

Hierarchical Compositional Representations of Object Structure

Aleš Leonardis*

¹ Visual Cognitive Systems Laboratory
Faculty of Computer and Information Science
University of Ljubljana
Tržaška 25, SI-1001 Ljubljana
Slovenia

² School of Computer Science
University of Birmingham
Edgbaston, Birmingham, B15 2TT
United Kingdom

ales.leonardis@fri.uni-lj.si, a.leonardis@cs.bham.ac.uk

Abstract. Visual categorisation has been an area of intensive research in the vision community for several decades. Ultimately, the goal is to efficiently detect and recognize an increasing number of object classes. The problem entangles three highly interconnected issues: the internal object representation, which should compactly capture the visual variability of objects and generalize well over each class; a means for learning the representation from a set of input images with as little supervision as possible; and an effective inference algorithm that robustly matches the object representation against the image and scales favorably with the number of objects.

In this talk I will present our approach which combines a learned compositional hierarchy, representing (2D) shapes of multiple object classes, and a coarse-to-fine matching scheme that exploits a taxonomy of objects to perform efficient object detection. Our framework for learning a hierarchical compositional shape vocabulary for representing multiple object classes takes simple contour fragments and learns their frequent spatial configurations. These are recursively combined into increasingly more complex and class-specific shape compositions, each exerting a high degree of shape variability. At the top-level of the vocabulary, the compositions represent the whole shapes of the objects. The vocabulary is learned layer after layer, by gradually increasing the size of the window of analysis and reducing the spatial resolution at which the shape configurations are learned. The lower layers are learned jointly on images of all classes, whereas the higher layers of the vocabulary are learned incrementally, by presenting the algorithm with one object class after another.

However, in order for recognition systems to scale to a larger number of object categories, and achieve running times logarithmic in the number of classes, building visual class taxonomies becomes necessary. We propose

* *This is a joint work with Sanja Fidler and Marko Boben.*

an approach for speeding up recognition times of multi-class part-based object representations. The main idea is to construct a taxonomy of constellation models cascaded from coarse-to-fine resolution and use it in recognition with an efficient search strategy. The structure and the depth of the taxonomy is built automatically in a way that minimizes the number of expected computations during recognition by optimizing the cost-to-power ratio. The combination of the learned taxonomy with the compositional hierarchy of object shape achieves efficiency both with respect to the representation of the structure of objects and in terms of the number of modeled object classes. The experimental results show that the learned multi-class object representation achieves a detection performance comparable to the current state-of-the-art flat approaches with both faster inference and shorter training times.