Master Internship 2018: 
Weakly-Supervised Detection for Laparoscopic Surgery Understanding

Key words: Deep Learning, Convolutional Neural Networks, Weakly Supervised Detection, Tracking.

1 Context

An important open problem in computer-assisted minimally invasive surgery is to automatically detect organs or medical tools that are visible to the endoscopic camera. There are several important applications for this, e.g. automatic indexing of videos or augmented-reality based surgical guidance - see Figure 1.

![Figure 1: Detection of organs in laparoscopic videos for augmented-reality based surgical guidance.](image)

2 Goals and Challenges

State-of-the-art methods for object detection in images or videos are currently based on deep Convolutional Neural Networks (ConvNets) [KSH12, GDDM14, DLHS16, MTCH17]. However, providing accurate annotations, e.g. Bounding Boxes (BB) or segmentation (pixel labeling) is very expensive [BRFL16]. The problem is exacerbated when dealing with videos, where providing BB annotations in each frame of the sequence is not a viable solution.

In this internship, we explore Weakly Supervised Learning (WSL) solutions to this problem, i.e. performing accurate bounding box predictions from coarse annotations, e.g. global labels. The main issue for
Fig. 2: **Negative evidence models.** The maps show local scores, *i.e.* heat maps, for prediction models trained from global labels, for a train model (left) and a bus model (right). Negative evidence pooling functions [DTC15, DTC16, MDTC17] exploit both max scores (yellow) and min scores (blue) to classify images. This enables using localized evidence of the absence of the class, *e.g.* train wheels strongly support the absence of the bus class.

Training WSL models is to design a global prediction from local evidences, *i.e.* a pooling function. Negative evidence models [DTC15, DTC16, MDTC17] - see Fig. 2 - have recently successfully been applied for weakly supervised object detection or image segmentation. We implement these negative evidence models for weakly supervised object detection and evaluate them in the M2CAI’16 dataset[^1] composed of laparoscopic videos acquired at IRCAI[^2]. A comparison with Global Average Pooling (GAP) [ZKA+16] will be carried out.

A second aspect when dealing with object detection in videos is to take advantage of temporal coherence. Object detection at a given image of the sequence should be supported by coherent detections in subsequent frames. A solution to this problem is to rely on visual tracking solutions. Tracking is currently intensively revisited by deep learning methods. Especially, we are interested in methods for online learning of the object appearance, *e.g.* [BVH+17].

### 3 Profile

This opportunity is dedicated to master (or engineering school) students with major in computer science or applied mathematics. Background in machine learning, deep learning and optimization are required. A good understanding of modern ConvNet architectures (AlexNet, VGG, ResNet) is desirable. Programming experience in python (Numpy, Scipy, matplotlib) is important, with skills to scale algorithms for videos. Prior experience with standard Deep Learning library (*e.g.* Tensorflow, Keras) would be a plus.

### 4 Modalities

**Internship duration:** 6 months, starting February-March 2018.

**Location:** Cnam Paris, 2 rue conté 75003.

**Salary:** 1100 € brut / month

**Supervision:** Nicolas Thome (Cnam), Alexandre Hostettler (IRCAD).

**Contact:** nicolas.thome@cnam.fr, alexandre.hostettler@ircad.fr.

[^1]: [http://camma.u-strasbg.fr/m2cai2016/](http://camma.u-strasbg.fr/m2cai2016/)
[^2]: [https://www.ircad.fr/fr/](https://www.ircad.fr/fr/)
References


