Topic of the PhD Thesis:	**A Multi-Task Neural Network for Real Time Object Detection and Tracking on Embedded Systems**
Key Words :	AI, neural network, multi-task, perception, object detection, tracking, embedded

Context

With the increasing demand for automated scene analysis, visual object tracking is a field in full development for intelligent transportation systems, video surveillance, medical imaging or human-computer interaction. Real-time object tracking involves many difficulties related to frequent occlusions, similarity of object appearance, potential change in object shape, lighting, or camera movement. Tracking algorithms can be classified into two distinct groups: detection free and tracking-by-detection approaches. Most modern multi-object tracking systems deal with the tracking-by-detection paradigm consisting of an object detector followed by a method for associating detections into tracks over frames. The choice of the appropriate matching method comes with the trade-off between method complexity and computing resources. Recent successes in popular 2D tracking benchmarks [16,17] indicate that top-scores can be achieved using a state-of-the-art detector and relatively simple associations. Nevertheless, matching solutions include, more and more often, additional information about learned appearance features (embedding), motion prediction or object's interaction to facilitate object association and re-identification (ReID) of lost tracks. Experiments show that the motion prediction model improves a tracking performance, especially when the cameras are moving fast, when people's poses are significantly deformed, or when objects have similar appearance (*e.g.* pedestrians with similar clothing at different locations in consecutive frames). According to the different implementations of these parts, *i.e.* detection, embedding extraction or motion prediction, tracking approaches can be divided into SDE, JDE and JDT series algorithms:

- Separate Detection and Embedding (SDE) completely separates detection from embedding extraction. Each of these computationally intensive tasks is performed sequentially. This design means that the system can fit any kind of detector without distinction and can be improved for each part separately (*e.*g. DeepSORT [1,2]);
- Joint Detection and Embedding (JDE), simultaneously generates the location and appearance features of targets in a single forward pass. It learns detection and ReID embedding in a shared neural network, and sets the loss function with a multi-task learning approach (*e.g.* JDE [3], FairMOT [4] or RetinaTrack [8]);
- Joint Detection and Tracking (JDT) extends JDE approach by co-similarity measure between tracked objects in past frames and detected new ones detected. This type of algorithm can also propose a network for joint detection and movement prediction (*e.g.* DEFT [5], Tracktor [9], CenterTrack [10], SiamMOT[11]).

In recent years, there has been remarkable progress of CNN-based object detection, ReID embedding's extraction and motion prediction. However, a little attention has been focused on combining several tasks in end-to-end architecture to enhance association of object sharing similar appearance and to reduce the inference time. Few existing algorithms are robust but cannot run easily in real time embedded systems with a low power and memory capacity. Consequently, state-of-the-art object-tracking approaches that meet embedded device requirements adopt rather SDE strategy [14], or periodically re-initialize object detector [12,13]. Then, to ensure tracking for the intermediate frames, the detector is coupled to a generic visual tracker (detection free approach) like correlation filter-based or optical flow-based.

Summary of the Thesis Topic

The LIAE lab at CEA has a strong expertise in hardware integration and software co-design that allows achieving the best HW/SW adequation for the targeted application and constrained resources. Several works are currently in progress in the laboratory to develop tools for the optimization and integration of neural networks (*i.e.* 8bits quantization, network pruning…), and thus to deliver hardware-efficient DNNs for embedded architecture.

This thesis work aims to exploring multi-task networks for JDE/JDT algorithms and their ability to be integrated within constraints of embedded architecture. We are looking for a fast and efficient multi-class multi-object tracking relying on a lightweight detection network that is unified with a subnetwork extracting instance-level embedding or/and motion prediction. Thus, the PhD candidate's exploration should allow the selection and implementation of an optimal multi-task CNN architecture that jointly solves object detection and tracking. The proposed tracker must meet requirements of embedded systems with limited power consumption and memory footprint but also cope with challenging tracking conditions - both the camera and the target(s) are moving.

Starting with the JDE framework, the very first research work, in the context of this thesis, may consist in exploiting the representational power of the intermediate feature maps of the object detector backbone to extract embedding. Next, the use of a more lightweight state-of-the-art object detector (*e.g.* FRDet [6], Mini-YOLOv3[7], SkyNet [15] or our own object detector…) with an additional embedding/motion prediction layer is expected. Since, the LIAE laboratory is developing mobile robotic platforms and holds a fully automated electric vehicle for autonomous driving; the work performed during this thesis will be used to feed the embedded perception bricks of such systems. It will be also presented at international conferences and scientific journals. Certain results may be patented.

About the CEA and LIST

The CEA (French Commission for Atomic and Renewable Energy) is a public research institute. It plays an important role in the research, development and innovation community. The CEA has four missions: security and defense, nuclear energy (fission and fusion), technology research for industry and fundamental research. With 16 000 employees, including technicians, engineers, researchers and support personnel, the CEA is involved in numerous research projects in collaboration with both academic and industrial partners.

