

THESE 2022

Réf : LVA-22-T2

Stratégies d'apprentissage d'IA frugales en annotations du domaine cible

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 (Computer Vision) selon quatre axes principaux :

- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns; détection d'anomalies; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- L'annotation intelligente (annotation à grande échelle de données visuelles 2D/3D de manière semi-automatique)
- La perception et la décision (processus de décision markovien, navigation)

Problématique

L'apprentissage supervisé de réseaux de neurones profonds a prouvé son efficacité pour résoudre de nombreuses tâches de vision par ordinateur utiles dans divers domaines applicatifs : du véhicule autonome à la vidéo-surveillance en passant par la médecine, l'agronomie et le contrôle industriel. Cependant, ce paradigme d'apprentissage nécessite une grande quantité de données représentatives ainsi que les annotations correspondant à la tâche à apprendre. Outre le coût de l'acquisition de données, le coût de leur annotation manuelle, fastidieuse, par des humains (voire des experts) rend l'apprentissage de l'IA rédhibitoire. Par conséquent, *comment apprendre avec peu ou pas d'annotations une tâche particulière* (ex : détecter des classes spécifiques d'objets) *dans un domaine cible spécifique* (ex : une modalité d'imagerie visible, IR, un contexte, un environnement, un point de vue particulier...) ?

Une stratégie consiste à utiliser directement un modèle d'IA appris par supervision sur des données annotées dans un domaine source, pour effectuer la même tâche dans le domaine cible. Par exemple, un détecteur d'objets appris sur le jeu d'images COCO permettra de détecter et reconnaître 80 types d'objets sur des images de type photos, semblables à ce jeu d'images. Mais si le contexte dans lequel on applique ce détecteur diffère, les performances de détection chuteront très probablement. Pour éviter cela, les jeux d'images sont généralement très grands pour incorporer une grande diversité d'instances des concepts à apprendre. Le but est alors d'apprendre un modèle le plus *générique* possible qui pourra s'appliquer à n'importe quel domaine.

Au contraire, une autre stratégie consiste à *spécialiser* le modèle pour l'adapter au mieux au domaine cible. Des méthodes d'adaptation de domaine non supervisée (*unsupervised domain adaptation*) permettent ainsi d'améliorer les performances du modèle sur le domaine cible. Par exemple, un modèle de réidentification de personnes ou de voitures entraîné sur des images de synthèse peut être adapté sans aucune supervision pour réidentifier des voitures sur des images réelles [Dubourvieux21, Dubourvieux22]. De même pour un modèle de segmentation [Hoyer22] (cf. Fig.1).

Une autre stratégie consiste à utiliser des données d'un domaine source non annotées pour la tâche cible mais très nombreuses afin d'apprendre des représentations génériques qui se transféreront bien à la tâche cible dans le domaine cible. C'est le cas, par exemple, d'un pré-entraînement effectué via des tâches prétextes (ex : classification d'images, auto-supervision par méthode contrastive...) sur les millions d'images du jeu ImageNet ou LUPerson pour ensuite apprendre la tâche de réidentification de personnes sur un autre jeu d'images [Fu21].

Il arrive parfois que le domaine cible ainsi que la tâche cibles soient très différents des jeux de données (domaines sources) dont on dispose. Se pose alors la question de savoir s'il est plus efficace de s'appuyer encore sur les connaissances de ces domaines sources ou s'il vaut mieux concentrer l'apprentissage sur les données du domaine cible de manière auto-supervisée. Certaines méthodes de réidentification non supervisées par des données sources présentent ainsi des performances à l'état de l'art [Ge20].

Cette même question se pose lorsque quelques annotations sont disponibles sur les données du domaine cible. Ainsi, des détecteurs d'objets peuvent être entraînés de manière semi-supervisée avec ces quelques annotations [Liu20, Xu21].

En résumé, lorsque l'ensemble d'images cibles non annotées est caractérisé comme appartenant à un domaine différent de l'ensemble d'images sources annotées [Ben-David06, Kashyap21], le choix du paradigme d'apprentissage pour apprendre une tâche donnée peut s'avérer stratégique [Pogrebnjakov21] : Faut-il s'appuyer sur un ou plusieurs ensemble(s) de données source(s) pour apprendre à résoudre la tâche donnée (paradigme de l'adaptation de domaine non supervisée [Zhang21]) ? Faut-il s'appuyer sur un ou des ensemble(s) de données sources non annotés et sur la résolution d'une tâche prétexte (paradigme de l'apprentissage auto-supervisé) ? Faut-il s'appuyer uniquement sur l'ensemble de données cible non annoté pour apprendre à résoudre la tâche donnée (paradigme de l'apprentissage auto-supervisé également) ? Faut-il combiner différentes stratégies ? La question de la stratégie à adopter peut également se poser lorsque l'ensemble d'images cibles contient une très petite partie annotée [Shin21, Wu22].



