

# Visual Knowledge Transfer Using Semantic Relatedness

Marcus Rohrbach

Department of Computer Science, TU Darmstadt

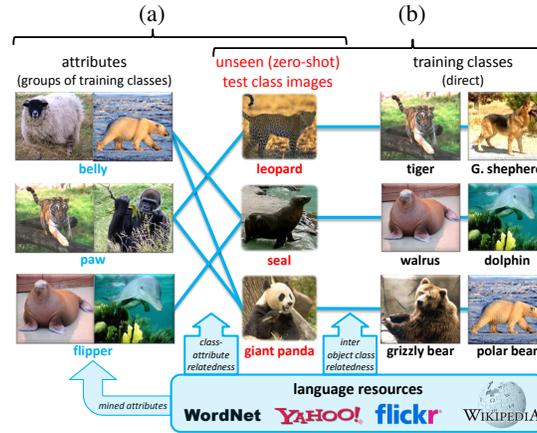
**Abstract.** Recognition of object classes has advanced to master difficult scenarios such as background clutter or multiple objects per image. Scaling to large numbers of classes, however, remains a challenge for state-of-the-art recognition approaches. In order to exploit similarities between classes, knowledge transfer between classes has been advocated. In this paper we examine the special case of transferring knowledge from known to unseen classes (zero-shot recognition). In previous work the decision which knowledge to transfer has been provided mostly by supervision in the form of manual associations between known and unseen classes or a few training examples, limiting the scalability of these approaches. We promote semantic relatedness to replace supervision in order to provide the missing link between the sources (known classes) and targets (unseen classes) of knowledge transfer. We provide a rigorous experimental evaluation of several state-of-the-art semantic relatedness measures and language resources which we evaluate on the challenging Animals with Attributes image dataset. This extensive evaluation provides insights into the different qualities and applicability of the different measures and resources.

## 1 Introduction

While remarkable recognition performance has been reported on a wide variety of object classes, scaling recognition to large numbers of classes remains a key challenge for state-of-the-art recognition approaches. The limiting factors are mainly the large training sets required to learn each object class and the restricted number of different classes systems are able to handle. It is generally believed that *knowledge transfer* between image classes is a promising way to tackle both problems. However, approaches so far require manual supervision [1, 2] or rely on a few bootstrap training examples [3, 4].

In this work we reduce the amount of manual supervision by using linguistic knowledge bases as an additional source of information. In the special case of zero-shot recognition we transfer knowledge from known object classes (during training) to unseen object classes (during testing) using two different models. The first model, suggested by [1], uses attributes as a level of indirection to generalize over classes. Fig. 1(a) visualizes the model and shows that one attribute (e.g. *paw*) is associated with several training classes (*tiger*, *gorilla*) and test classes (*leopard*, *giant panda*). The binary object-attribute associations are mined from language resources, such as WordNet [5], Wikipedia, Flickr tags and descriptions, or the World Wide Web queried by Yahoo. The second model uses inter-object class relatedness to transfer knowledge. As shown in Fig. 1(b) unseen classes are directly associated with the most similar training classes, e.g. *seal* with *walrus* and *dolphin*. Again, associations are mined from language.

The main contributions of this work are as follows. First, we provide a thorough evaluation of a variety of language resources and semantic relatedness (SR) measures



**Fig. 1.** Visual knowledge transfer from training classes to unseen test class images (zero-shot recognition). The required associations are provided by semantic relatedness measures based on different language resources (blue-colored lines). Two models are distinguished: **(a)** transfer from attributes to unseen classes using *class-attribute relatedness*; **(b)** transfer from most similar classes using *inter-object class relatedness*.

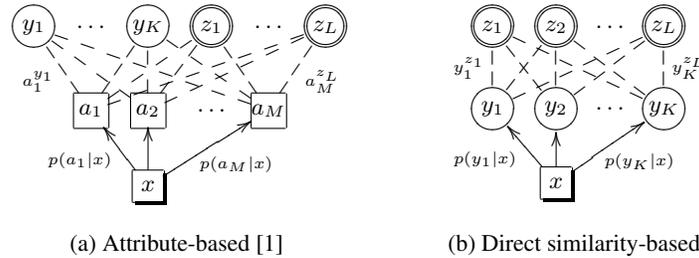
which allow significantly reducing supervision in zero-shot object detection. Second, we compare two different models for zero-shot detection including several levels of automation for the attribute model to further reduce manual intervention. Third, we discuss the major differences of different language resources and give insights which might also be applicable to other vision tasks. Large parts of this work have been published in [6].

