

A Shape-Based Object Class Model for Knowledge Transfer

Michael Stark Michael Goesele Bernt Schiele
Computer Science Department, TU Darmstadt, Germany

Abstract

Object class models trained on hundreds or thousands of images have shown to enable robust detection. Transferring knowledge from such models to new object classes trained from a few or even as little as one training instance however is still in its infancy. This paper designs a shape-based model that allows to easily and explicitly transfer knowledge on three different levels: transfer of individual parts' shape and appearance information, transfer of local symmetry between parts, and transfer of part topology. Due to the factorized form of the model, knowledge can either be transferred for the complete model or just partial knowledge corresponding to certain aspects of the model. The experiments clearly demonstrate that the proposed model is competitive with the state-of-the-art and enables both full and partial knowledge transfer.

1. Introduction

Object class detection has made impressive progress in recent years. Most models rely on robust local features and powerful learning approaches such as SVMs requiring substantial amounts of training data per object class. In order to scale to larger numbers of object classes than it is possible today it is widely believed that information from one or a set of object classes should be transferred to the learning of new object classes. This would reduce the required training data for new object classes and might even facilitate 1-shot learning of novel classes. This idea of transfer learning has long been argued for both from a psychological point of view [1, 25] as well as from a computer vision point of view [4, 9, 10, 14, 22, 24, 38]. While these approaches have shown to enable object class learning from small numbers of training instances none of these models—as of today—has reached wide-spread use.

The starting point and main contribution of this paper is therefore to take a fresh look at the problem and to explicitly design a novel object model that directly lends itself to transfer learning. We start with the observation that there are at least three different types of knowledge that should be

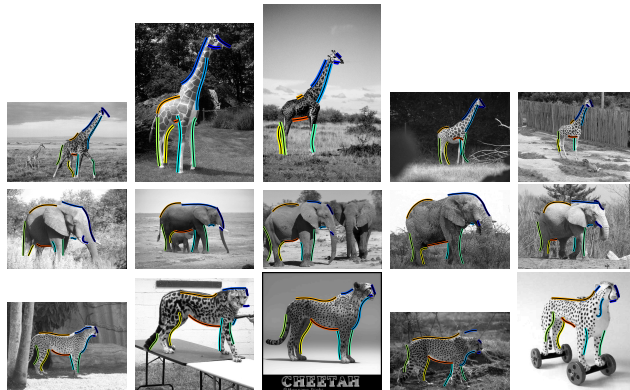


Figure 1. Animal detections using 1-shot models.

transferable between object models. First, the appearance or shape of an object part should be transferable (e.g., the shape of a leg or wheel). Second, local symmetries between parts are often shared by different object classes (e.g., the symmetry between front- and back-legs for quadrupeds). And third, the layout of the different parts is often at least partially shared among different object classes (e.g., the layout of head, torso and body for quadrupeds and birds, see also Sec. 5.2). In the following, we devise a part based model with a separate factor for each of these properties that allows, e.g., to transfer the layout of an object model either fully or only partially, constrained to an appropriate subset of object parts. The main contributions of our paper are:

- We propose a novel *shape-based object model for knowledge transfer* that can be factored into per-part components and enables transfer of full or partial knowledge.
- We demonstrate the *importance of symmetries*, a primitive rarely used for object detection, for both, object model learning as well as knowledge transfer.
- We experimentally show state-of-the-art performance of our object model on the ETHZ Shape Classes data set.
- We demonstrate that our model enables transfer of information on a quadrupeds database where we transfer the full layout and symmetry information. In addition, we also show successful partial information transfer in two interesting and quite different cases.

The remainder of this paper is organized as follows: After a review of related work we first introduce our model (Sect. 2) and validate its performance (Sect. 3). We then describe the knowledge transfer approach (Sect. 4) and demonstrate results for full and partial model transfer (Sect. 5). We conclude with an outlook on future work (Sect. 6).

1.1. Related Work

Transferring knowledge such as appearance, shape or symmetries between object classes is an important topic due to its potential to enable efficient learning of object models from a small number of training examples. It provides the basis for scalability to large numbers of classes. Broadly speaking, related work in knowledge transfer falls into three different categories: *distance metric learning*, *joint learning of multiple object classes*, and *use of prior information*.

The main idea of *distance metric learning* is to learn a representation for a set of a priori known classes in the form of a distance metric among them [14, 32]. This metric can then be used directly to classify instances of an unknown class. [3] replaces features from known classes with ones from a new class, implicitly re-using the learned distance metric. These approaches have shown to improve *1-shot learning* mainly for simple objects and handwritten chars.

In the context of *joint learning of multiple object classes*, machine learning has developed the notion of multiple task learning. This allows learners to benefit from the similarity of multiple, different learning tasks [5]. A second line of research is based on joint training of multiple classifiers, which draw from a common pool of features [2, 33], thereby enabling feature sharing. While these approaches clearly reduce the amount of necessary training data per object class, knowledge transfer happens rather implicitly. Explicit and controlled transfer of knowledge is not supported.

The *use of prior information* is most related to this work, and comes in multiple flavors. Levi et al. [22] use models of unrelated object classes to prime feature selection of an unknown class. Bart et al. [4] directly use similarities to known classes to represent new classes. Zweig and Weinsshall [38] propagate knowledge along a hierarchy, by learning and combining classifiers for individual levels of the hierarchy to yield a more effective classifier for specific leaf classes. Fei-Fei et al. [9] transfer information via a Bayesian prior on object class models, using knowledge from known classes as a generic regularizer for newly learned models.

