Research Internship in Computer Vision
"Scalable Multi Objects Tracking for Retail"

Stage - Paris

Context:

This end-of-studies internship is a research internship. The student will not be fully integrated into the mission's team on the subject, in order to be detached from the short-term issues inherent from the mission; he will still have to interact with the data scientists. He/She will be supervised in the R&D department, with the ideal objective at the end of the internship of publishing a scientific article in a conference with an international peer review committee, the targeted format can be a short paper, demo paper format or full paper. At the end of this internship, the student will be considered for a position as a CIFRE PhD student, according to the opportunities and wishes of the student.

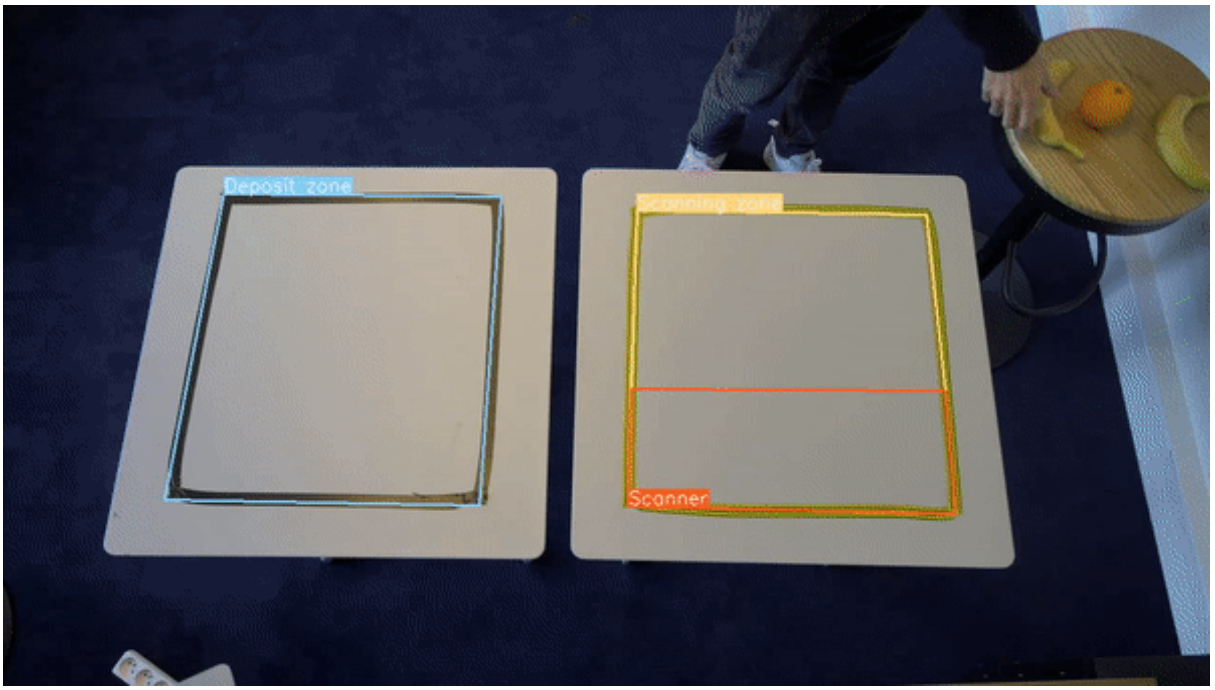
Presentation of Artefact

Artefact is a new generation consulting firm. Specialized in Data, AI, Data Marketing and Digital Marketing, Artefact is recognized as one of the leaders in these fields. In strong growth since its creation in 2017, Artefact today has more than 1100 employees in 17 offices around the world. The mission of the Artefactors? Transforming our customers' data into business impact. In this sense, we collaborate with more than 1,000 customers, leaders in their industries (Orange, Carrefour, Samsung, AccorHotels, L'Oréal, LVMH, Decathlon, etc.) and are privileged partners of Google (GMP/GCP), AWS as well as Microsoft. Azure.