The remainder of the paper is structured as follows. After a review of related work (Sect. 2) we present the models and approaches used for zero-shot knowledge transfer (Sect. 3) and introduce the different language resources and relatedness measures (Sect. 4). Then we discuss our experimental results in Sect. 5 and conclude with an outlook in Sect. 6.

## 2 Related work

Sharing and transferring knowledge between classes has recently become an important research direction for scalable recognition. We distinguish between two broad types of knowledge transfer: First, sharing knowledge across classes on a separate layer of attributes, and, secondly, directly transferring knowledge between classes using inter-class similarity. We conclude this section by discussing approaches which use vision and language resources in combination.

Attributes have been used in a wide variety of vision applications. Approaches range from elementary visual properties such as colors or geometric patterns [7–9] to high level attributes such as *gender* for face verification [10], background scenes [11], or parts [2]. As attribute activations can characterize object classes without using reference exemplars they can be used for *zero-shot* classification of previously unseen object classes: [12] formally discusses zero-shot learning and compares the performance of a



**Fig. 2.** Two models for zero-shot object classification. See Sect. 3 for discussions.

linguistic knowledge base (Google Trillion-Word-Corpus) to manual labels. However, they apply it to a very different domain, namely neural decoding of thoughts. In the domain of computer vision, [1] presents zero-shot classification schemes based on attributes, where the associations between attributes and object classes are obtained using manual supervision by human subjects. [13] advocates a paradigm shift from “naming” (by object classes) to “describing” (by attributes), distinguishing among common, discriminating, unusual, and unexpected attributes for object classes. While we use the model proposed by [1] for attribute-based knowledge transfer, we replace manual supervision with information extracted automatically from language resources. In particular, our attribute-based approach is the first which is able transfer knowledge to unseen visual object classes without any additional supervision.

In contrast to attribute-based approaches, which employ an additional layer of indirection, direct approaches use inter-class similarities to model relationships. One direction is to model inter-class similarities in a semantic hierarchy [14–17]. A second direction is to directly transfer knowledge from known to unseen classes by transferring class priors [2, 18] or by using one or few example images of these new classes [4, 19, 3]. Our second model for knowledge transfer is also based on such direct similarities but we extend it to zero-shot classification by using information obtained from language resources instead.

Literature combining vision and language resources ranges from estimating the “visualness” of words [20, 21], over creating visual ontologies [16, 17] to joint models of images and accompanying text [22, 23]. Similar to our work, search engine hit counts have been used to determine semantic relatedness between objects and background scenes [11] or to find the color of objects [7]. Although these approaches are still limited in the use of language resources and semantic relatedness measures applied, they show the benefit of using visual and language information together. This work extends several of these ideas and provides an in-depth study of a range of SR measures applied to different language resources for zero-shot object class detection.

### 3 Models and approaches for zero-shot knowledge transfer

We examine knowledge transfer in a zero-shot scenario, i.e. we assume that we have a set of known classes  $y_1, \dots, y_K$  and a set of unseen classes  $z_1, \dots, z_L$ . For this we build upon two different kinds of models, namely *attribute-based classification* [1] with

an intermediate level of attributes  $a_1, \dots, a_M$  (Fig. 2(a)) and *direct similarity-based classification* [3], transferring knowledge directly from the known classes  $y_1, \dots, y_K$  to the unseen classes  $z_1, \dots, z_L$  (see Fig. 2(b)).

### 3.1 Attribute-based classification

The attribute-based model aims to learn common visual features across classes defined by attributes. As these trained attribute classifiers generalize over the set of training classes they can be used to classify new unseen classes. This is visualized in Fig. 1(a), where the attribute *paw* generalizes over the training classes *tiger* and *gorilla* which can then be transferred to the test classes *leopard* and *giant panda*.

To formally model these associations we use the direct attribute prediction (DAP) model suggested by [1]. This is shown in Fig. 2(a): an attribute  $a_m$  is associated with training ( $y_1, \dots, y_K$ ) and test classes ( $z_1, \dots, z_L$ ), class-attribute associations shown with dashed lines. The class-attribute associations ( $a_m^{y_k} / a_m^{z_l}$ ) can either be active or inactive resulting in a group of classes associated with each attribute. E.g., for *giant panda* in Fig. 1(a), attributes are either active (*belly* and *paw*) or inactive (*flipper*).