While this paper clearly falls into the last category, we stress the following key differences from related work: most importantly, our approach is designed to allow an explicit, controlled transfer of prior knowledge. In particular, it facilitates knowledge transfer at both the level of a full model, and selected aspects of a model. Being based on an assembly of parts, their spatial layout and symmetry relations provides a rich source of independently transferable properties,

ranging from fairly general (overall spatial layout of parts) to very specific (local part shape). We strongly believe that both the explicit as well as the partial transferability of prior information are key ingredients to make knowledge transfer a common tool for object class modeling and learning.

Concerning object recognition, our work is most related to part-based methods such as the Constellation Model [11] or the Implicit Shape Model [20]. While the non-parametric scene-object-part model of [31] requires less supervision than ours, its appearance-based, visual word part representation is limited compared to our flexible combination of local shape and semi-local symmetry relations. [36] gives a fundamental treatment of probabilistic shape modeling and Gestalt principles, including symmetry. [27] evaluates the accuracy of several symmetry detection algorithms. In contrast to early attempts [7, 26], our paper shows the successful application of a particular kind of symmetry relations [6, 29] to object class detection in real images.

2. The Model

Our model is inspired by the *Constellation Model* [11], but goes beyond this model in several ways. First, it relies entirely on shape information. Second, we propose a Data-Driven Markov Chain Monte Carlo (DDMCMC) [37] technique for efficient inference, which increases the number of features the system can handle by several orders of magnitude. Third, we enrich the original formulation comprising object parts, their relative scales, and spatial layout, by pair-wise symmetry relations between parts. Pair-wise relations even between simplistic line features have proven to be powerful cues for recognition [21], which we confirm in our experiments. Fourth, we demonstrate that knowledge can be effectively transferred between different model instances, on two different levels.

2.1. Local Shape Features

We introduce a novel flavor of local shape features, which constitute a discrete, over-complete representation of image contours. The shape features are based on the *Contour Segment Network* (CSN) [13], and its associated local companions, *k-Adjacent Segments* (*k-AS*) [12]. We suggest important additions to these techniques, as detailed below.

Starting from an edge probability map of the Berkeley natural boundary detector [23], a collection of discrete, roughly straight contour segments is formed, and subsumed in a network topology (the CSN), based on spatial proximity and edge continuity constraints. Since, by design, the CSN can be assumed to provide an over-segmentation of image edges, meaning that object parts are likely to be fragmented into several segments, we *simultaneously* include *k-AS* with $k \in \{1, \dots, K\}$ into our representation, to increase the chance of having a shape feature available that matches one-to-one to an object part. In practice, we use $K = 5$.

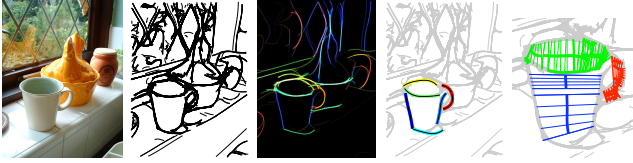


Figure 2. From left to right: Original image, local shape features, color-coded part likelihoods, detection hypothesis, selected symmetry lines and axes.

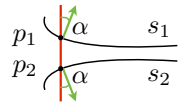
Further, we *unify* the representation of k -AS for varying k by fitting a parametric B-spline curve to all constituent edgel chains, using the exact same parameterization, independent of k . This offers the additional benefit of retaining the original *curvature* information and increasing the discriminative power of the features compared to the original k -AS represented by straight line approximations.

In our implementation, we first transform all constituent edgel chains of a given k -AS into a translation and scale invariant space, using Procrustes analysis [8]. We use the resulting spline parameters as a low-dimensional local shape description. In all experiments, we use pairs of quadratic B-splines, resulting in an 8-dimensional descriptor. We prune the set of features based on the goodness of fit of the splines. Fig. 2 shows all 1640 local shape features of an image.

2.2. Semi-Local Symmetry Relations

As shown in the literature [13, 30], local shape features based on contour segments tend to be more generic in nature than local texture features, and hence provide relatively weak discrimination among object parts and background clutter, if used in isolation. We therefore include another powerful perceptual cue into our model, which relates pairs of local shape features by identifying and describing *symmetries* between them. In particular, we use a B-spline-based implementation [29] of *Smoothed Local Symmetry* (SLS). SLS were originally proposed by [6] in the context of planar shape analysis.

SLS relate two parametric shapes by determining pairs of points that fulfill a *local symmetry* constraint: A point p_1 on shape s_1 is locally symmetric to a point p_2 on s_2 , if the respective angles between the connecting line between p_1 and p_2 , and the normal vectors at p_1 and p_2 , are equal. The set of all locally symmetric point pairs and their associated connecting lines (the *symmetry lines*) then define the *symmetry axis* between the shapes: it consists of the mid-points of the symmetry lines. Fig. 2 (right) depicts several selected symmetry lines and axes between local shape features of a mug (blue: symmetries between side-wall features, green: between rim features, red: between handle features).



