

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIS-2024-32**
(à rappeler dans toute correspondance)

Lieu : Palaiseau

Département/Dir./Serv. : DTIS/IVA

Tél. : 01 80 38 65 69

Responsable(s) du stage : S. Herbin

Email. : stephane.herbin@onera.fr

DESCRIPTION DU STAGE

Thématique(s) : Intelligence Artificielle et Décision

Type de stage : Fin d'études bac+5 Master 2 Bac+2 à bac+4 Autres

Intitulé : Apprentissage Continu pour l'Adaptation de Modèles Multi-Modaux Vision & Langage

Sujet :

Contexte

Les modèles de réseaux de neurones multi-modaux ont récemment montré des performances impressionnantes dans une variété de tâches associant vision par ordinateur et traitement du langage naturel. Leur capacité d'adaptation à de nouvelles tâches ou à des domaines d'application spécifiques reste cependant un défi. Ce stage vise à explorer des méthodes d'apprentissage continu, c'est-à-dire par mise à jour incrémentale du modèle, pour réaliser cette adaptation.

Mots clés : Intelligence Artificielle, Deep Learning, Apprentissage continu, modèles neuronaux vision & langage

Problématique

Les réseaux de neurones pré-entraînés de représentation d'images ou de texte tels qu'utilisés actuellement en IA (par exemple CLIP « Contrastive Language-Image Pre-training » [1]), souvent dénommés modèles de fondation (« foundational models » [2]), semblent posséder des propriétés inégalées de généralisation et de transfert vers d'autres tâches [3]–[8]. Ces modèles ne sont cependant pas toujours efficaces pour représenter un domaine applicatif spécifique (par exemple les images satellitaires ou médicales sont en général très différentes des images web utilisées pour l'apprentissage des modèles de fondation) et demandent à être adaptées pour rendre compte de manière plus complète et précise de l'information utiles contenue dans les données. L'objectif du stage est de développer une approche permettant de réaliser cette adaptation pour des tâches associant vision et langage comme la capacité de répondre à des questions ouvertes sur le contenu d'une image (« visual question answering ») ou la classification « zero-shot », c'est-à-dire sans données annotées d'apprentissage [9]–[14].

On se propose de suivre une stratégie d'apprentissage continu [15]–[17] qui exploite les données du domaine visé dès qu'elles sont disponibles et met à jour incrémentalement le modèle. L'intérêt d'une approche continue est de ne pas avoir à stocker en mémoire la totalité des données lors de chaque phase d'adaptation; la difficulté est de gérer un flux de données [18], [19] et des annotations ou des récompenses asynchrones en limitant ce que l'on appelle l'oubli catastrophique [20], c'est-à-dire la perte des compétences acquises dans les phases d'apprentissage antérieures.

Deux stratégies d'apprentissage sont envisageables : partir d'une représentation pré-entraînée donnée que l'on vient spécialiser par réglage fin par exemple en apprenant un « prompt » conditionnel [21]–[23], ou construire une nouvelle représentation spécifique au domaine mais aidée par un modèle pré-entraîné qui lui fournit des capacités de généralisation [24], [25]. Les données non annotées pourront également être exploitées par une stratégie semi-supervisée complémentaire [26]–[28].

Travaux envisagés

Après une phase de bibliographie le travail du stage consistera principalement à adapter un article sélectionné dans l'état de l'art, à coder la méthode et à l'évaluer sur les bases de données usuelles du

domaine.

Le stage a vocation à se poursuivre en thèse.