Figure 1: Exemples de domaines sources (à gauche) et cibles (à droite) pour les tâches de (a) réidentification de personnes (datasets PersonX et Market-1501); (b) réidentification de véhicules (datasets VehicleX et VehicleID); (c) segmentation sémantique (datasets SYNTHIA et BDD100K).

Objectifs de la thèse

L'objectif de cette thèse est d'étudier, pour une tâche cible à résoudre dans un domaine cible, les différentes stratégies d'apprentissage frugal en données annotées et les facteurs qui ont le plus d'impact sur leur efficacité. Notamment, on cherchera à caractériser l'écart interdomaines (*domain discrepancy*) existant entre des données sources annotées disponibles et les données cibles non annotées afin de définir de manière théorique et empirique le choix de la stratégie d'apprentissage à appliquer. Il sera intéressant de déterminer si ce choix dépend également de la taille relative et absolue du jeu d'images cibles et du nombre d'annotations, ou encore s'il dépend de la tâche à résoudre (détection d'objets/réidentification d'objets, ensemble ouvert ou fermé pour les annotations...).

L'idée à terme est de pouvoir prédire une stratégie d'apprentissage en fonction de l'ensemble des données cibles et sources disponibles pour apprendre efficacement à résoudre une tâche de vision par ordinateur et faciliter ainsi un déploiement dans un nouveau contexte.

Références:

- [Ben-David06] Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19.
- [Dubourvieux21] Dubourvieux et al. (2021) Improving Unsupervised Domain Adaptive Re-Identification via Source-Guided Selection of Pseudo-Labeling Hyperparameters. *IEEE ACCESS*.
- [Dubourvieux22] Dubourvieux et al. (2022) A formal approach to good practices in Pseudo-Labeling for Unsupervised Domain Adaptive Re-Identification. *arXiv:2112.12887*
- [Fu21] Fu et al. (2021) Unsupervised Pre-training for Person Re-identification. *CVPR*.
- [Ge20] Yixiao Ge et al. (2020) Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID. *NeurIPS*.
- [Hoyer22] Hoyer et al. (2022) HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation. *arXiv:2204.13132*
- [Kashyap21] Kashyap et al.. (2021, June). Domain Divergences: A Survey and Empirical Analysis. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1830-1849).
- [Liu20] Liu et al. (2020) "Unbiased teacher for semi-supervised object detection." in *ICLR*.
- [Pogrebnjakov21] Pogrebnjakov, N., & Shaghaghian, S. (2021, August). Predicting the Success of Domain Adaptation in Text Similarity. In *Proceedings of the 6th Workshop on Representation Learning for NLP (Repl4NLP-2021)* (pp. 206-212).
- [Shin21] Shin, I., Kim, D. J., Cho, J. W., Woo, S., Park, K., & Kweon, I. S. (2021). Labor: Labeling only if required for domain adaptive semantic segmentation. In *IEEE.CVF ICCV* (pp. 8588-8598).
- [Wu22] Wu, T. H. et al. (2022). D2ADA: Dynamic Density-aware Active Domain Adaptation for Semantic Segmentation. *arXiv preprint arXiv:2202.06484*.
- [Xu21] Xu et al. (2021) "End-to-end semi-supervised object detection with soft teacher." in *ICCV*.
- [Zhang21] Zhang, Y. (2021). A survey of unsupervised domain adaptation for visual recognition. *arXiv preprint arXiv:2112.06745*.



CEA List
Laboratoire de Vision et Apprentissage
pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
<http://www.kalisteo.eu>

Contacts Romaric AUDIGIER
Angelique LOESCH
Tél +33 (0)1 69 08 01 06
E-mail Romaric.audigier@cea.fr
Angelique.loesch@cea.fr

Niveau demandé :	Ingénieur, Master 2
Durée :	3 ans
Rémunération :	entre 1800 € et 2000 €.
Compétences requises : <ul style="list-style-type: none">- Vision par ordinateur- Apprentissage automatique (deep learning)- Reconnaissance de formes- Python- La maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch).	