Starting from the spline-based representation of SLS, we

now devise a semi-local symmetry descriptor, which captures both the shape of the symmetry axis and the lengths of the symmetry lines, in order to characterize the symmetry. The first is achieved by representing the axis as a local shape feature, exactly as described in Sect. 2.1. We compute a fixed number of symmetry lines (usually 10) and record a profile of their respective lengths, as we traverse the symmetry axis from end to end. We then reduce the dimensionality of the resulting length profile vector by PCA (usually to 3). Fig. 7 (b) depicts length profiles as bar plots corresponding to the symmetry axes denoted by gray lines in Fig. 7 (a).

2.3. Probabilistic Model

We now describe the probabilistic model that subsumes individual part shapes S , binary symmetry relations B , relative part scales R , and their overall spatial layout X . We borrow from the notation of [11] where appropriate.

During detection, our goal is to find an *assignment* of all P model parts to local shape features, which we denote the detection hypothesis $H = (h_1, \dots, h_P)$. That is, h_p contains a local shape feature identifier assigned to part p . We formulate the detection problem as a maximum a posteriori hypothesis search over the distribution $p(X, R, S, B, H|\theta)$, which is the joint posterior distribution of H and image evidence, given a learned model θ . It factors as follows:

$$p(X, R, S, B, H|\theta) = \underbrace{p(S|H, \theta)}_{\text{Local Shape}} \underbrace{p(B|H, \theta)}_{\text{Symm. Rel.}} \underbrace{p(X|H, \theta)}_{\text{Layout}} \underbrace{p(R|H, \theta)}_{\text{Rel. Scale}} \underbrace{p(H|\theta)}_{\text{Prior}} \quad (1)$$

In all experiments, we assume a uniform prior $p(H|\theta)$.

Local Part Shape. Local part shape $S(h_p)$ is modeled by a Gaussian density on spline parameters (see Sect. 2.1).

$$p(S|H, \theta) = \prod_{p=1}^P \mathcal{N}(S(h_p)|\theta). \quad (2)$$

Binary Symmetry Relations. We instantiate the binary relation component of our model with a joint density over SLS descriptors, as described in Sect. 2.2. It comprises all pairs of parts, excluding self- and duplicate pairings. For each pair, it factors into two Gaussian densities, where one governs the SLS axis spline parameters $B_a(h_i, h_j)$, and one the PCA-projection of the corresponding symmetry line length profile $B_l(h_i, h_j)$.

$$p(B|H, \theta) = \prod_{i=1}^{P-1} \prod_{j=i+1}^P p(B(h_i, h_j)|\theta)$$

$$p(B(h_i, h_j)|\theta) = \mathcal{N}(B_a(h_i, h_j)|\theta) \mathcal{N}(B_l(h_i, h_j)|\theta) \quad (3)$$

Spatial Layout and Relative Scales. We model the spatial layout of constituent model parts as a joint Gaussian distribution over their coordinates $X(H)$ in a translation- and scale-invariant space (the *constellation*), again using Procrustes analysis [8]. The model allocates independent Gaussians for the relative scale $R(h_p)$ of each part, i.e., the ratio

between part and constellation scale.

$$p(X|H, \theta) p(R|H, \theta) = \mathcal{N}(X(H)|\theta) \prod_{p=1}^P \mathcal{N}(R(h_p)|\theta) \quad (4)$$

2.4. Learning and Inference

Learning. We learn maximum likelihood model parameters θ for all model components using supervised training. Supervision is provided by labeling contour segments in training images (see Sect. 2.1), which in practice amounts to a few mouse clicks per object instance.

Inference. During detection, we search for $H_{\text{MAP}} = \arg \max_H p(H|X, R, S, B, \theta)$, the maximum a posteriori hypothesis. This is equivalent to $\arg \max_H p(X, R, S, B, H|\theta)$. We approximate H_{MAP} by drawing samples from $p(X, R, S, B, H|\theta)$ using the Metropolis-Hastings (MH) algorithm [17]. We use the Single Component update variant of MH, since it allows to separately update individual components of the target density, conditioned on the remaining portion of the current state of the Markov chain. This opens the possibility to guide the sampling towards high density regions by data-driven, bottom-up proposals [34, 37]. Similar to [19], we define P independent proposal distributions of the form $q_p(S(h_p)|\theta) = \mathcal{N}(S(h_p)|\theta)$, based on the likelihoods of the local shape part model. Fig. 2 depicts a joint, color-coded visualization of all part proposals for a mug model consisting of 7 parts (two side-walls, two rim parts, one bottom part, two handle parts), together with an example detection based on exactly these proposals. Notably, the combined part likelihood is much sparser than the corresponding visualization of all local shape features to the left of Fig. 2.

We obtain the following acceptance ratio for changing the current hypothesis $H = (H_{-p}, h_p)$ to $H' = (H_{-p}, h'_p)$, where H deviates from H' only in component h_p , and H_{-p} denotes the other components that are kept.

$$\alpha = \min \left\{ 1, \frac{p(X, R, S, B, h'_p|H_{-p}, \theta) q_p(S(h_p)|\theta)}{p(X, R, S, B, h_p|H_{-p}, \theta) q_p(S(h'_p)|\theta)} \right\} \quad (5)$$

Note that most of the terms in this ratio actually cancel due to the factorization of our model (namely the ones not involving the part under consideration p). This implies in particular that the number of pair-wise relations that have to be computed per iteration grows only linearly, and not quadratically, with increasing number of parts P . Further, since the sampling process is guided by data-driven proposals, the number of pair-wise relations considered is orders of magnitudes smaller than the number of all possible pairings. We exploit this fact by computing SLS in a lazy fashion, and subsequently caching them, which greatly improves runtime behavior. For a typical image with several thousands of features, our model typically (re-) considers at most a few tens of thousands of pairs, not tens of millions.

