Building Visual Class Hierarchies from Flickr

Pierre Benard (benard@inrialpes.fr)
INRIA Grenoble, LEAR / LJK

Marcin Marszałek (marszale@inrialpes.fr)
INRIA Grenoble, LEAR / LJK

Cordelia Schmid (schmid@inrialpes.fr)
INRIA Grenoble, LEAR / LJK

Abstract

We propose a method to automatically extract vision-oriented semantic information from Flickr. Building on image tags - small pieces of semantics provided by Flickr users - we show how to construct a rich class hierarchy that reflects visual similarities between classes. In our automatically built class hierarchies we observe semantic relationships similar to ones present in expert ontologies, but we also discover visual context links and scene-type groupings. Experiments show that our vision-oriented hierarchies outperform ontology-based hierarchies in terms of modeling the visual similarities between object classes.

1. Introduction

We take a novel approach in modeling the inter-class relationships — we propose to build a class hierarchy by mining a large database of visual labels.

Input. Flickr — a large, non-expert database of personal photos. Association rules [1] from image tags can be mined.

Method. An association rule could indicate a relation between classes. We reduce the graph to a directed acyclic graph.

Result. We discover class hierarchies that are both vision-oriented and semantic. This can help object recognition [3].

Figure 2: Flickr tag clusters - densely connected components in the association rules graph mined from image tags (see fig. 1)

2. Class hierarchy

Semantic. High-level concepts are automatically assigned to high-level nodes (concept organization like in WordNet). This allows reasoning, see our CVPR'07 paper [2].

Visual. The hierarchy reflects how Flickr users see the world. Four kinds of relationships are dominant:

- high-level concept grouping (*pet*, *animal*)
- inclusion of parts (*eye*, *nose*)
- contextual links (*grass*, *fence*)
- scene-type associations (*city*, *farm*)

3. Results

Qualitative. We observe that the Flickr hierarchy reflects visual and contextual similarities between the categories. It also expresses more relationships than the WordNet hierarchy.

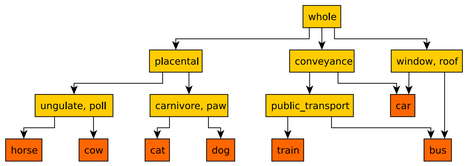

Figure 3: Automatically retrieved WordNet hierarchy

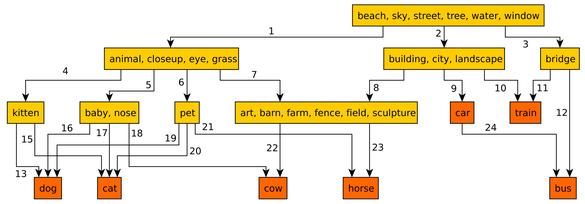

Figure 4: Automatically retrieved Flickr hierarchy

Quantitative. We evaluate if classes which are close in the hierarchy are visually similar. The capacity to distinguish between clusters of classes is measured.