Following the probabilistic formulation of DAP in [1], let  $a^y = (a_1^y, \dots, a_M^y)$  be a vector of binary associations  $a_m^y \in \{0, 1\}$  between attributes  $a_m$  and training object classes  $y$ . A classifier for attribute  $a_m$  is trained by labeling all images of all classes for which  $a_m^y = 1$  as positive and the rest as negative training examples. Now this classifier can estimate the posterior probability  $p(a_m|x)$  of that attribute being present in image  $x$ . Mutual independence yields  $p(a|x) = \prod_{m=1}^M p(a_m|x)$  for multiple attributes.

In order to transfer attribute knowledge to an unseen class  $z$ , we again assume a binary vector  $a^z$  with  $p(a|z) = 1$  for  $a = a^z$  and  $p(a|z) = 0$  otherwise. The posterior probability of class  $z$  being present in image  $x$  is then obtained by marginalizing over all possible attribute associations  $a$ , using Bayes' rule  $p(z|a^z) = \frac{p(a^z|z)p(z)}{p(a^z)} = \frac{p(z)}{p(a^z)}$ :  
 $p(z|x) = \sum_{a \in \{0,1\}^M} p(z|a)p(a|x) = \frac{p(z)}{p(a^z)} \prod_{m=1}^M p(a_m|x)^{a_m^z}$ . Assuming identical class priors  $p(z)$  and a factorial distribution for  $p(a) = \prod_{m=1}^M p(a_m)$ , we obtain

$$p(z|x) \propto \prod_{m=1}^M \left( \frac{p(a_m|x)}{p(a_m)} \right)^{a_m^z} \quad (1)$$

Attribute priors can be approximated by empirical means over the training classes  $p(a_m) = \frac{1}{K} \sum_{k=1}^K a_m^{y_k}$  or set to  $\frac{1}{2}$  [1]. Classifying an image  $x$  according to test classes  $z_L$  uses MAP prediction  $\operatorname{argmax}_{l=1, \dots, L} p(z_l|x)$ .

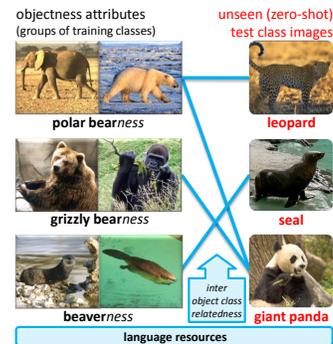
This leaves us with estimating the class-attribute associations for the unseen classes  $a_m^z$  and the known classes  $a_m^y$ . [1] based the associations on judgments of 10 test subjects [24, 25]. Although [1] achieved promising performance with these associations, it requires a large amount of human involvement when moving to a new domain or sets of classes. In the following we suggest three approaches to decrease this required supervision. First, we reduce the human involvement by mining the associations automatically using semantic relatedness measures. Second, the attributes are also mined automatically to achieve a fully unsupervised setting. Finally, our third approach uses the class terms themselves as *objectness attributes*.

**Mining class-attribute associations.** To mine class-attribute associations one can use widely and freely available language resources paired with semantic relatedness (SR) measures. In contrast to domain specific knowledge bases which might require specific queries and measures, SR measures are generally applicable to a large range of textual resources such as the World Wide Web and thus have a very broad coverage and availability (see Sect. 4). When the class-attribute associations were determined manually by [24], human judges were provided with full text attribute descriptions. One example is the property “*living in the New World (North and South America)*” which was abbreviated by [24] to “*newworld*”. Neither the description nor the abbreviation is appropriate as input for SR measures. We thus map the attribute descriptions/abbreviations to a concise term, in this example we used the term *america*. In this process a large amount of information is obviously lost or changed and the attributes loose precision and accuracy which increases the amount of noise (see results in Sect. 5).

**Mining attributes.** The amount of supervision is reduced significantly by mining the class-attribute associations but it still requires the definition of a set of attributes when moving to another domain. Important qualities of the attributes are the ability to discriminate between object classes and to be visually distinctive, but at the same time shared among several classes to enable knowledge transfer. To meet these criteria we suggest using part attributes (e.g. *flipper* for animals, *wheel* for vehicles): part-based approaches have successfully been used in vision in different domains and most objects consist of parts. We mine WordNet for the explicitly encoded parts of all 50 class terms and recursively all sub- and super concepts. In total we found 74 part terms<sup>1</sup>, see e.g. Fig. 1(a).

