PHD - Object Detection from Few Multispectral Examples

Academic lab: IRISA (https://www.irisa.fr/)
Compagny: ATERMES (https://www.atermes.fr/)
CIFRE PHD

Context

ATERMES is an international mid-sized company, based in Montigny-le-Bretonneux with a strong expertise in high technology and system integration from the upstream design to the long-life maintenance cycle. It specializes in offering system solution for border surveillance. Its flagship product BARIER(TM) (“Beacon Autonomous Reconnaissance Identification and Evaluation Response”) provides ready application for temporary strategic site protection or ill-defined border regions in mountainous or remote terrain where fixed surveillance modes are impracticable or overly expensive to deploy. As another example, SURICATE is the first of its class optronic ground "RADAR" that covers very efficiently wide field with automatic classification of intruders thanks to multi-spectral deep learning detection.

The collaboration between ATERMES and IRISA was initiated through a first PhD thesis (Heng Zhang, defended December 2021, https://www.theses.fr/2021REN1S099/document). This successful collaboration led to multiple contributions on object detection in both mono-modal (RGB) and multi-modal (RGB+THERMAL) scenarios. Besides, this study allowed to identify remaining challenges that need to be solved to ensure multispectral object detection in the wild.

Objectives

The project aims at providing deep learning-based methods to detect objects in outdoor environments using multispectral data in a low supervision context, e.g., learning from few examples to detect scarcely-observed objects. The data consist of RGB and IR (Infra-red) images which are frames from calibrated and aligned multispectral videos.

Few-shot learning [1][2], semi-supervised learning [3][4] and continual learning [5][6] are among the most widely-used frameworks to tackle this task. For the first approach based on few-shot object detection (FSOD), the recent trend has relied on using meta learning or transfer learning approaches [1]. Yet, realistic settings including scarce objects may exist a domain shift that makes the task more challenging. The second approach based on semi-supervised learning considers a large amount of unlabeled data in the training process to foster the representation capacity of deep models, improving the peformance of object detection from a small amount of labeled samples. As the third approach, continual learning [5] aims to maintain the performance of the deep models on old categories and avoid the "catastrophic forgetting" phenomenon when learning new object categories. It has been also integrated into a FSOD task [7] to ensure that few-shot object detectors could learn new object concepts without forgetting previous object categories that still exist in prediction phase. Last but not least, with the dramastically rapid evolution of research in AI, another challenge to tackle is the investigation of modern AI models, and more specifically foundation models which involves multimodal transformers [8][9]. Indeed, these large machine learning models trained on a vast quantity of data at scale have been designed to be adapted to a wide range of downstream tasks (including object detection, see for instance UniDetector [10]) or CLIP2 [11]. These models leading to zero-shot object detection could very well be the ultimate answer for the task of having a true scene understanding.

It should be noted that most developed methods dealing with object detection in low-supervision scenarios based on the above approaches have been proposed to work with RGB images in the computer vision domain. They should be adapted and improved to deal with scarce object detection from multispectral images. To tackle the case where objects of interest are also rare at train time, anomaly detection approaches [12][13] can be also be considered and combined with *a priori* expert knowledge to identify new object classes.

In addition to the (private) data from ATERMES, the PhD candidate will be able to work with public datasets such as KAIST (https://soonminhwang.github.io/rgbt-ped-detection/data/), FLIR (https://drive.google.com/file/d/1xHDMGl6HJZwtarNWkEV3T4O9X4ZQYz2Y/view), [FLIRV2] (https://www.kaggle.com/datasets/samdazel/teledyne-flir-adas-thermal-dataset-v2), VEDAI (https://downloads.greyc.fr/vedai/) or MIL (https://www.mi.t.u-tokyo.ac.jp/static/projects/mil_multispectral/) to benchmark the developed frameworks and compare them to the state-of-the-art in the vision and machine learning communities.

In addition the model developped are intended to be integrated into the equipments (intelligent cameras) and thus process in almost real time the images.

Working Plan

The PhD candidate will work part time (80%) at IRISA (with 1 day per week in Rennes and the rest of the time in the Vannes IRISA facility) and part time (20%) in ATERMES in Paris (which corresponds to 2 days every 2 weeks). The exact schedule will be flexible: it might be preferable to spend more time in the company at the beginning of the thesis to learn about the system and understand the data and be full time in the lab while writing the PhD dissertation.

- T0-T0+8: The PhD candidate will survey the recent literature about deep learning under low supervision scenarios in the broad sense, with a specific focus on methods adapted to the (multispectral) object detection problem.
- T0+9 - T0+24: During this period, the candidate will propose original contributions to tackle the problem of low supervision for multispectral object detection. We expect contributions related to few-shot/zero-shot learning, incremental and/or continual learning.
- T0+24 – T0+32: During this period, the candidate will integrate its contributions to the system developed by ATERMES.
- T0+33 - T0+ 36: The last period will be dedicated to writing the PhD dissertation

Required background and skills

- MSc or Engineering degree with excellent academic track and proven research experience in the following fields: computer science, applied maths, signal processing and computer vision;
- Experience with machine learning, in particular deep learning;
- Skills and **proved experience** in programming (Python is mandatory and knowledge about frameworks such as Pytorch is a real plus);
- Excellent communication skills (spoken/written English) is required ;
- Ambition to publish at the best level in the computer vision community (CVPR, ICCV, TPAMI, ...) during the thesis.