Problem and data :

The objective is to develop a model for analyzing videos acquired in stores. The final industrial framework is to support customers when they pass through the self-check-out area, when scanning the items and placing them in the drop-off area. The objective is to detect the store's products present in the video stream, to be able to count the number of these products placed in the drop-off area, and ultimately to be able to identify the items as precisely as possible, within the limit of the typology identifiable on the video stream.

To illustrate the use case, we performed a demonstration video in our premises:



The student will have access during his internship to the data acquired during the mission with our client, which will be in real condition, in store. The dataset will cover the self-check-out areas of a handful of stores, and for each checkout an acquisition time of several hundred hours.

The problems of this internship will be clarified at the start. A first effort will be to onboard the core of the model, which aims at detecting objects and counting the deposited objects. This will be done in collaboration with Artefact data scientists developing the model. In a second step, the student will focus on longer-term questions and with a research objective. Here are listed the possible streams of research, and are not intended to be fully covered during the internship.

Improvement of the detection model with the aim of a non-invasive customer experience: The scenario is to deploy the algorithm and be integrated in daily experience for the entire population. Thus, there is a strong challenge to ensure a non-invasive experience. Typically, guarantees will be needed that the model does not raise alerts too frequently, which would block users in their checkout process. This non-invasive issue will have to be balanced with the business objectives of detecting cases of fraud, voluntary or involuntary.

Optimization of models for real-time inference: A challenge in this industrial context is to minimize inference costs in production, while maintaining real-time analysis of video streams. The literature mostly seems to use GPUs for inference, which is expensive as it would require one GPU per case. The student may consider optimizing the architecture of the network, the use of separable convolutions, or even proposing a pipeline that does not use a detection neural network at every frame but only sparsely. For this last point, rules adapted to our data and use-cases can be defined, since our use case shows

some properties that can be leveraged to optimize the pipeline and maintain satisfactory accuracy. Other algorithmic alternatives would typically be to pre-segment the image to only keep windows with segmented pixels.

Automatic learning for new products detection: The scaling of the solution also includes updating product detection in the case of new products on the shelves. Indeed, the range of products on the shelves evolves very regularly, and in the final regime the overall system must be able to detect the types of products, and not just their presence. It is unrealistic to carry out additional manual annotations each time there is a new product or new packaging. Therefore, a problem will be to develop a strategy where the model automatically learns to detect new products. This is made possible by the information fed back from the scan area machine, which indicates which product was last scanned. A difficulty behind this strategy will be to be able to identify the right level of typology of the said product, because it is in practice impossible to perform detection at the granularity of a product - it would for example be more pragmatic to have a "can" label and not the brand of the drink. A second difficulty would be to define a target architecture for such an automatic update, with the challenge of efficiently arranging the different stages of the pipeline.

Candidatures :

To apply, please write an e-mail to charlotte.silo@artefact.com and emmanuel.malherbe@artefact.com.

Literature :

Follmann, Patrick, Tobias Bottger, Philipp Hartinger, Rebecca Konig, and Markus Ulrich. "MVTec D2S: densely segmented supermarket dataset." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 569-585. 2018.

I. -I. Panagos, A. P. Giotis and C. Nikou, "Multi-object Visual Tracking for Indoor Images of Retail Consumers," *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2022, pp. 1-5, doi: 10.1109/IVMSP54334.2022.9816269.

Wang, Zhongdao, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang. "Towards real-time multi-object tracking." In *European Conference on Computer Vision*, pp. 107-122. Springer, Cham, 2020.

Kalake, Lesole, Wanggen Wan, and Li Hou. "Analysis based on recent deep learning approaches applied in real-time multi-object tracking: a review." *IEEE Access* 9 (2021): 32650-32671.

Voigtlaender, Paul, Michael Krause, Aljosa Osep, Jonathon Luiten, Berin Balachandar Gnana Sekar, Andreas Geiger, and Bastian Leibe. "Mots: Multi-object tracking and segmentation." In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pp. 7942-7951. 2019.