**Objectness as attributes.** Our third approach is based on the idea that each attribute is represented by a subset of all classes which contain similar visual image features. Thus such groups can be formed by grouping similar classes. More specifically we use the class names themselves as attributes, for instance *beaver* in Fig. 3 groups object classes which are similar to a beaver. This attribute could thus be interpreted as *beaverness* of an object. We denominate this approach as *objectness*.

Regarding the underlying model there is no change, but SR measures are now required to measure inter-object class similarities rather than object-attribute relatedness. In line with this change objectness attributes frequently group very similar classes compared to the more generic AwA or part attributes, e.g. the attribute *grizzly bearness* in Fig. 3 groups several bears (*grizzly bear*, *gorilla*, and *giant panda*) in contrast to the attribute *belly* (see Fig. 1(a)), which groups the very diverse classes *sheep*, *polar bear*, *seal*, and *giant panda*.



**Fig. 3.** Objectness

<sup>1</sup> All software for computing object class-attribute associations from linguistic language resources and obtained intermediate results (lists of mined attributes, object class-attribute associations) will be made publicly available on our web pages prior to DAGM 2010.

### 3.2 Direct similarity-based classification.

Rather than using the indirection of an attribute layer we can classify the unseen classes  $z_1, \dots, z_L$  by using the learned models of the most similar training class  $y_1, \dots, y_K$  as illustrated in Fig. 2(b). Again, the similarity  $y_k^z$  between the test class  $z$  and training class  $y_k$  can be mined from language resources (Fig. 1(b)).

The direct similarity model can be interpreted as the DAP model with  $M = K$  attributes, where each attribute corresponds exactly to one training class  $y_k$ . We thus train classifiers for each class  $y_k$  to provide estimates of  $p(y_k|x)$  for a test image  $x$ . Now we can estimate the posterior of a test image  $x$  by applying these changes to Equation 1:  $p(z|x) \propto \prod_{k=1}^K \left( \frac{p(y_k|x)}{p(y_k)} \right)^{y_k^z}$ . Instead of using only binary associations  $y_k^z$  we found empirically that continuous weights improve performance. We weight the most similar classes using the continuous similarities  $w_y^z$  between  $z$  and  $y$  as normalized weights  $y_k^z = \frac{w_{y_k}^z}{\sum_{i=1}^K w_{y_i}^z}$  and set  $K = 5$  in all our experiments.

## 4 Semantic relatedness measures and language resources

In this section we introduce several semantic relatedness (SR) measures to extract similarity information from the most widely used language resources.

**WordNet** [5] is the largest machine readable expert-created language ontology. It contains over 100,000 concepts which are organized in a hierarchical graph structure. We employ the measure proposed by Lin [26] which estimates SR between two concepts by comparing their individual depth in the hierarchy with the depth of their lowest common subsumer.

**Wikipedia** is the largest community built online encyclopedia. The state-of-the-art Explicit Semantic Analysis (ESA) measure on Wikipedia [27] represents each term as a vector of frequencies over all articles. Similarity of two terms  $t_1, t_2$  is computed by the cosine between the two respective vectors  $sim_{ESA}(c_1, c_2) = \frac{\vec{c}_1 * \vec{c}_2}{|\vec{c}_1| * |\vec{c}_2|}$ .

**Yahoo Web.** The World Wide Web is presumably the largest publicly available source of textual information. Due to this fact it has been extensively use in diverse Natural Language Processing applications, mainly by using the hit counts of search engines to determine co-occurrence information [28]. In our study we use Yahoo to gather hit count ( $HC$ ) information and the Dice coefficient to measure the similarity of term pairs  $sim_{DICE}(t_1, t_2) = \frac{HC(t_1, t_2)}{HC(t_1) + HC(t_2)}$ .

**Yahoo Holonyms.** When using part attributes we can explicitly use part-whole *holonym* relations. This can be achieved by querying the web for *holonym patterns*, such as “spots of Dalmatins” or “cat’s paw”. We use nine holonym patterns<sup>2</sup> suggested by [29], excluding the “in” patterns as they tend to denote non-visible parts. The measure is based on DICE, setting  $HC(t_1, t_2) := \sum_{i=1}^9 HC(pattern_i(t_1, t_2))$ .

