

PROPOSITION DE SUJET DE THESE

Intitulé : Apprentissage continu pour l'acquisition en ligne de capacités d'interprétation sémantique d'un système perceptif dynamique

Référence : **TIS-DTIS-2024-27**
(à rappeler dans toute correspondance)

Début de la thèse : Octobre 2024

Date limite de candidature : Mai 2024

Mots clés

Intelligence artificielle, Apprentissage continu, apprentissage actif, apprentissage en monde ouvert, systèmes perceptifs, modèles neuronaux vision & langage, apprentissage par renforcement, robotique

Profil et compétences recherchées

Ecole d'ingénieur ou Master 2 avec spécialisation en intelligence artificielle, vision par ordinateur et/ou robotique.

Présentation du projet doctoral, contexte et objectif

On s'intéresse dans cette thèse à la conception de systèmes perceptifs visuels ayant la capacité d'interagir avec leur environnement pour acquérir de l'information ou des données et réaliser des tâches d'interprétation, c'est-à-dire produire une description textuelle pertinente et utile de la scène observée. De telles tâches peuvent être utilisées typiquement pendant une exploration de l'environnement par un robot à la recherche d'un objet ou d'une direction de déplacement sans obstacle et potentiellement aidé par un agent distant avec lequel échanger des informations ou des connaissances.

Pour construire un tel système perceptif on se propose de développer une approche par apprentissage automatique incrémental (De Lange et al., 2022; Masana et al., 2023; van de Ven et al., 2022; L. Wang et al., 2023; Zhou et al., 2023) pour acquérir progressivement de nouvelles capacités sémantiques ou étendre le domaine d'utilisation. La difficulté d'une telle approche est d'intégrer au système ces nouvelles compétences sans oublier celles qui avaient été acquises précédemment (French, 1999).

Les systèmes perceptifs que l'on envisage interagissent avec leur environnement, soit pour acquérir de nouvelles données perceptives, soit pour demander à un agent externe des informations ou des précisions sur le contenu utile de la scène observée: ils combinent potentiellement exploration active et prise d'information sémantique pour s'adapter à un nouveau contexte, ou apprendre de nouvelles capacités d'interprétation. L'un des objectifs de la thèse sera de réaliser cette adaptation par un apprentissage incrémental en ligne (Gallardo et al., 2021; Gu et al., 2022; Gunasekara et al., 2023; Mai et al., 2022) en pilotant activement les sources de données (Aloimonos, 2013; Ayub and Fendley, 2022; Cheng et al., 2018) ou d'information (« active learning » (Aggarwal et al., 2020; Buck et al., 2018; Ren et al., 2021; Teso et al., 2023; Vo et al., 2022; Zhan et al., 2022), apprentissage en monde ouvert (Boult et al., 2019; Joseph et al., 2021; Kim et al., 2022; Z. Wang et al., 2023; Zohar et al., 2023)). Les stratégies de contrôles de sources de données et d'information pourront être réalisées par des approches d'apprentissage par renforcement (Khetarpal et al., 2022; Kumar et al., 2023).

Une évolution récente des techniques d'intelligence artificielle est le développement de modèles neuronaux multi-modaux généralistes tels que CLIP (Cherti et al., 2022; Liang et al., 2022; Radford et al., 2021) qui permettent d'associer vision et langage dans un espace de représentation commun pour y produire des inférences d'interprétation sémantique. Cette gamme de modèles possède des capacités de généralisation et d'association remarquables mais des domaines de compétence et d'utilisation mal cernés : ils demandent une étape d'adaptation pour être utilisés dans un contexte applicatif plus spécialisé (Hu et al., 2023). Les modèles multi-modaux ont également la capacité

d'introduire un contexte sous la forme de « prompt » (Gu et al., 2023a, 2023b; Jia et al., 2022) conditionnant globalement les traitements pour les adapter à différentes situations. Un autre objectif de la thèse sera d'utiliser ces modèles multi-modaux pour les systèmes perceptifs.

Les approches seront principalement développées et validées à partir d'environnements de simulation récents produisant des données réalistes pour diverses conditions d'acquisition contrôlables (Gong et al., 2023; Li et al., 2022; Qiu et al., 2023; Zhu et al., 2023).

En résumé, l'ambition de la thèse est de concevoir une démarche d'apprentissage incrémental permettant à un système perceptif de piloter activement les sources de données et d'information utiles pour acquérir de nouvelles capacité d'interprétation en exploitant des modèles neuronaux multi-modaux.