In the section of the CEA focused on technology research for industry, the LIST institute is focused on intelligent digital systems. This institute has a culture of innovation and has as a mission to transfer these technologies to industrial partners. The DSCIN department specializes in complex digital and embedded systems for Artificical Intelligence (AI), High-Performance Computing (HPC) and Cyber security applications.

The focus of the LIAE laboratory is embedded Artificial Intelligence, perception combined with multimodal sensors (including vision sensors). The lab is located in the Paris region (Palaiseau).

The academic registration, the candidate will be affiliated with the Doctoral School InfoMaths (ED 512) at the Lyon university. The PhD student will be attached to the LIRIS team of Ecole Centrale de Lyon.

Contacts

Supervisor:	Martyna POREBA Michal SZCZEPANSKI	Email : martyna.poreba(at)cea.fr michal.szczepanski(at)cea.fr
Directory	Liming CHEN	École centrale de Lyon (ECL) Email : liming.chen(at)ec-lyon.fr
Qualifications :	• Master degree in informatics or electronics • Programming skills in Python, C/C++ • Familiarity with AI lib (tensorflow or pytorch) and OpenCV • Basic experience with Linux environment, Cmake, and Git	

	Former experience on using computer vision, deep learning for embedded devices will be appreciable.
To Apply :	Please send your application to Martyna POREBA including: - A detailed CV - A cover letter - Marks and ranking for the three previous academic years - Contact information for two references
Contact :	Martyna POREBA CEA DRT/DSCIN/LIAE Institut Carnot CEA List - CEA Saclay - Nano-INNOV Bat 862 - PC 172 F91191 Gif-Sur-Yvette Cedex Téléphone: +33 1 69 08 00 42 Email: martyna.poreba(at)cea.fr **More information on : http://www-instn.cea.fr/**
Start and End Dates	Octobre 2022 – Octobre 2025

References :

[1] Wojke, N., Bewley, A., Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric. 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645—3649.

[2] Wojke, N., Bewley, A. (2018). Deep Cosine Metric Learning for Person Re-identification. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 748-756.

[3] Wang, Z., Zheng, L., Liu, Y., Li, Y., Wang, S. (2020). Towards Real-Time Multi-Object Tracking. In: Vedaldi A., Bischof H., Brox T., Frahm JM. (eds) Computer Vision – ECCV 2020. Lecture Notes in Computer Science, vol. 12356. Springer, Cham.

[4] Zhang, Y., Wang, Ch., Wang X., Zeng W., Liu, W. (2020). FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. arXiv preprint: Computer Vision and Pattern Recognition.

[5] Chaabane, M., Zhang, P., Beveridge, R., O'Hara, S. (2021). DEFT: Detection Embeddings for Tracking. arXiv preprint arXiv:2102.02267

[6] Oh, S. T., You J-H., Kim Y-K. (2020). FRDet: Balanced and Lightweight Object Detector based on Fire-Residual. CoRR, preprint arXiv; 2011.08061.

[7] Mao, Q-Ch, Sun,H-M., Liu, Y-B.,Jia, R. (2019). Mini-YOLOv3: Real-Time Object Detector for Embedded Applications. IEEE Access, vol. 7, pp. 133529-133538.

[8] Lu, Z., Rathod, V., Votel, R., Huang, J. (2020). RetinaTrack: Online Single Stage Joint Detection and Tracking.2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14656-14666.

[9] Bergmann, Ph., Meinhardt, T., Leal-Taixe. L. (2019). Tracking without bells and whistles. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[10] Zhou, X., Wang, D., Krahenbuhl Ph. (2020). Tracking Objects as Points ECCV, 2020

[11] Shuai, B., Berneshawi, A., Li, X., Modolo, D., Tighe, J. (2021). SiamMOT: Siamese Multi-Object Tracking. CVPR 2021.

[12] Yang, Y. (2020). FastMOT: High-Performance Multiple Object Tracking Based on YOLO, Deep SORT, and Optical Flow, Software v1.0.0,

[13] Nousi, P., Mademlis, I., Karakostas, I., Tefas, A., Pitas. I. (2019). Embedded UAV Real-Time Visual Object Detection and Tracking. 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR), pp. 708-713.

[14] Danish, Brazauskas, Bricheno, Lewis, Mortier. (2020). DeepDish: multi-object tracking with an off-the-shelf Raspberry Pi. Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking.

[15] Zhang, X., Lu, H., Hao, C., Li, J., Cheng, B., Li, Y., Rupnow, K., Xiong, J., Huang, T., Shi,H., Hwu, W-M., Chen, D. (2020). SkyNet: a hardware-efficient method for object detection and tracking on embedded systems. Conference on Machine Learning and Systems.

[16] Leal-Taixé, L., Milan, A., Reid, I., Roth, S. & Schindler, K. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. arXiv:1504.01942 [cs], 2015., (arXiv: 1504.01942).

[17] Geiger, A., Lenz, P. & Urtasun, R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012.