**Yahoo Img / Flickr Img.** The web tends to be very noisy as unrelated terms might appear on the same web page leading to noisy hit count statistics. Using web image

<sup>2</sup> Nine holonym patterns: (1-2) **whole’s part**[s], (3-4) **wholes’ part**[s], (5-6) **part**[s] of a **whole**, (7-8) **part**[s] of the **whole**, (9) **parts** of **wholes**.

search instead allows reducing this noise as the evaluated text refers to the same entity, the image. Additionally we hope to get more visually relevant hit count results. We use Yahoo image search and search Flickr’s image tags and descriptions.

## 5 Experiments

In this section we evaluate the different language resources and SR measures (Sect. 4) for the various approaches for zero-shot classification (Sect. 3) on the Animals with Attributes (AwA) dataset [1].

The AwA dataset consists of 50 mammal object classes paired with an inventory of 85 attributes and corresponding object class-attribute associations [25, 24]. We follow the experimental protocol of [1], using the provided split into 40 training and 10 test classes (24,295 training, 6,180 test images). We also use the provided pre-computed feature descriptors, namely, RGB color histograms, SIFT, rgSIFT, PHOG, SURF, and local self-similarity histograms. In contrast to [1], we concatenate all features to a single vector instead of training independent SVMs. For computational reasons we depart slightly from [1] in our main experiments by down-sampling all training images to the minimum of 92 available images per class and using histogram intersection kernel SVMs instead of  $\chi^2$  kernel SVMs. For each attribute in the attribute-based model and each class in the direct model we train a SVM with an intersection kernel over the concatenated feature vectors. We use libSVM with the built-in probability estimates and a fixed cost parameter  $C=10$ , which has been found on a subset of the training classes by grid-search and cross-validation. To binarize the continuous SR values we threshold with the mean over all matrix values. We normalize the matrix values by dividing by column and row sums prior to binarization.

**Reproduction of the results in [1].** When using all available training images and a SVM with  $\chi^2$ -kernel, our implementation achieves 80.3% mean AUC and a multi-class classification accuracy of 40.3% which is very similar to 80.7% and 40.5% reported in [1]. Using intersection kernel and only 92 training images per class results in slightly lower performance with a mean AUC of 78.5% (Table 5, first row & first column) and accuracy of 34.7%. All following results are produced with this computationally more manageable setting.

**1. AwA attributes - mined object class-attribute associations.** We start by replacing the manual object class-attribute associations for the AwA attributes [1] with mined association, shown in the first row in Table 5. Comparing the different language resources and SR measures introduced in Sect. 4, we find the image-based measures (Yahoo Img and Flickr Img) and Wikipedia to perform best with a mean AUC of 71.0%, 70.1%, and 69.7%, respectively. This is expected as the image-based measures are based on image related text and thus inherently capture important correlations between terms and visual attributes. Wikipedia has shown to provide almost noise-free resource for computing semantic relatedness [27] which can apparently be transferred to our task.

WordNet (60.5%) and Yahoo Web (60.4%) perform worst with a significant drop in performance of about 10% compared to the first three language resources. We explain this drop for Yahoo Web by the increased level of noise, mainly incidental co-occurrences on web pages, compared to image search and Wikipedia. While WordNet is noise-free, the problem originates from the employed measure which is based on

	Manual	WordNet	Wikipedia	Yahoo Web	Yahoo Img	Flickr Img	Y. Holonyms
1. AWA attributes	78.5	60.5	69.7	60.4	<b>71.0</b>	70.1	–
2. Mined attributes	–	59.8	66.0	57.4	65.2	64.6	<b>69.9</b>
3. Objectness as attributes	–	71.2	66.4	66.7	<b>74.1</b>	68.5*	–
4. Direct similarity	–	73.4	76.6	77.7	<b>78.8</b>	77.8	–

**Table 1.** Mean Area under ROC curve (AUC) of the ten test classes in % for zero-shot classification on the AWA data set. The respective best result of all knowledge bases is shown in bold. \*See Sect. 5 for discussions.

proximity in WordNet’s hypernym hierarchy: Object classes and attributes are inherently different in nature and tend to lie in different subtrees of the hypernym hierarchy.

