement supervisé, à moindre coût.

Les nombreuses méthodes d'apprentissage actif proposées depuis plus de deux décennies se classent généralement en trois grandes familles selon le critère de *sélection de données* employé [1]. Certaines approches définissent et mesurent l'incertitude des données pour présenter à l'annotateur les plus incertaines. D'autres approches se basent sur la diversité des données à annoter de façon à représenter au mieux la distribution de l'ensemble des données non-labellisées. Enfin, des approches cherchent à sélectionner les données qui causeraient le plus grand changement aux paramètres ou aux sorties du modèle courant, si leurs labels étaient connus. Des stratégies hybrides ont également été proposées pour tirer parti de la complémentarité de ces approches.

Cependant, plusieurs problèmes limitent l'utilisation de ces méthodes. Le premier est le passage à l'échelle. De nombreuses méthodes, développées sur des modèles de taille modeste et de petits jeux de données, ne sont pas efficaces lorsque des réseaux de neurones profonds sont utilisés sur des datasets plus gros. Autres conséquences d'utilisation de jeux de données de grande taille : les données peuvent ne plus être disponibles dans leur globalité mais arrivent en flux dans lequel il faut pouvoir prédire l'importance de chaque élément ; une annotation fine des données parmi les nombreux labels possibles devient trop coûteuse pour l'humain [2].

Par ailleurs, l'ensemble des données d'entraînement ne représente généralement pas bien l'ensemble des cas possibles que le modèle doit traiter. Il est parfois difficile de disposer de nombreuses données présentant les multiples et subtils attributs propres à un label spécifique. Ainsi, certaines approches [3] se basent sur des modèles génératifs, tels que les auto-encodeurs variationnels conditionnels (*Conditional Variational Auto-Encoder*), pour générer des exemples synthétiques.

Enfin, de nombreuses méthodes d'apprentissage actif sont spécifiques à une tâche donnée et ne se transposent pas directement sur d'autres tâches. Alors que la classification d'images est largement traitée dans le cadre de l'apprentissage actif, les tâches de détection et de segmentation d'objets ont été très peu explorées [4, 5]. Aussi, déterminer les régions les plus informatives de l'image pour une hypothétique présence d'objets demeure

difficile. Certaines méthodes proposent des approches génériques, indépendantes de la tâche à résoudre [1] tandis que d'autres utilisent divers paradigmes d'apprentissage, comme l'auto-apprentissage via *pseudo-labeling* [6] (voir Figure 1) ou l'apprentissage incrémental (*continual learning*) [7], pour faciliter l'apprentissage actif.

L'objectif de ce stage sera d'explorer le paradigme d'apprentissage actif appliqué aux tâches de détection et segmentation d'objets dans les images. Après évaluation des performances et des limitations des méthodes de l'état de l'art sur des jeux de données de taille et de nature différentes, on proposera de nouvelles méthodes ou des améliorations de méthodes existantes. Les résultats seront évalués sur les datasets académiques ainsi que sur des données de domaines bien distincts issues de problématiques industrielles. La publication de brevets et articles sur les travaux menés sera encouragée.

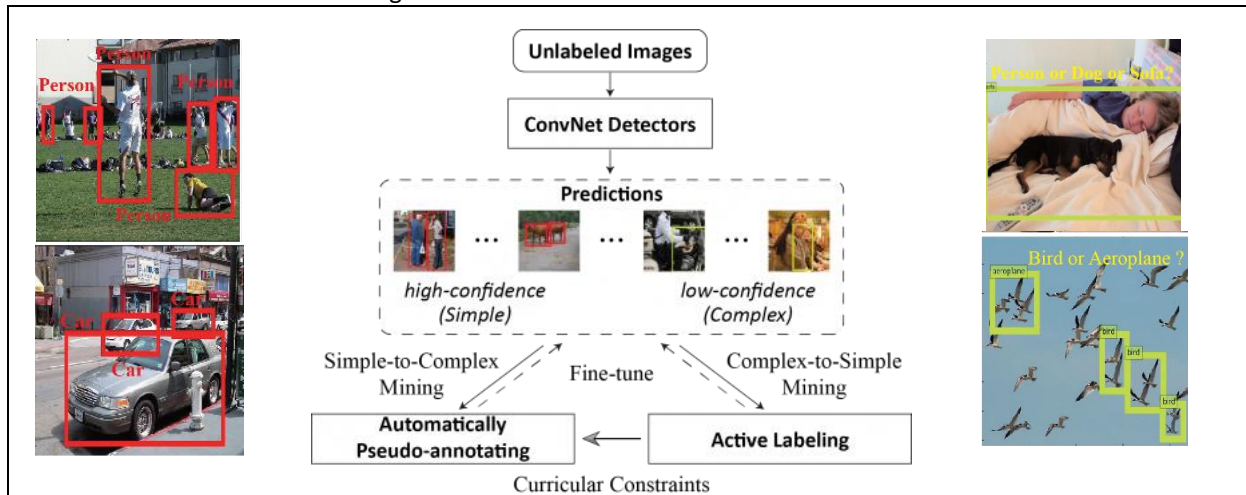


Figure 1 : Framework proposé par [6] pour l'apprentissage d'un détecteur d'objets via pseudo-labellisation de régions prédites avec une confiance élevée et labellisation active de régions prédites avec une confiance faible.

Mots-clés :

vision par ordinateur, apprentissage profond, apprentissage semi-supervisé, apprentissage actif, détection d'objets, segmentation sémantique d'instances.

Références

- [1] Yoo and Kweon. Learning Loss for Active Learning. In *CVPR* 2019.
- [2] Hu et al. Active Learning With Partial Feedback. In *ICLR* 2019.
- [3] Yu and Grauman. Thinking Outside the Pool: Active Training Image Creation for Relative Attributes. In *CVPR* 2019.
- [4] Kao et al. Localization-Aware Active Learning for Object Detection. In *ACCV* 2018.
- [5] Roy et al. Deep active learning for object detection. In *BMVC* 2018.
- [6] Wang et al. Cost-Effective Object Detection: Active Sample Mining With Switchable Selection Criteria. In *IEEE Tr. on Neural Networks and Learning Systems*, Vol. 30, No. 3, March 2019.
- [7] Brust et al. Active Learning for Deep Object Detection. In *VISAPP* 2019.

Profil du candidat/de la candidate

Niveau demandé :	Ingénieur, Master 2
Ce stage ouvre la possibilité de poursuite en thèse et ingénieur R&D dans notre laboratoire.	
Durée :	6 mois
Rémunération :	entre 700 € et 1300 € suivant la formation.
Compétences requises :	
<ul style="list-style-type: none"> - Vision par ordinateur - Apprentissage automatique (deep learning) - Reconnaissance de formes - C/C++, Python - La maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch) est un plus. 	