Detection. We detect object instances by running m independent Markov chains, and memorizing the per-chain highest-scoring hypotheses. In all experiments, we run $m = 50$ chains, for a maximum number of 1000 iterations, yielding runtimes of under a second per Markov chain. We use the greedy non-maximum suppression described in [16] to prune overlapping hypotheses.

3. Shape Classes Experiments

We evaluate the performance of our model on a standard shape benchmark [13] and report detection results on 4 of the 5 classes of the ETHZ Shape Classes data set (see Fig. 3 and 4). We use the test protocol of [12]: experiments are conducted in 5-fold cross-validation. For each class, we learn 5 different models by sampling 5 subsets of half of the class images at random. The test set for a model then consists of all other images in the data set (taken from all 5 classes). Performance is measured as the average detection rate at 0.4 false positives per image (FPPI). We compare the results with two state-of-the-art methods, one shape-based [12], and one based on topic-decompositions of HOG-like features [16]. For [12], we consider the results based on *learned* models rather than the ones based on hand-drawings, as they are comparable to our approach. For the same reason we do not compare against [28, 35].

As shown in Fig. 3, our model without symmetry significantly outperforms previous results on *bottles* and is slightly better on *giraffes*. For *mugs* however the performance is lower and for *swans* it is between [12] and [16]. On average it outperforms both methods. We attribute this state-of-the-art performance to the combination of robust, discriminative local shape features with a flexible spatial model.

Adding symmetry relations (SLS) significantly increases performance for two classes (11% for *swans*, 8% for *mugs*) and also slightly for *bottles* (3%). As a consequence our model performs better than previous work on all four classes. Using symmetries attains 89.9% on average, 6.4% better than the next best related method.

4. Knowledge Transfer

In the following, we highlight two different levels of inter-model knowledge transfer supported by our model. First, we show that a full model, learned for a known class A , can be readily transferred to a new but related object class B . An object model for this new class B is obtained from one or a few training instances plus the transferred knowledge from the previously learned object class A . Second, we also show *partial knowledge transfer* by restricting a previously learned model A to a proper subset of parts, retaining all knowledge about their spatial configuration, relative scales, and symmetry relations. The resulting partial model can be transferred to a new class B for which only a few training instances are available.

Results	Bottle	Giraffe	Mug	Swan	average
Ferrari et al. [12]	83.2 (7.5)	58.6 (14.6)	83.6 (8.6)	75.4 (13.4)	75.2 (11.0)
Fritz et al. [16]	76.8 (6.1)	90.5 (5.4)	82.7 (5.1)	84.0 (8.4)	83.5 (6.3)
Our model	91.0 (3.8)	91.7 (4.1)	76.6 (9.9)	77.7 (5.8)	84.3 (5.9)
Our model, SLS	94.4 (3.8)	91.7 (2.6)	84.5 (4.7)	88.8 (6.9)	89.9 (4.5)

Figure 3. ETHZ Shape Classes results: average detection rates, standard deviations given in brackets.

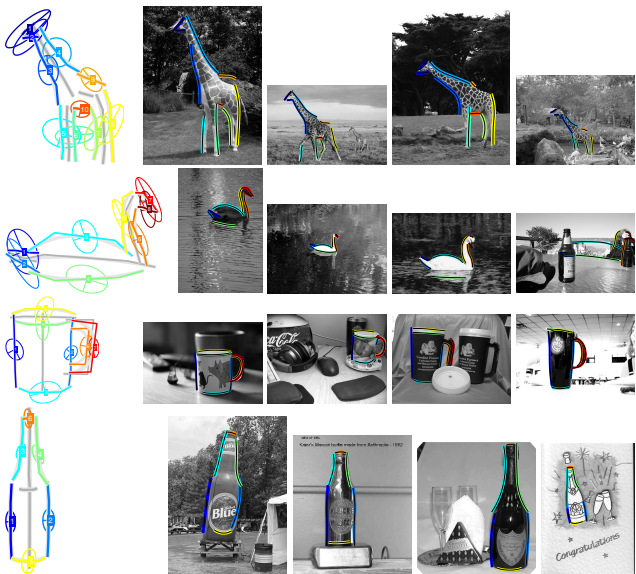


Figure 4. Learned ETHZ Shape Classes models (left) and example detections (right). For models, mean local part shapes and selected mean symmetry axes are drawn at mean positions and relative scales. Covariances of part positions are shown as ellipses.

4.1. Full Model Transfer

Our approach to combine prior knowledge and data is inspired by, but not strictly adhering to, the Bayesian paradigm. Instead of deriving a posterior distribution over models, given data D , from a prior and corresponding likelihood $p(\theta|D) \propto p(\theta) \times p(D|\theta)$, we follow the simpler route of directly combining and manipulating components of models that we have learned. These manipulations are valid because of the specific factorization and parametric forms of involved distributions. In particular, since all distributions are Gaussian, we can manipulate means and covariances separately, and can restrict models to subsets of parts by marginalizing out the ones we are not interested in.

