Weakly supervised semantic and instance segmentation

Supervisor's names:
Alexandre Baussard (Full professor) alexandre.baussard@utt.fr
Pierre Beauseroy (Full professor) pierre.beauseroy@utt.fr

Description of the research project

In recent years, machine learning and especially machine vision has made substantial advances in various domains such as object detection, classification and image segmentation. If we shortly describe each of these general tasks, we have:

- Image Classification: Classify the main object category within an image.
- Object Detection: Identify the object category and locate the position using a bounding box for every known object within an image.
- Semantic Segmentation: Identify the object category of each pixel for every known object within an image. Labels are class-aware.
- Instance Segmentation: Identify each object instance of each pixel for every known object within an image. Labels are instance-aware.

Illustrative examples are presented in Fig.1. Thus, one can say that semantic and instance segmentation are among the hardest possible vision tasks. Consequently, they are still open issues and one of the most ongoing area of works.

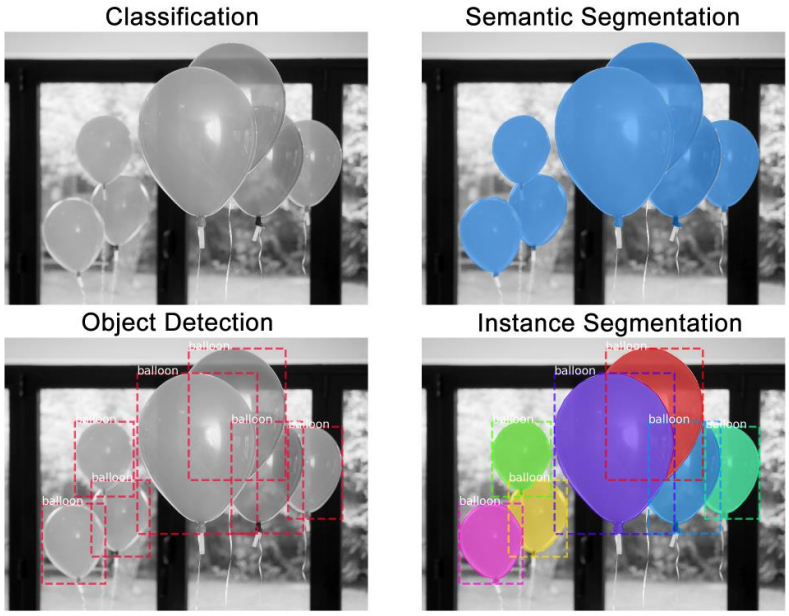

Fig. 1: Illustration of the differences between classification, detection, semantic segmentation and instance segmentation.

These advances in machine vision are mainly based on deep learning methods. Results obtained on public benchmark data sets by applying these methods, generally Convolutional Neural Nets (CNN), outperform other approaches by a large margin in

terms of accuracy and efficiency. However, one of the main weaknesses of deep learning approaches is that they need a large number of training samples for high quality results. It means that a large number of pixel-level labelled training data is necessary while manually pixel-level annotation is a time-consuming, expensive and laborious task. To overcome this problem, weakly supervised methods have been recently proposed. They use weak annotation, for example image-level labels and/or bounding boxes, which are more accessible because of availability in existing public data sets or because of the low annotation costs. Due to incomplete annotations, most trained models focus on discriminative parts rather than the whole object area, providing inaccurate segmentation results. That is why additional information for object localization and shape is usually used. To add localization information, techniques based on CNN can be employed, where some works show the localization ability of the network to identify rough locations of discriminative parts of each class in an image. To add shape cue, several techniques have been proposed. Superpixels, i.e. a group of connected pixels with similar colors or gray levels, or fully connected CRF (Conditional Random Field) are used as post-processing techniques to smooth the segmentation. Despite these methods improve results, they assume that object boundaries are well-defined, which is not always the case. The aim of this doctoral project is to investigate the applicability of a weakly supervised image or signal segmentation method in cases where labels are given at image or signal level.

Additional Bibliography

1. l. Michaelis, I. Ustyuzhaninov, M. Bethge, A.S. Ecker, One-Shot Instance Segmentation, arXiv:1811.11507, 2018.
2. A. Khoreva, R. Benenson, J. Hosang, M. Hein, B. Schiele, Simple Does It: Weakly Supervised Instance and Semantic Segmentation, arXiv:1603.07485, 2016.
3. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2921-2929).
4. Pathak, D., Krahenbuhl, P., & Darrell, T. (2015). Constrained convolutional neural networks for weakly supervised segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1796-1804).