

I have seen enough: Transferring parts across categories

David Novotny¹
david@robots.ox.ac.uk

Diane Larlus²
diane.larlus@xrce.xerox.com

Andrea Vedaldi¹
<http://www.robots.ox.ac.uk/~vedaldi>

¹ Visual Geometry Group
University of Oxford

² Computer Vision Group
Xerox Research Centre Europe



Figure 1: We study the transferability of object parts and ask a question: What is the minimal amount of supervision so that part detectors trained on source classes perform well on unseen target classes? To provide an answer, we experiment with a novel Animal Parts dataset.

Recent progress in image understanding, while dramatic, has been primarily fueled by the availability of increasingly large quantities of labeled data. However, it is unclear whether manual supervision will be able to keep up with the demands of increasingly sophisticated and data-hungry algorithms. This raises the obvious question: *When is supervision enough?*

In this paper we examine this question from the viewpoint of learning shareable semantic parts, focusing on two research problems.

- (i) We conduct the first careful investigation of part transferability across a large set of visually dissimilar classes. To this end, the problem is viewed from the *Domain Adaptation* (DA) perspective where domains correspond to animals.
- (ii) We investigate how many annotations are required to train a part detector. We consider an *Active Learning* scenario, studying which images should be chosen and how many are needed to saturate performance of the part detector.

ImageNet Animal Parts dataset. A thorough evaluation of the transferability of parts requires a suitable dataset with a large number of classes. Because existing datasets are insufficient for our task, we annotated “foot” and “eye” part key-points in 14711 existing ImageNet images from 100 animal classes. Example annotations can be seen in Figure 1.

Proposed methods. Our part detectors rely on convolutional neural networks. We adapt the widely used *uncertainty* sampling principle to our scenario. We further propose an efficient ensemble of detectors which is optimized for generalization to new classes and can be used to guide active learning.

Main results

Part transferability. We challenge the common assumption that parts are visually “shareable” across different classes and therefore it suffices to learn them from a limited number of classes to understand them equally well in all cases. Given an unseen target class, by comparing performances of detectors trained on its semantically nearest/farthest classes we have verified the relevance of semantic distance for cross-domain transfer. Furthermore, we have shown that learning parts from a sufficiently diverse set of classes allows satisfactory transfer to novel classes.

Have I seen enough? We also ask how many source images need to be annotated in the source domains such that the performance of a part detector saturates as quickly as possible. Results indicate that our proposed method outperforms other baselines on foot detection while performing on par for eyes. Besides the relative merits of compared methods, the performance reaches 98% of the accuracy of the fully annotated scenario by providing only a few thousand annotations, showing that excellent performance can be achieved by annotating a small representative subset of classes and images.