Bibliographie indicative

- Aggarwal, U., Popescu, A., Hudelot, C., 2020. Active Learning for Imbalanced Datasets. Presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1428–1437.
- Aloimonos, Y., 2013. Active perception. Psychology Press.
- Ayub, A., Fendley, C., 2022. Few-Shot Continual Active Learning by a Robot. Advances in Neural Information Processing Systems 35, 30612–30624.
- Boult, T., Cruz, S., Dhamija, A., Gunther, M., Henrydoss, J., Scheirer, W., 2019. Learning and the unknown: Surveying steps toward open world recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 9801–9807.
- Buck, C., Bulian, J., Ciaramita, M., Gajewski, W., Gesmundo, A., Housby, N., Wang, W., 2018. Ask the Right Questions: Active Question Reformulation with Reinforcement Learning, in: International Conference on Learning Representations.
- Cheng, Y., Gan, Z., Li, Y., Liu, J., Gao, J., 2018. Sequential attention gan for interactive image editing via dialogue. arXiv preprint arXiv:1812.08352.
- Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L., Jitsev, J., 2022. Reproducible scaling laws for contrastive language-image learning. <https://doi.org/10.48550/arXiv.2212.07143>
- De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., Tuytelaars, T., 2022. A Continual Learning Survey: Defying Forgetting in Classification Tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 3366–3385. <https://doi.org/10.1109/TPAMI.2021.3057446>
- French, R.M., 1999. Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences 3, 128–135. [https://doi.org/10.1016/S1364-6613\(99\)01294-2](https://doi.org/10.1016/S1364-6613(99)01294-2)
- Gallardo, J., Hayes, T.L., Kanan, C., 2021. Self-Supervised Training Enhances Online Continual Learning. arXiv:2103.14010 [cs].
- Gong, R., Huang, J., Zhao, Y., Geng, H., Gao, X., Wu, Q., Ai, W., Zhou, Z., Terzopoulos, D., Zhu, S.-C., Jia, B., Huang, S., 2023. ARNOLD: A Benchmark for Language-Grounded Task Learning With Continuous States in Realistic 3D Scenes. <https://doi.org/10.48550/arXiv.2304.04321>
- Gu, J., Beirami, A., Wang, X., Beutel, A., Torr, P., Qin, Y., 2023a. Towards Robust Prompts on Vision-Language Models. <https://doi.org/10.48550/arXiv.2304.08479>
- Gu, J., Han, Z., Chen, S., Beirami, A., He, B., Zhang, G., Liao, R., Qin, Y., Tresp, V., Torr, P., 2023b. A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models. <https://doi.org/10.48550/arXiv.2307.12980>
- Gu, Y., Yang, X., Wei, K., Deng, C., 2022. Not Just Selection, but Exploration: Online Class-Incremental Continual Learning via Dual View Consistency. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7442–7451.
- Gunasekara, N., Pfahringer, B., Gomes, H.M., Bifet, A., 2023. Survey on Online Streaming Continual Learning. Presented at the Thirty-Second International Joint Conference on Artificial Intelligence, pp. 6628–6637. <https://doi.org/10.24963/ijcai.2023/743>
- Hu, X., Zhang, K., Xia, L., Chen, A., Luo, J., Sun, Y., Wang, K., Qiao, N., Zeng, X., Sun, M., Kuo, C.-H., Nevatia, R., 2023. ReCLIP: Refine Contrastive Language Image Pre-Training with Source Free Domain Adaptation.
- Jia, M., Tang, L., Chen, B.-C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.-N., 2022. Visual Prompt Tuning, in: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (Eds.), Computer Vision – ECCV 2022, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 709–727. https://doi.org/10.1007/978-3-031-19827-4_41
- Joseph, K.J., Khan, S., Khan, F.S., Balasubramanian, V.N., 2021. Towards Open World Object Detection. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5830–5840.
- Khetarpal, K., Riemer, M., Rish, I., Precup, D., 2022. Towards Continual Reinforcement Learning: A Review and Perspectives. Journal of Artificial Intelligence Research 75, 1401–1476. <https://doi.org/10.1613/jair.1.13673>