Supervision team

The PhD will be co-supervised by Prof. Elisa Fromont (LACODAM team, IRISA/INRIA Rennes) and Prof. Sébastien Lefèvre (OBELIX team, IRISA Vannes). The supervision team will be completed by Dr. Minh-Tan Pham (Ass. Prof., OBELIX team) and Bruno Avignon (CSO, ATERMES).

Application Procedure

Your application (CV+cover letter+academic transcripts) should be sent **before 31/12/2023 (but the sooner the better)** to the 4 email addresses:
elisa.fromont@irisa.fr; sebastien.lefevre@irisa.fr; minh-tan.pham@irisa.fr; bavignon@atermes.fr

Applications will be treated and interviews will be conducted along the way.
The candidate will be hired with a CIFRE (https://www.anrt.asso.fr/fr/le-dispositif-cifre-7844) contract by ATERMES.
The expected gross salary is around 3500€ per month for 3 years.
The contract will start at the beginning of 2024. Atermes can hire the candidate (as an engineer, CDI) before the beginning of the CIFRE contract if necessary.

Bibliography

---

1. Köhler, M., Eisenbach, M. and Gross, H.M. (2023) Few-Shot Object Detection: A Comprehensive Survey (https://doi.org/10.1109/TNNLS.2023.3265051). IEEE Transactions on Neural Networks and Learning Systems, early access

2. Huang, G., Laradji, I., Vazquez, D., Lacoste-Julien, S. and Rodriguez, P. (2022) A Survey of Self-Supervised and Few-Shot Object Detection (https://doi.org/10.1109/TPAMI.2022.3199617). IEEE Transactions on Pattern Analysis and Machine Intelligence, 45:4, 4071-4089

3. Wang, Y., Liu, Z., & Lian, S. (2023) Semi-supervised Object Detection: A Survey on Recent Research and Progress (https://doi.org/10.48550/arXiv.2306.14106). arXiv preprint arXiv:2306.14106

4. Chen, B., Li, P., Chen, X., Wang, B., Zhang, L., & Hua, X. S. (2022) Dense learning based semi-supervised object detection (https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.00477). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4805-4814

5. De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G. and Tuytelaars, T. (2022) A continual learning survey: Defying forgetting in classification tasks (https://doi.org/10.1109/TPAMI.2021.3057446). IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7):3366-3385

6. Menezes, A.G., Gustavo de Moura, C.A., and de Carvalho, A. (2023) Continual Object Detection: A review of definitions, strategies, and challenges (https://doi.org/10.1016/j.neunet.2023.01.041). Neural Networks, 161:476-493

7. Fan, Z., Ma, Y., Li, Z. and Sun, J. (2021) Generalized few-shot object detection without forgetting (https://doi.ieeecomputersociety.org/10.1109/CVPR46437.2021.00450). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4525-4534

8. Zhang, R., Hu, X., Li, B., Huang, S., Deng, H., Qiao, Y., Gao, P. and Li, H., (2023). Prompt, generate, then cache: Cascade of foundation models makes strong few-shot learners (https://openaccess.thecvf.com/content/CVPR2023/papers/Zhang_Prompt_Generate_Then_Cache_

In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15211-15222).

9. Li, J., Li, D., Savarese, S. and Hoi, S., (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. (https://arxiv.org/pdf/2301.12597.pdf) arXiv preprint arXiv:2301.12597.

10. 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 (https://doi.org/10.48550/arXiv.2303.11749). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11433-11443

11. Zeng, Y., Jiang, C., Mao, J., Han, J., Ye, C., Huang, Q., Yeung, D.Y., Yang, Z., Liang, X. and Xu, H., 2023. CLIP2: Contrastive Language-Image-Point Pretraining from Real-World Point Cloud Data (https://openaccess.thecvf.com/content/CVPR2023/html/Zeng_CLIP2_Contrastive_Language-Image-Point_Pretraining_From_Real-World_Point_Cloud_Data_CVPR_2023_paper.html) In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15244-15253).

12. Gangloff, H., Pham, M.T., Courtrai, L. and Lefèvre, S. (2022) Variational Autoencoder with Gaussian Random Field prior: application to unsupervised animal detection in aerial images (https://hal.science/hal-03774853/document). HAL preprint hal-03774853

13. Gangloff, H., Pham, M.T., Courtrai, L. and Lefèvre, S. (2022) Leveraging Vector-Quantized Variational Autoencoder Inner Metrics for Anomaly Detection (https://ieeexplore.ieee.org/document/9956102). In Proceedings of the 26th International Conference on Pattern Recognition (ICPR), 435-441