Let $m_A(\mu_A, \Sigma_A)$ and $m_B(\mu_B, \Sigma_B)$ be two models, where m_A is the *base model*, i.e., the model which we want to transfer, and m_B a model learned from k training instances of class B . We denote m_B a k -shot model. Now the question arises which knowledge should be transferred from m_A to obtain a more powerful model for class B . Consider, e.g., the case that class A corresponds to *horses* and class B

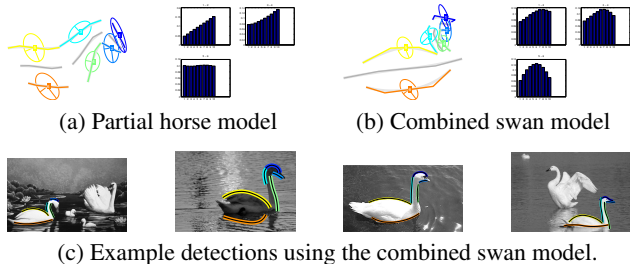


Figure 5. Partial transfer models (a)(b), and I -shot detections (c).

corresponds to *giraffes*. While the mean of the overall object shape is different the variation in object shape is similar as both classes belong to the class *quadrupeds*. Therefore we derive a combined model $m_{AB}(\mu_{AB}, \Sigma_{AB})$ for class B by taking μ_{AB} to be μ_B , and Σ_{AB} to be a weighted combination of Σ_A and Σ_B . For $k = 1$, we set $\Sigma_{AB} = \Sigma_A$. The experiments in Sect. 5.1 show results of this procedure.

4.2. Partial Model Transfer

The factorization of our model into separate components for local part shape, relative scales, symmetry relations, and the overall spatial layout facilitates keeping subsets of parts, while discarding others. For part shape as well as relative scale components, we keep all relevant part contributions. For symmetry relations, we keep all contributions involving at least two relevant parts. For spatial layout, we can marginalize out all irrelevant parts.

To realize the importance of *partial knowledge transfer* consider the following example (see also Sect. 5.2). Let us assume class A again corresponds to *horses* and class B corresponds to *swans*. As the first is a quadruped and the second is a bird, one might see little opportunity for knowledge transfer, since global object shape is different.

However, there is indeed *partial knowledge* that can be transferred, namely the topology of a subset of parts (head, neck, and torso). As this information is contained in the *horse* model, we may first extract the corresponding relevant portion by marginalization, and then transfer this partial knowledge. The experimental section shows the usefulness of such partial knowledge transfer, which we argue to be a very general and versatile concept, as many parts and constellations of parts reoccur across many object classes. Therefore, use of such *partial knowledge transfer* about constellation, local shape and symmetries of object parts and part ensembles is a powerful tool to enable scalability to large numbers of object classes.

5. Knowledge Transfer Experiments

We demonstrate the ability of our approach to effectively transfer knowledge between models by a series of recognition experiments based on the animal quadruped classes *horse*, *elephant*, *cheetah*, and *giraffe* for which we com-

binned images from the Mammal Images Benchmark [15], the Corel data base, INRIA Horses [18], and additional images from the web. Images show quadrupeds roughly pose-aligned, but at varying scales, and contain considerable background clutter (see Fig. 1). While all quadrupeds share a common topology (head, neck, torso, and four legs), they vary significantly in the concrete embodiment, leading to variations in both the appearance of individual body parts as well as their spatial layout. In addition, we use the *swan*, *mug*, and *bottle* classes from the ETHZ Shape Classes data set in Sect. 5.2 for partial knowledge transfer.

All experiments follow this protocol: Models are learned from a set of training images of a given class and evaluated on a test set consisting of images containing at least one instance of that class, and a comparable number of background images not containing any class instances. Performance is evaluated in a retrieval setting where we run detection for each test image and record the highest scoring hypothesis. For each n between 1 and the number of test images, we plot the fraction of images belonging to the class in the n highest scoring ones.

5.1. Full Model Transfer

Using the quadruped classes, we show that prior knowledge about the general stature of a quadruped can be used to bootstrap specialized quadruped detectors. In particular, we learn a *base model* from all 170 positive INRIA *horse* images, which we assume to yield a reasonable model of quadruped stature (see Fig. 7). We then use k training images of another quadruped class and learn a k -shot model from these images. The models are combined as described in Sect. 4 and the combined model is evaluated as above. We found experimentally that the weighting of the individual models has little impact on performance and thus report all results for uniform weighting.

Fig. 6 gives recognition results for the classes *elephant*, *cheetah*, and *giraffe* without and with symmetry relations. Each plot compares the performance of combinations of the base model with k -shot models learned from $k \in \{1, 5, 10, 25\}$ training images, the *full model* learned from all available training images of the target class, and the base model alone. The curves for $k \in \{1, 5, 10\}$ are averaged over 5 different random choices of k training images among the full 25 training images available for each class.

