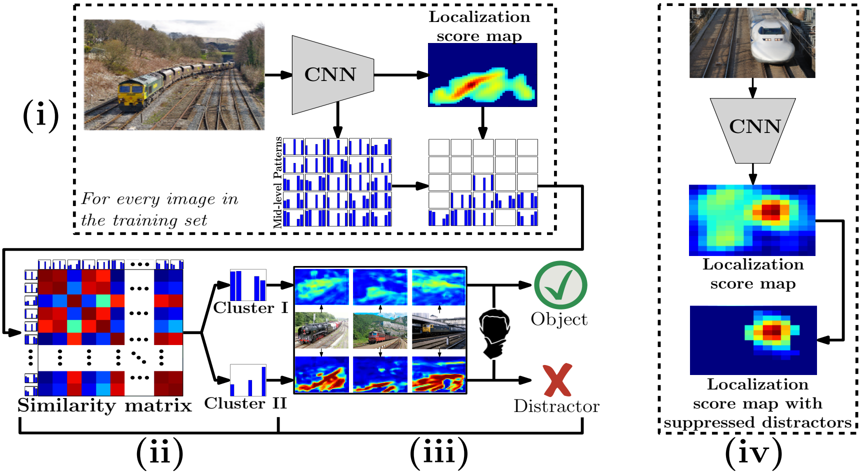
# Improving Weakly-Supervised Object Localization By Micro-Annotation

Alexander Kolesnikov
akolesnikov@ist.ac.at
Christoph H. Lampert
chl@ist.ac.at

IST Austria
Am Campus 1
3400 Klosterneuburg
Austria



Object localization is a crucial step needed for building automatic systems for visual scene understanding. This task can be successfully tackled using fully-supervised learning methods, but these require annotations in a form of bounding boxes or per-pixel segmentation masks that are time-consuming and expensive to acquire. Therefore, it is important to develop weakly-supervised object localization learning techniques, which require much cheaper forms of annotation, *e.g.* image-level class labels.

Analyzing the current methods for weakly-supervised object localization we arrive at the conclusion that they tend to fail for object classes that consistently co-occur with the same background elements (distractors), *e.g. trains* on *tracks*. We overcome these failures by developing a new procedure that determines semantic parts that constitute the object detection and then discards distractor parts. The main steps of our approach are (see Figure above) (i) represent all predicted foreground regions of all images by mid-level features learned by a deep neural network, (ii) cluster these features using spectral clustering (the number of clusters is determined automatically), (iii) visualize the clusters and let a human annotator select which ones actually corresponds to the object class of interest. The information about clusters and their annotation can then be used to better localize objects: (iv) for any (new) image, predict a foreground map using only the image regions that match clusters labeled as ’object’.

Note, that the proposed method requires virtually negligible amount of additional supervision: an annotator has to answer a few binary questions (typically 2 or 3) per semantic class. Huge datasets, such as *ILSVRC*, can be annotated by one annotator in just a few hours.

The proposed approach can be readily used in combination with many existing localization methods. In this work we combine it with the current state-of-the art methods for weakly-supervised bounding box prediction [2] and for weakly-supervised semantic segmentation [1], showing improved results on the challenging *ILSVRC* 2014 and *PASCAL VOC* 2012 datasets.

[1] A. Kolesnikov and C. H. Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. *ECCV*, 2016.

[2] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *CVPR*, 2016.