# Weakly-supervised one-shot object detection

**Environment**: QARMA (machine learning) Team at Laboratoire d'Informatique et Systèmes (LIS)
**Location**: Ecole Centrale Marseille (ECM), Technopôle de Château-Gombert, Marseille.
**Supervisors**: Ronan Sicre (LIS – ECM), Stéphane Ayache (LIS – Polytech).
**Salary**: legal minimum
**Keywords**: computer vision, deep learning, one-shot learning, weakly-supervised object detection.
**Contact:** ronan.sicre@lis-lab.fr

Computer vision and deep learning received a lot of attention lately due to the great improvements brought by Deep Neural Networks (DNN). Over the last decade, these networks are addressing more complex tasks and are reducing their requirement for large amounts of annotated data.

Over the last years several work study weakly-supervised object detection, which aims at localizing objects in images simply with image level information and without any localization information. Most recent approaches aim at learning models using masking method, meaning that the input image is masked before going through the DNN [1]. Then, the model should learn to remove the background "mask-in" or the object of interest "mask-out". These methods can also generate saliency maps localizing the object that can be useful for interpretability purposes. However, these systems require large annotated dataset.

At the same time, several work study the problem of few-shot, one-shot and zero-shot learning. These methods train models with few, one, or zero example. Furthermore, a lot of effort is recently dedicated to self-supervised models. These methods allow the learning of good and transferable representation from data without any sort of supervision. For instance, iBOT [2] combines contrastive learning on augmentation, teacher-student distillation and token masking to learn such representations.
Few works propose to combine few shot and detection. StarNet [3] propose to study weakly-supervised few-shot detection, where a query image will geometrically match its object instances with the ones of support images of the categories. Also, OSD [4] proposes one shot detection where a query object is matched to a set of images to obtain localization, using co-attention and co-excitation.

Our goal is to address a similar problem. First a self-supervised backbone will be used to obtain tensorial and vectorial representations. Then detection can be obtained by performing image level optimization with "mask-in" criteria. In other, the optimization goal is to produce a saliency map that will mask the image background and try to obtain a vectorial representation as close as possible to the category support image. Such image level optimization is similar to adversarial attacks, where the optimization is performed for only one image.

[1] Lu, Weizeng, et al. "Geometry constrained weakly supervised object localization." ECCV 2020.
[2] Zhou, Jinghao, et al. "ibot: Image bert pre-training with online tokenizer." ICLR 2022
[3]Karlinsky, Leonid, et al. "Starnet: towards weakly supervised few-shot object detection." AAAI 2021.
[4] Hsieh, Ting-I., et al. "One-shot object detection with co-attention and co-excitation." NeurIPS 2019.