- Kim, D., Lin, T.-Y., Angelova, A., Kweon, I.S., Kuo, W., 2022. Learning Open-World Object Proposals Without Learning to Classify. *IEEE Robotics and Automation Letters* 7, 5453–5460. <https://doi.org/10.1109/LRA.2022.3146922>
- Kumar, S., Marklund, H., Rao, A., Zhu, Y., Jeon, H.J., Liu, Y., Van Roy, B., 2023. Continual Learning as Computationally Constrained Reinforcement Learning. <https://doi.org/10.48550/arXiv.2307.04345>
- Li, C., Zhang, R., Wong, J., Gokmen, C., Srivastava, S., Martín-Martín, R., Wang, C., Levine, G., Lingelbach, M., Sun, J., Anvari, M., Hwang, M., Sharma, M., Aydin, A., Bansal, D., Hunter, S., Kim, K.-Y., Lou, A., Matthews, C.R., Villa-Renteria, I., Tang, J.H., Tang, C., Xia, F., Savarese, S., Gweon, H., Liu, K., Wu, J., Fei-Fei, L., 2022. BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation. Presented at the 6th Annual Conference on Robot Learning.
- Liang, W., Zhang, Y., Kwon, Y., Yeung, S., Zou, J., 2022. Mind the gap: Understanding the modality gap in multi-modal contrastive representation learning, in: NeurIPS.
- Mai, Z., Li, R., Jeong, J., Quispe, D., Kim, H., Sanner, S., 2022. Online continual learning in image classification: An empirical survey. *Neurocomputing* 469, 28–51. <https://doi.org/10.1016/j.neucom.2021.10.021>
- Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A.D., van de Weijer, J., 2023. Class-Incremental Learning: Survey and Performance Evaluation on Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 5513–5533. <https://doi.org/10.1109/TPAMI.2022.3213473>
- Qiu, Y., Nagasaki, Y., Hara, K., Kataoka, H., Suzuki, R., Iwata, K., Satoh, Y., 2023. VirtualHome Action Genome: A Simulated Spatio-Temporal Scene Graph Dataset With Consistent Relationship Labels. Presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3351–3360.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning Transferable Visual Models From Natural Language Supervision, in: Proceedings of the 38th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 8748–8763.
- Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B.B., Chen, X., Wang, X., 2021. A Survey of Deep Active Learning. *ACM Comput. Surv.* 54, 180:1–180:40. <https://doi.org/10.1145/3472291>
- Teso, S., Alkan, Ö., Stammer, W., Daly, E., 2023. Leveraging explanations in interactive machine learning: An overview. *Frontiers in Artificial Intelligence* 6.
- van de Ven, G.M., Tuytelaars, T., Tolias, A.S., 2022. Three types of incremental learning. *Nat Mach Intell* 1–13. <https://doi.org/10.1038/s42256-022-00568-3>
- Vo, H.V., Siméoni, O., Gidaris, S., Bursuc, A., Pérez, P., Ponce, J., 2022. Active Learning Strategies for Weakly-supervised Object Detection. <https://doi.org/10.48550/arXiv.2207.12112>
- Wang, L., Zhang, X., Su, H., Zhu, J., 2023. A Comprehensive Survey of Continual Learning: Theory, Method and Application. <https://doi.org/10.48550/arXiv.2302.00487>
- Wang, Z., Li, Y., Chen, X., Lim, S.-N., Torralba, A., Zhao, H., Wang, S., 2023. Detecting Everything in the Open World: Towards Universal Object Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11433–11443.
- Zhan, X., Wang, Q., Huang, K., Xiong, H., Dou, D., Chan, A.B., 2022. A Comparative Survey of Deep Active Learning. <https://doi.org/10.48550/arXiv.2203.13450>
- Zhou, D.-W., Wang, Q.-W., Qi, Z.-H., Ye, H.-J., Zhan, D.-C., Liu, Z., 2023. Deep Class-Incremental Learning: A Survey. <https://doi.org/10.48550/arXiv.2302.03648>
- Zhu, H., Kapoor, R., Min, S.Y., Han, W., Li, J., Geng, K., Neubig, G., Bisk, Y., Kembhavi, A., Weihs, L., 2023. EXCALIBUR: Encouraging and Evaluating Embodied Exploration. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14931–14942.
- Zohar, O., Wang, K.-C., Yeung, S., 2023. PROB: Probabilistic Objectness for Open World Object Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11444–11453.

Collaborations envisagées

Laboratoire d'accueil à l'ONERA

Département : Traitement de l'Information et Systèmes

Lieu (centre ONERA) : Palaiseau

Contact : Stéphane Herbin

Tél. : 01 80 38 65 69 Email : stephane.herbin@onera.fr

Directeur de thèse

Nom : Stéphane HERBIN

Laboratoire : DTIS

Tél. : 01 80 38 65 69

Email : stephane.herbin@onera.fr

Pour plus d'informations : <https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche>