

STAGE 2022

Réf : LVA-22-S6

Self-Supervised Interactive Segmentation

Presentation of the host laboratory

Based in Paris-Saclay campus, CEA-LIST is one of four technological research institutes of CEA TECH, the technological research direction of CEA. Dedicated to intelligent digital systems, it contributes to the competitiveness of companies via research and knowledge transfers. The expertise and competences of the 800 research engineers and technicians at CEA-LIST help more than 200 companies in France and abroad every year on subjects categorized over 4 programs and 9 technological platforms. 21 start-ups have been created since 2003.

The Computer Vision and Machine Learning for scene understanding Laboratory addresses computer vision subjects with a stronger emphasis on four axes:

- Recognition (detection or segmentation of objects and persons)
- Behavior analysis (action and gesture recognition, anomalous behavior of individuals or crowds)
- Smart annotation (large scale annotation of 2D and 3D data using semi-supervised methods)
- Perception and decision-making (Markovian decision processes, navigation)

The intern will join a team composed of 30 researchers (research engineers, PhD students, interns) and will be able to interact with peers working on related subjects and methods.

Context

As deep learning gains popularity, the need for large amounts of annotated images has never been greater. Annotating images is a tedious and time-consuming task, especially in the field of image segmentation where human annotators have to draw complex polygons around all sorts of objects. Interactive segmentation can considerably reduce the amount of time needed to annotate a dataset by making the process of annotating images much easier for annotators. In another area, video or image editing software also use interactive segmentation to help artists precisely select objects within images. However, such interactive segmentation algorithms heavily rely on large annotated datasets to train and are therefore highly dependent on the type and quality of those given annotations. To overcome this issue, several methods choose to exploit synthetically created datasets from « png » images in order to obtain perfectly delineated objects [1]. While they already obtain a significant gain in accuracy, such images come in limited numbers. Meanwhile, self-supervised learning techniques have recently proved very successful in the extraction of meaningful feature maps for downstream tasks such as segmentation [2].

Objectives of this internship

Based on these observations, the objectives of the internship are :

- Analyze existing self-supervised « pre-tasks » and choose the most relevant to generate features which discriminate well objects in a scene
- Conceive and develop an interactive segmentation algorithm, which combines such features with sparse human interactions in order to generate fine-grained segmentation masks. To achieve such a goal, the intern may look into classical image segmentation tools and adapt their use of either texture or edge information (e.g. Magic Wand, Intelligent Scissors, Graph-Cut) to « deep » embeddings.
- Evaluate the developed method on standard interactive segmentation benchmarks (SBD, MsCOCO, DAVIS).

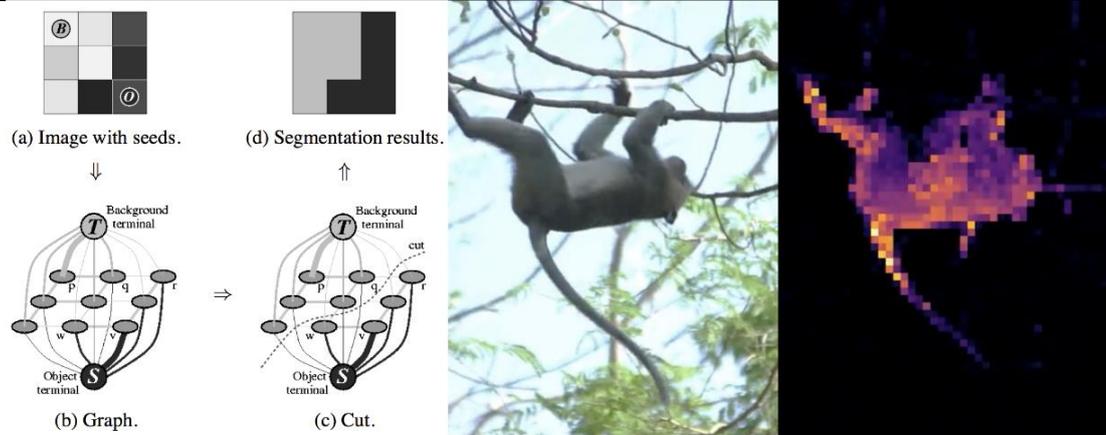


Figure 1 - LEFT Image segmentation as a graph cut [3] : the user provides some prior knowledge about the object and background through the O (Object) and B (Background) and the labelling of the pixels is then solved through a cut on the graph. RIGHT Attention heatmap generated with a Self-supervised Vision Transformer (DINO) [4]

Keywords

instance segmentation, deep learning, self-supervised learning.

References

- [1] Marco Forte, Brian Price, Scott Cohen, Ning Xu, François Pitié. Getting to 99% Accuracy in Interactive Segmentation (2020), <https://arxiv.org/abs/2003.07932>
- [2] Xie, Zhenda and Lin, Yutong and Zhang, Zheng and Cao, Yue and Lin, Stephen and Hu, Han. Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning (CVPR 2021)
- [3] Boykov and Jolly, "Interactive Graph Cuts" (ICCV 2001)
- [4] Emerging Properties in Self-Supervised Vision Transformers (ICCV 2021)
- [5] Duchenne et al. Segmentation by transduction (CVPR 2008).

Required level:	Engineer, Master 2
This internship opens the possibility of pursuing a thesis and R&D engineer in our laboratory.	
Duration :	6 months
Remuneration:	between 700 € and 1300 € depending on the training.
Required Skills :	
<ul style="list-style-type: none"> - Computer vision - Machine learning (deep learning) - Shape recognition - Proficiency in programming (Python) - Mastery of a deep learning framework (in particular PyTorch or Tensorflow) 	