For a deeper understanding we examine the mined class-attribute associations which are the bases for the knowledge transfer. Judging if the associations are meaningful is not always easy, nevertheless we provide examples for the visual attribute “striped” to give an impression of the quality of mined similarities. In the following we list the four top ranked mammal classes for “striped” in decreasing order: Manual: *zebra, tiger, skunk, raccoon*; WordNet: *elephant, seal, mouse, bat*; Wikipedia: *zebra, skunk, tiger, Chihuahua*; Yahoo Web: *zebra, collie, Dalmatian, polar bear*; Yahoo Img: *zebra, skunk, tiger, Persian cat*; Flickr Img: *skunk, tiger, zebra, leopard*.

Overall we found that mined associations (at best 71.0%) perform considerably worse (7.5%) than manual defined associations (78.5%). Although we acknowledge this drop as significant we want to emphasize that the information given to the human judges was more descriptive than the simplified terms used for querying the language resources, e.g. “nest” abbreviates the attribute “keeping their young in a designated, enclosed area”. In connection with the significant reduction in manual supervision we consider the results a promising contribution to visual knowledge transfer.

**2. Mined attributes (and associations).** In contrast to the manually defined AWA attributes used in the first experiment, we use the mined (part) attributes which do not require any supervision (Table 5, 2<sup>nd</sup> row). Best is Yahoo Holonyms with an mean AUC of 69.9% reaching performance close to manual attributes. This variant of Yahoo Web is specifically targeted to part-whole relations by searching for specific patterns which reduce incidental co-occurrences compared to Yahoo Web.

Next are Wikipedia, Yahoo Img, and Flickr Img (66.0%, 65.2%, 64.6%). Last are, with a significant drop, WordNet 59.8% and Yahoo Web 57.4%. Disregarding Yahoo Holonyms, all language resources drop in performance compared to AWA attributes in the first experiment but the relative performance is in line with our previous observations. We explain this drop by the smaller number of attributes (74 instead of 85) and the decreased diversity of attributes (only part attributes).

**3. Objectness as attributes.** In our third experiment with use all 50 class names as attributes (Table 5, 3<sup>rd</sup> row). Yahoo Img is again best (74.1%), but in contrast to the first two experiments, WordNet follows closely as second best (71.2%): As objectness

requires the SR measures to compare similarity of objects rather than object-attribute associations the WordNet hierarchy provides an adequate model for similarity. Next are Yahoo Web (66.7%) and Wikipedia (66.4%). Although the average performance of Flickr Img is slightly above (68.5%), one test class was not associated with any attributes resulting in chance-level performance for this class. We explain this by insufficient statistics in the user provided text for co-occurring object class terms.

**4. Direct similarity.** Our fourth experiment evaluates our second, direct model (Sect. 3.2), which uses the five most similar training classes. Disregarding WordNet (73.4%) all language resources perform very similar and sometimes even on par with manually defined object class-attribute associations (Table 5, 4<sup>th</sup> row: Yahoo Img 78.8%, Flickr Img 77.8%, Yahoo Web 77.7%, Wikipedia 76.6%). This can be explained by the observation that the 5 most similar classes for a given class are very similar among the language resources and quite reliable. The direct similarity model additionally eliminates the need for an attribute layer and thus uses appropriate training data.

**Images of known training classes in the test set.** In all experiments so far the training and test class set have been disjoint as suggest by [1]. This means that a zero-shot classifier under test never has to distinguish known classes from the unseen ones. We expect this to be more difficult as it requires to classify those classes as negatives used as positive during training. To test this effect we add all images not used for training (due to downsampling) to the test set as negatives and report results for the best knowledge base for each of the above experiments.

The performance drops for objectness (74.1% to 67.7%) and direct similarity (78.8% to 76.0%) for added negatives. In contrast to this, the performance is stable for manual associations (78.5% to 78.9%) or increases for mined associations on AwA attributes (71.0% to 73.2%) and mined attributes (69.9% to 70.7%). We attribute this to the more general character of AwA attributes and mined attributes: While the inter-object class relatedness based approaches (objectness and direct similarity) group very similar classes (Fig 3) the object-attribute relatedness approaches tend to generalize over more diverse classes (Fig. 1(a)) which reduces the influence of specific (positive) training classes as negatives during testing.