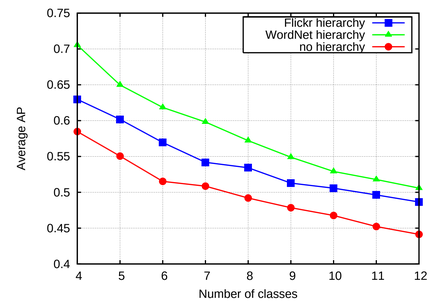

Figure 5: Results on the Pascal VOC'07 challenge dataset

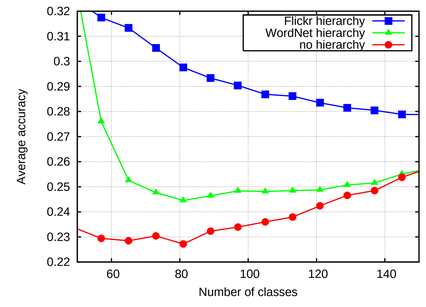

Figure 6: Results on the Caltech 256 dataset

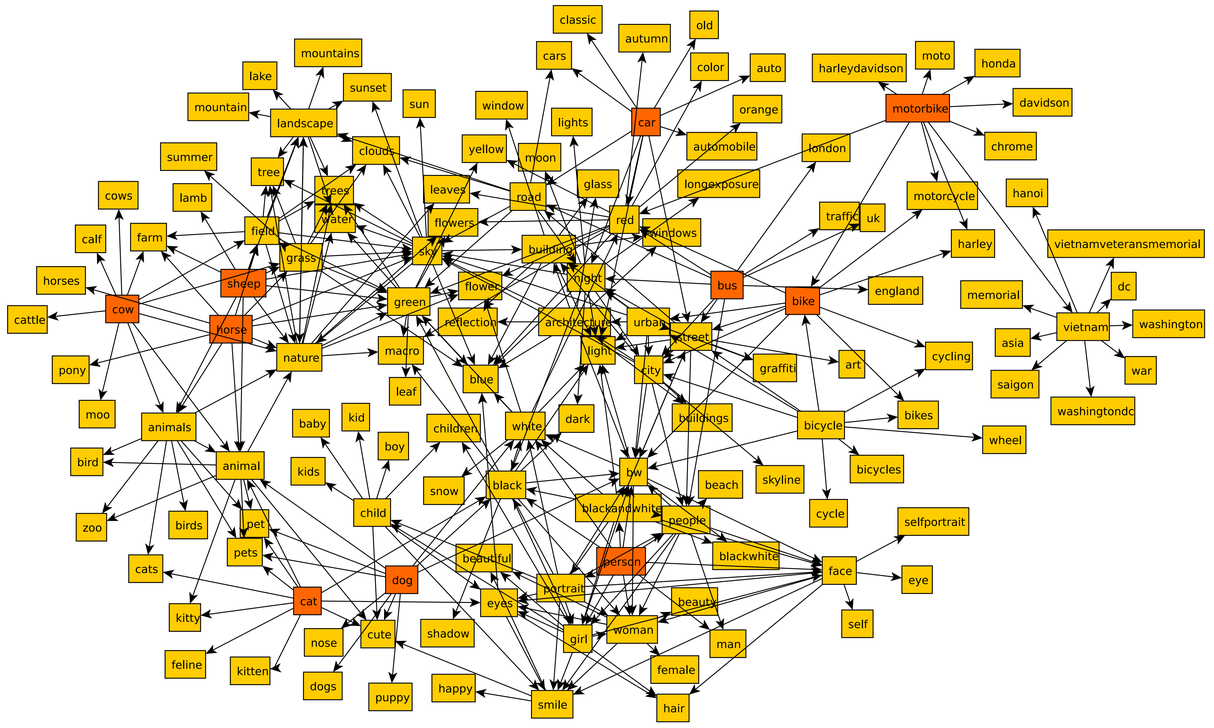

Figure 1: Association rules mined from Flickr image tags. The query consisted of 10 VOC'06 class labels. Note the unstructured form of the graph.

4. Discussion

We check how representative are the discovered inter-class relationships for popular benchmark datasets.

(a) Eyes and noses of cats and dogs (VOC'07), links 4, 6

(b) Cars and buses in a city (VOC'07), l. 9, 24

(c) A train on a bridge (VOC'07), link 11

(d) Cows and horses next to a fence (VOC'07 and Caltech-256), links 22, 23

(e) A horse among cars (VOC'07), link 8

(f) A horse in a city (Caltech-256), link 8

(g) Dogs and cats may be difficult to distinguish (VOC'07), *kitten* node

Figure 7: Examples illustrating the discovered relationships. Note the corresponding links in the class hierarchy.

5. Summary

- Semantic and visually oriented class hierarchies can be automatically extracted from Flickr
- Flickr inter-class relationships go beyond grouping classes to high-level concepts or building objects from parts — contextual links and scene-types are also present
- The knowledge extracted from Flickr can explain the content of popular recognition benchmarks
- It can be used to predict the inter-class similarities and outperforms WordNet based semantic hierarchies for a large number of classes

References

[1] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *SIGMOD*, 1993.

[2] M. Marszałek and C. Schmid. Semantic hierarchies for visual object recognition. In *CVPR*, 2007.

[3] A. Zweig and D. Weinshall. Exploiting object hierarchy: Combining models from different category levels. In *ICCV*, 2007.