We first observe in Fig. 6 that the base model learned entirely from *horse* images performs surprisingly well on *elephants* and *cheetahs* despite major differences in appearance. It can therefore be transferred directly even without a single training image. This can be explained by the fact that the *horse* model already captures a fair amount of variations in the shape and spatial layout of *elephants* and *cheetahs*. This is also confirmed in Fig. 7 and 8: means and covariances of part shape as well as constellations of full *elephant*

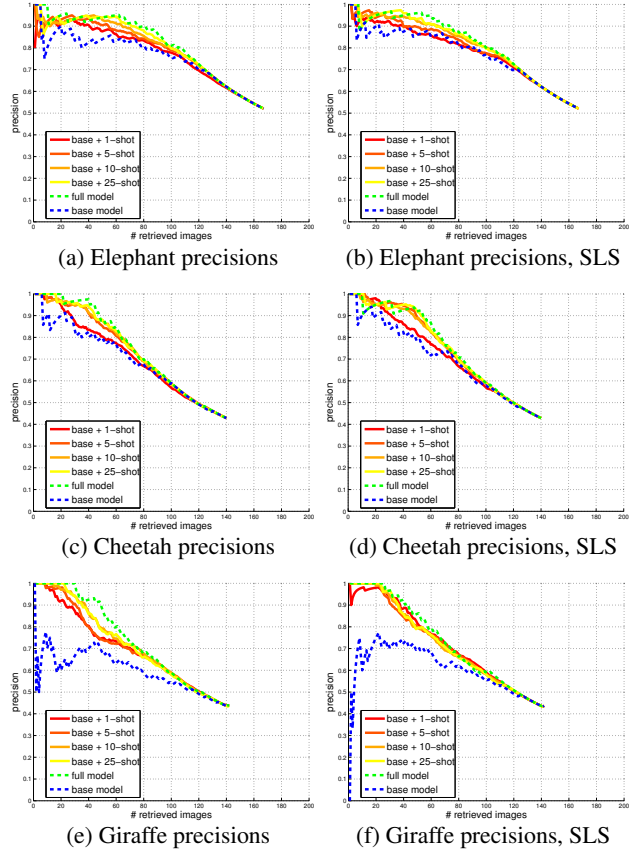


Figure 6. Full model transfer recognition results without symmetry relations (left) and with symmetry relations (SLS, right).

and *cheetah* models are visually close to the horse base model (Fig. 7). Furthermore, all shown symmetry distance models share common properties, namely, an almost linear increase in distance between head parts (1-2), quadratic dependency between pairs of leg parts (5-6 and 7-8), and the almost flat shape of the torso (9-10).

Adding training images clearly improves precision and adapts models to the target classes. A small number of training images (5 for *cheetah* and 10 for *elephants*) is sufficient to achieve a performance that is largely equivalent to the corresponding full model. Fig. 8 confirms this observation: combinations of 5-shot and base models (middle column) are visually close to the corresponding full models (right column) and can thus be expected to behave comparably.

Interestingly, the base model performs poorly for the *Giraffe* class as the full *giraffe* model differs quite strongly from the *horse* base model (e.g., the neck parts, see Fig. 8(c)). Note however, that even a single training image is sufficient to boost the performance to almost the level of the full model. This is particularly pronounced for added symmetry relations, and explained by the high degree of similarity among all symmetry distance models.

In order to understand the role of the base model, we

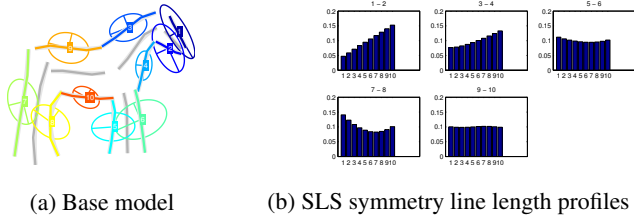
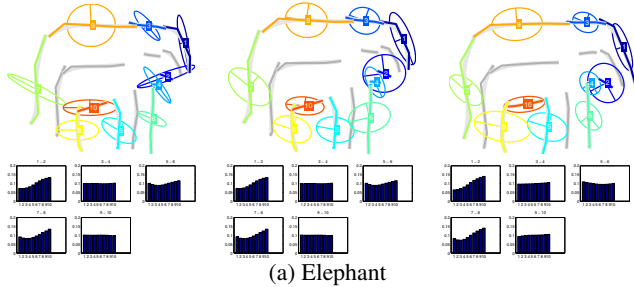
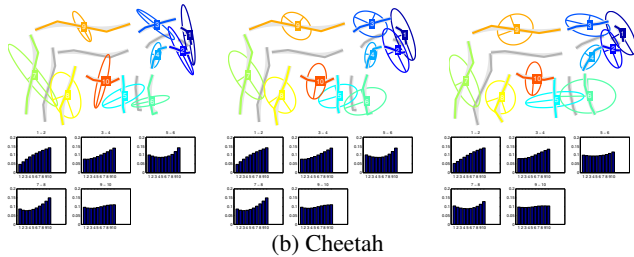


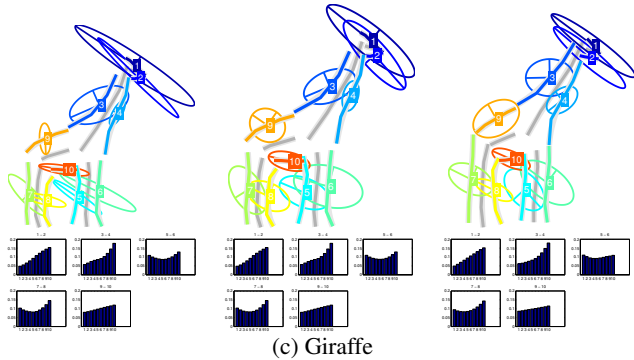
Figure 7. The horse base model used in the k -shot experiments of Sect. 5. For clarity, we show only a subset of symmetry relations: numbers above plots in (b) refer to pairs of part numbers in (a).