## 6 Summary and Conclusions

Reducing supervision is vital for enabling knowledge transfer for a large number of classes. In this work we propose several approaches to fully replace manual intervention by tapping into linguistic knowledge bases<sup>1</sup>. In particular, overall best performance is achieved by the hit count based measure on Yahoo Image search outperforming most other measures on attribute-, objectness- and direct-similarity based approaches, reaching a performance on par with manually defined associations for the direct-similarity based approach. Due to a smaller coverage Flickr image is always slightly inferior. While Wikipedia performs similarly compared to image search, Yahoo Web and WordNet are especially inferior for attribute-based associations.

As part of future work we plan to apply our findings to data sets with thousands of classes where manual supervision is infeasible.

## References

1. Lampert, C.H., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: CVPR. (2009)
2. Stark, M., Goesele, M., Schiele, B.: A shape-based object class model for knowledge transfer. In: ICCV. (2009)
3. Bart, E., Ullman, S.: Single-example learning of novel classes using representation by similarity. In: BMVC. (2005)
4. Fink, M.: Object classification from a single example utilizing class relevance pseudometrics. In: NIPS. (2004)
5. Fellbaum, C.: WordNet: An Electronical Lexical Database. The MIT Press (1998)
6. Rohrbach, M., Stark, M., Szarvas, G., Gurevych, I., Schiele, B.: What Helps Where – And Why? Semantic Relatedness for Knowledge Transfer. In: CVPR. (2010)
7. Millet, C., Grefenstette, G., Bloch, I., Moëlllic, P.A., Hède, P.: Automatically populating an image ontology and semantic color filtering. In: OntoImage. (2006)
8. Ferrari, V., Zisserman, A.: Learning visual attributes. In: NIPS. (2007)
9. Wang, G., Forsyth, D.: Joint learning of visual attributes, object classes and visual saliency. In: ICCV. (2009)
10. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: ICCV. (2009)
11. Delezoide, B., Pitel, G., Borgne, H.L., Grefenstette, G., Mollic, P.A., Millet, C.: Object/background scene classification in photographs using linguistic statistics from the web. In: OntoImage. (2008)
12. Palatucci, M., Pomerleau, D., Hinton, G., Mitchell, T.: Zero-shot learning with semantic output codes. In: NIPS. (2009)
13. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.A.: Describing objects by their attributes. In: CVPR. (2009)
14. Zweig, A., Weinshall, D.: Exploiting object hierarchy: Combining models from different category levels. In: ICCV. (2007)
15. Marszalek, M., Schmid, C.: Semantic hierarchies for visual object recognition. CVPR (2007)
16. Popescu, A., Millet, C., Moëlllic, P.A.: Ontology driven content based image retrieval. In: CIVR, ACM (2007)
17. Wang, H., Jiang, X., Chia, L.T., Tan, A.H.: Ontology enhanced web image retrieval: aided by wikipedia & spreading activation theory. In: MIR. (2008)
18. Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. PAMI (2006)
19. Thrun, S.: Is learning the n-th thing any easier than learning the first. In: NIPS. (2006)
20. Barnard, K., Yanai, K.: Mutual information of words and pictures. In: ITA. (2006)
21. Boiy, E., Deschacht, K., Moens, M.F.: Learning visual entities and their visual attributes from text corpora. In: DEXA. (2008)
22. Barnard, K., Duygulu, P., Forsyth, D.A., de Freitas, N., Blei, D.M., Jordan, M.I.: Matching words and pictures. JMLR (2003)
23. Li, L.J., Socher, R., Fei-Fei, L.: Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In: CVPR. (2009) 2036–2043
24. Osherson, D.N., Stern, J., Wilkie, O., Stob, M., Smith, E.E.: Default probability. CS (1991)
25. Kemp, C., Tenenbaum, J.B., Griffiths, T.L., Yamada, T., Ueda, N.: Learning systems of concepts with an infinite relational model. In: AAAI. (2006)
26. Lin, D.: An information-theoretic definition of similarity. In: ICML. (1998)
27. Zesch, T., Gurevych, I.: Wisdom of crowds versus wisdom of linguists - measuring the semantic relatedness of words. JNLE (2010)
28. Kilgarriff, A., Grefenstette, G.: Introduction to the special issue on the web as corpus. CL'03
29. Berland, M., Charniak, E.: Finding parts in very large corpora. In: ACL. (1999)