S03 – Open-Set Object Detection: Can VLM understand all of the unknown world?

6-month internship @ CEA List

Internship context

Based in Saclay (Essonne), the LIST is one of the two institutes of CEA Tech, the technological research division of the CEA. Dedicated to intelligent digital systems, its mission is to carry out technological developments of excellence on behalf of industrial partners in order to create value.

Within the LIST, the Laboratory of Vision and Learning for Scene Analysis (LVA) conducts research in the field of computer vision and artificial intelligence for the perception of intelligent and autonomous systems. The laboratory's research themes include visual recognition, behavior and activity analysis, large-scale automatic annotation, and perception and decision models. These technologies are applied in major sectors such as security, mobility, advanced manufacturing, healthcare, and sports...

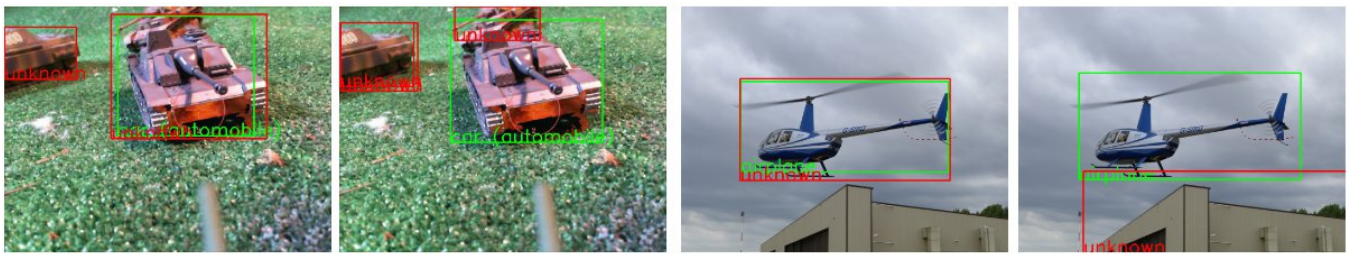
Missions

In the last decade, deep learning models have demonstrated significant success to detect a close set of known annotated classes. However, their behaviour is uncertain when subjected to never-before-seen classes. Being able to localise and classify an object from new classes different from those known during training, is a challenge called Open Set Object Detection (OSOD). It is of paramount importance to ensure safe deployment of AI models as trusted AI is crucial in critical applications such as autonomous driving. Indeed, one of the most important issues for these applications is the lack of supervisory signals on data not seen in training, which can generate erroneous but confident predictions on out-of-distribution (OOD) data [1].

Example of multiple object detection paradigms

Works in OSOD demonstrated different behaviours of probability-based methods (contrastive learning) [3] and pseudo-labeling based approaches [2]. Recently, Visual Language Large Models (VLLM) have shown tremendous improvements for several computer vision applications, while being computationally expensive. It is important to study their capacities in the OSOD context and their efficiency when faced to new domains.

In fact, the behaviour of OSOD methods under more realistic conditions, such as those of the open world, and possible confrontation with new specific contexts, such as those of aerial images [5], have not been studied. These operational constraints, therefore pose new challenges.



OpenDet *OW – DETR++* *OpenDet* *OW – DETR++*

Example of OSOD detection from OpenDet [3] and OW–DETR++ (ours) [2]

Internship objectives

Several areas of study within OSOD will be explored in this internship:

- Study state-of-the art of OSOD approaches;
- Study the efficiency of recent VLM in the context of Open World;
- Design an object detector aware of the existence of the unknown;
- Evaluate them on recent benchmarks under different scenarios;
- Apply these methods to new contexts (e.g. areal images);
- Depending on the obtained results, the contributions of this internship may lead to an international conference or workshop publication.

References

[1] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In IEEE CVPR, 2015.

[2] Hejer Ammar, Nikita Kiselov, Guillaume Lapouge, Romaric Audigier, Open-set object detection: towards unified problem formulation and benchmarking, The 3rd workshop of Out-Of Distribution Generalization in Computer Vision Foundation Models, ECCVW24

[3] Han, Jiaming and Ren, Yuqiang and Ding, Jian and Pan, Xingjia and Yan, Ke and Xia, Gui-Song, Expanding low-density latent regions for open-set object detection, CVPR 2022

[4] Gupta, Akshita and Narayan, Sanath and Joseph, KJ and Khan, Salman and Khan, Fahad Shahbaz and Shah, Mubarak, Ow-detr: Open-world detection transformer, CVPR 2022

[5] XIA, Gui-Song, BAI, Xiang, DING, Jian, et al. DOTA: A large-scale dataset for object detection in aerial images. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 3974-3983.

Qualifications

- Students in their 5th year of studies (M2)
- Computer vision skills
- Machine learning skills (deep learning, LLM, VLM...)
- Python proficiency in a deep learning framework (especially PyTorch or TensorFlow)

Job-related benefits

Join CEA List and LVA as an intern to:

- Work in one of the most innovative research organizations in the world (ranked in the global top 100, top 3 in France), addressing societal challenges to build the world of tomorrow
- Discover a rich ecosystem: privileged connections between the industrial and academic sectors
- Conduct research in an environment where autonomy and creativity are recognized, and where valorizing results is encouraged (publication of scientific articles, patents, and sharing of open-source code whenever possible).
- Join a young and dynamic team made up of research engineers, PhD students, post-doctoral researchers, and interns.
- Benefit from an internal computing infrastructure equipped with around 300 state-of-the-art GPUs.
- Receive a stipend between €1300 and €1400 per month.
- Have the opportunity to continue with a PhD or as a research engineer after the internship.
- Have the possibility of remote work.
- Receive a 75% (instead of 50%) reimbursement on public transportation costs, and benefit from the "mobili-jeune" aid to reduce rent costs...

To apply, contact the laboratory with a CV and cover letter: lva-stages@cea.fr