(a) Elephant



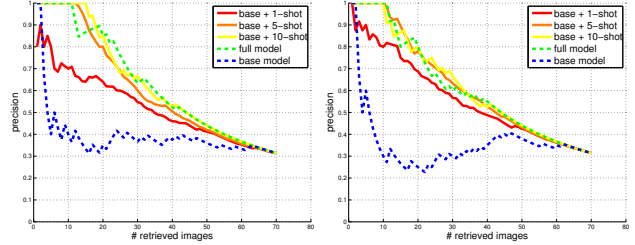
(b) Cheetah



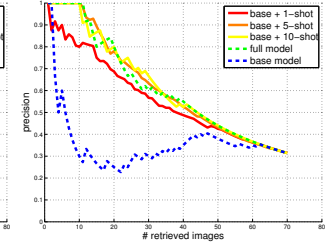
(c) Giraffe

Figure 8. Animal models: left: 5-shot model, middle: 5-shot model combined with base model, right: full model. Note the similarities between the models on the right and in the middle.

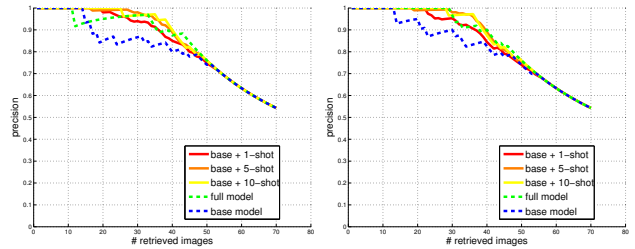
further compared the performance of combined l -shot models with l -shot models using isotropic regularization, which we determined empirically on a separate data set (ETHZ Shape). Even though these models can perform on a similar scale as the combined models, they tend to be slightly worse on average, and introduce the disadvantage of having to choose a suitable regularizer, while regularization comes for free with a base model.



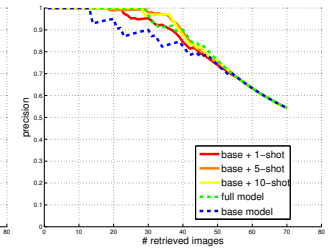
(a) Swan precisions



(b) Swan precisions, SLS



(c) Bottle precisions



(d) Bottle precisions, SLS

Figure 9. Partial model transfer recognition results without symmetry relations (left) and with symmetry relations (SLS, right).

To summarize, knowledge transfer with a suitable base model clearly reduces the number of required training images in all cases. k -shot models including symmetry relations between parts are often superior. Also, the variance of the curves including symmetry relations exhibit less variation, in particular for *giraffes* (Fig. 6(f)) clearly showing the importance of symmetry relations for knowledge transfer.

5.2. Partial Model Transfer

For partial model transfer, we restrict a base model to a proper subset of parts and combine this restricted base model with a k -shot model of a new class. As mentioned before we can transfer partial knowledge of a *horse* base model to the *swan* class. Therefore, in a first experiment, we restrict a *horse* model to head, neck, and torso parts and then combine this restricted base model with k -shot models of *swans* (see Fig. 5). In the second experiment we transfer partial knowledge of a *mug* base model to the *bottle* class. For this we restrict the *mug* model to the sidewall and bottom parts, discarding handle and upper rim parts and combine this with k -shot models of *bottles*. As before we report retrieval performance for *swan* and *bottle* images respectively.

From Fig. 9(a) and (b), it is immediately apparent that the restricted *horse* base model performs only at chance level for the *swan* retrieval, both with and without symmetries. Strikingly, adding a *single image* of a *swan* drastically improves detection rate (base + 1-shot). As before, adding only a handful of images to the restricted base model yields performance close to the full model. Likewise, adding symmetries to the model is highly beneficial. In particular, the combined *swan* 1-shot model benefits significantly ($\approx 10\%$) from including symmetry relations.

In the second experiment (Fig. 9(c) and (d)) the *mug* base model already enables to retrieve *bottle* images quite well. This is due to the fact that the two classes not only share several common parts, but their shape is also similar. In this case, already a single training example is sufficient to reach the performance level of the corresponding full models.

From these experiments we can conclude that our model does indeed allow for partial knowledge transfer and enables to train object models from few training images. In cases where object classes share many similarities (*mug-bottle-transfer*) as little as one training instance can suffice. For larger variations between objects (*horse-swan-transfer*) using only five training instances can yield a good model.

6. Conclusions and Future Work

While pioneering work on knowledge transfer for object class model training exists, none of it has been adopted widely. Despite this fact, we strongly believe that knowledge transfer is an important ingredient to enable learning and recognition of large numbers of object classes. As demonstrated by our results, our shape-based model enables explicit knowledge transfer between object classes thereby drastically simplifying training for new object classes. The model's ability to transfer individual components makes our approach applicable to a large number of scenarios. Its competitive results on the ETHZ Shape Classes confirm the validity of the object model formulation for object class detection. The use of local symmetries improves the performance both for detection and model transfer significantly although symmetries are so far seldom used for object detection.