Bibliographie indicative

- [1] A. Radford et al., « Learning Transferable Visual Models From Natural Language Supervision », in Proceedings of the 38th International Conference on Machine Learning, PMLR, juill. 2021
- [2] R. Bommasani et al., « On the Opportunities and Risks of Foundation Models ». arXiv, 12 juillet 2022. doi: 10.48550/arXiv.2108.07258.
- [3] L. Ericsson, H. Gouk, et T. M. Hospedales, « How Well Do Self-Supervised Models Transfer? », présenté à Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [4] E. Fini, V. G. T. da Costa, X. Alameda-Pineda, E. Ricci, K. Alahari, et J. Mairal, « Self-Supervised Models Are Continual Learners », Conference on Computer Vision and Pattern Recognition, 2022,
- [5] L. Jing et Y. Tian, « Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey », IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no 11, 2021,.
- [6] M. Mendieta, B. Han, X. Shi, Y. Zhu, et C. Chen, « Towards Geospatial Foundation Models via Continual Pretraining ». arXiv, 31 août 2023. doi: 10.48550/arXiv.2302.04476.
- [7] O. Ostapenko et al., « Continual Learning with Foundation Models: An Empirical Study of Latent Replay ». Disponible sur: <http://arxiv.org/abs/2205.00329>
- [8] V. Thengane, S. Khan, M. Hayat, et F. Khan, « CLIP model is an Efficient Continual Learner ». arXiv,2022. doi: 10.48550/arXiv.2210.03114.
- [9] R. D. Chiaro, B. Twardowski, A. D. Bagdanov, et J. van de Weijer, « RATT: Recurrent Attention to Transient Tasks for Continual Image Captioning », ArXiv, juill. 2020
- [10] A. Lechat, « Apprentissage incrémental semi-supervisé pour les applications de vision artificielle », These de doctorat, Normandie, 2021.Disponible sur: <https://www.theses.fr/2021NORMC256>
- [11] A. Lechat, S. Herbin, et F. Jurie, « Adaptation du problème de questions-réponses visuelles à un contexte d'apprentissage continu », in actes du GRETSI, Lille, 2019.
- [12] M. Nikandrou, L. Yu, A. Suglia, I. Konstas, et V. Rieser, « Task Formulation Matters When Learning Continually: A Case Study in Visual Question Answering », 2022, doi: 10.48550/ARXIV.2210.00044.
- [13] X. Zhang, F. Zhang, et C. Xu, « VQACL: A Novel Visual Question Answering Continual Learning Setting », IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023
- [14] Z. Zheng, M. Ma, K. Wang, Z. Qin, X. Yue, et Y. You, « Preventing Zero-Shot Transfer Degradation in Continual Learning of Vision-Language Models ». arXiv, 11 août 2023. doi: 10.48550/arXiv.2303.06628.
- [15] E. Belouadah, A. Popescu, et I. Kanellos, « A comprehensive study of class incremental learning algorithms for visual tasks », Neural Networks, vol. 135, p. 38-54, 2021.
- [16] J. A. Mendez et E. Eaton, « How to Reuse and Compose Knowledge for a Lifetime of Tasks: A Survey on Continual Learning and Functional Composition ». arXiv, 2022. doi: 10.48550/arXiv.2207.07730.
- [17] L. Wang, X. Zhang, H. Su, et J. Zhu, « A Comprehensive Survey of Continual Learning: Theory, Method and Application ». arXiv, 31 janvier 2023. doi: 10.48550/arXiv.2302.00487.
- [18] N. Gunasekara, B. Pfahringer, H. M. Gomes, et A. Bifet, « Survey on Online Streaming Continual Learning », IJCAI; 2023
- [19] Z. Mai, R. Li, J. Jeong, D. Quispe, H. Kim, et S. Sanner, « Online Continual Learning in Image Classification: An Empirical Survey », arXiv preprint arXiv:2101.10423, 2021.
- [20] R. M. French, « Catastrophic forgetting in connectionist networks », Trends in Cognitive Sciences, vol. 3, no 4, p. 128-135, 1999, doi: [https://doi.org/10.1016/S1364-6613\(99\)01294-2](https://doi.org/10.1016/S1364-6613(99)01294-2).
- [21] J. Gu et al., « A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models ». arXiv, 24 juillet 2023. doi: 10.48550/arXiv.2307.12980.
- [22] S. W. Lei et al., « Symbolic Replay: Scene Graph as Prompt for Continual Learning on VQA Task », Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no 1, Art. no 1, juin 2023.
- [23] Z. Wang et al., « Learning to Prompt for Continual Learning ». arXiv, 21 mars 2022. doi:

10.48550/arXiv.2112.08654.

[24] Z. Wang et al., « CLIP-TD: CLIP Targeted Distillation for Vision-Language Tasks ». arXiv, 28 décembre 2022. doi: 10.48550/arXiv.2201.05729.

[25] C. Yang et al., « CLIP-KD: An Empirical Study of Distilling CLIP Models ». arXiv, 24 juillet 2023. doi: 10.48550/arXiv.2307.12732.

[26] J. Gallardo, T. L. Hayes, et C. Kanan, « Self-Supervised Training Enhances Online Continual Learning », arXiv:2103.14010 [cs], oct. 2021

[27] A. Lechat, S. Herbin, et F. Jurie, « Pseudo-Labeling for Class Incremental Learning », in BMVC 2021: The British Machine Vision Conference, 2021.

[28] A. Lechat, S. Herbin, et F. Jurie, « Semi-Supervised Class Incremental Learning », in ICPR, 2020.

Est-il possible d'envisager un travail en binôme ? **Non**

Méthodes à mettre en oeuvre :

- | | |
|---|--|
| <input type="checkbox"/> Recherche théorique | <input type="checkbox"/> Travail de synthèse |
| <input checked="" type="checkbox"/> Recherche appliquée | <input type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale | <input type="checkbox"/> Participation à une réalisation |

Possibilité de prolongation en thèse : **Oui**

Durée du stage : Minimum : 4 mois Maximum : 6 mois

Période souhaitée : Printemps-Eté 2024

PROFIL DU STAGIAIRE

Connaissances et niveau requis : - Connaissance en apprentissage automatique, vision par ordinateur et traitement du langage naturel - Expérience en programmation Python et en utilisation de bibliothèques telles que PyTorch ou TensorFlow.	Ecoles ou établissements souhaités : Ecole d'ingénieur ou Master 2 Recherche avec spécialisation en intelligence artificielle ou apprentissage automatique.
--	--