Since both, the model as well as the inference method, can be easily extended to larger number of parts and to include other complementary features, we believe that it presents large opportunities for future work.

Acknowledgements. This work has been funded, in part, by EU project CoSy (IST-2002-004250), and DFG Emmy Noether grant GO1752/3-1.

References

- [1] W.-K. Ahn and W. Brewer. Psychological studies of explanation-based learning. In G. DeJong, editor, *Investigating Explanation-Based Learning*. Kluwer, 1993.
- [2] Y. Amit, M. Fink, N. Srebro, and S. Ullman. Uncovering shared structures in multiclass classification. In *ICML*, 2007.
- [3] E. Bart and S. Ullman. Cross-generalization: Learning novel classes from a single example by feature replacement. In *CVPR*, 2005.
- [4] E. Bart and S. Ullman. Single-example learning of novel classes using representation by similarity. In *BMVC*, 2005.
- [5] S. Ben-David and R. Schuller. Exploiting task relatedness for multiple task learning. In *COLT*, 2003.
- [6] M. Brady and H. Asada. Smoothed local symmetries and their implementation. In *IJRR*, 1984.
- [7] R. Brooks. Model-based three dimensional interpretation of two dimensional images. *PAMI*'83.
- [8] T. Cootes. An introduction to active shape models, 2000.
- [9] L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *PAMI*, 28(4):594–611, 2006.
- [10] A. Ferencz, E. Learned-Miller, and J. Malik. Building a Classification Cascade for Visual Identification from One Example. In *ICCV'05*.
- [11] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, 2003.
- [12] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detection with deformable shape models learnt from images. In *CVPR*, 2007.
- [13] V. Ferrari, T. Tuytelaars, and L. J. V. Gool. Object detection by contour segment networks. In *ECCV*, 2006.
- [14] M. Fink. Object classification from a single example utilizing class relevance pseudo-metrics. In *NIPS*, 2004.
- [15] M. Fink and S. Ullman. From aardvark to zorro: A benchmark for mammal image classification. *IJCV*, 77(1-3):143–156, 2008.
- [16] M. Fritz and B. Schiele. Decomposition, discovery and detection of visual categories using topic models. In *CVPR*, 2008.
- [17] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo In Practice*. 1996.
- [18] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *CVPR*, 2004.
- [19] M. W. Lee and I. Cohen. Proposal maps driven mcmc for estimating human body pose in static images. In *CVPR*, 2004.
- [20] B. Leibe, A. Leonardis, and B. Schiele. An implicit shape model for combined object categorization and segmentation. In *Toward Category-Level Object Recognition*, 2006.
- [21] M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *CVPR*, 2007.
- [22] K. Levi, M. Fink, and Y. Weiss. Learning from a small number of training examples by exploiting object categories. In *Workshop of Learning in Computer Vision*, 2004.
- [23] D. R. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 26(5):530–549, 2004.
- [24] E. Miller, N. Matsakis, and P. Viola. Learning from One Example Through Shared Densities on Transforms. In *CVPR*, 2000.
- [25] Y. Moses, S. Ullman, and S. Edelman. Generalization across changes in illumination and viewing position in upright and inverted faces. Tech. rep., Weizmann Institute, 1993.
- [26] R. Nevatia and T. Binford. Description and recognition of curved objects. *USC Computer Vision*, 1977.
- [27] M. Park, S. Lee, P.-C. Chen, S. Kashyap, A. A. Butt, and Y. Liu. Performance evaluation of state-of-the-art discrete symmetry detection algorithms. In *CVPR*, 2008.
- [28] S. Ravishanker, A. Jain, and A. Mittal. Multi-stage contour based detection of deformable objects. In *ECCV*, 2008.
- [29] P. Saint-Marc, H. Rom, and G. Medioni. B-spline contour representation and symmetry detection. *PAMI*, 15(11):1191–97, 1993.
- [30] M. Stark and B. Schiele. How good are local features for classes of geometric objects. In *ICCV*, 2007.
- [31] E. Sudderth, A. Torralba, W. T. Freeman, and A. Willsky. Describing visual scenes using transformed objects and parts. *IJCV*, 2008.
- [32] S. Thrun. Is learning the n-th thing any easier than learning the first. In *NIPS*, 2006.
- [33] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. In *CVPR*, 2004.
- [34] Z. Tu, X. Chen, A. Yuille, and S. Zhu. Image parsing: Unifying segmentation, detection and recognition. *IJCV*, 2005.
- [35] Q. Zhu, L. Wang, Y. Wu, and J. Shi. Contour context selection for object detection: A set-to-set contour matching approach. In *ECCV'08*.
- [36] S.-C. Zhu. Embedding gestalt laws in markov random fields. In *PAMI*, 1999.
- [37] S.-C. Zhu, R. Zhang, and Z. Tu. Integrating bottom-up/top-down for object recognition by data driven markov chain monte carlo. In *CVPR*, 2000.
- [38] A. Zweig and D. Weinshall. Exploiting object hierarchy: Combining models from different category levels. In *ICCV*, 2007.