Hierarchical Transfer of Semantic Attributes

Ziad Al-Halah Rainer Stiefelhagen
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology
{ziad.al-halah, rainer.stiefelhagen}@kit.edu

1. Introduction

In the prevailing approach, attributes are learned from all seen classes and then reused to describe or classify an unseen one. However, this doesn't account for the high intra-attribute variance. Using all the seen classes helps in learning visual semantics in a very abstract manner. Hence, subsets of classes that share similar attributes cannot be distinguished easily. Eventually, the fine properties of the attribute that help in discriminating a group from another are lost when it is learned from all the classes. Consider for example the attribute *beak*. The global attribute model would learn that a beak is an elongated extension at a certain position relative to the head; *i.e.* ignoring the distinctive long thin beak shape of the hummingbird species or the wide curved-end of the albatross species. In other words, the global model does not take advantage of the rich information already available in the source dataset. This results in transferring less discriminative attributes to the novel class. On the other hand, capturing these specific properties of *beak* relative to each subgroup of birds is beneficial. It gives us the option to select the most proper type of *beak* to share with the unseen class. Accordingly, knowing that both *Gull* and *Albatross* are *Seabirds*, it is intuitive and probably more discriminative to describe the beak of the *California-Gull* as an *albatross-like-beak*.

2. Approach

Hierarchical representation of concepts and objects is part of the human understanding of the surrounding world. This helps us to better learn the commonality as well as the differences in and across groups. The key idea of our approach is to take advantage of the embedded structure in the object category space and extend the notion of global attributes to include different levels of abstraction. The object hierarchy groups the classes based on their overall visual similarity; thus provides a natural way to guide the transfer process to share information from the knowledge sources that will most likely contain relative information. In the following, we describe the three main steps of our Hierarchical Attribute Transfer approach (HAT) [2].

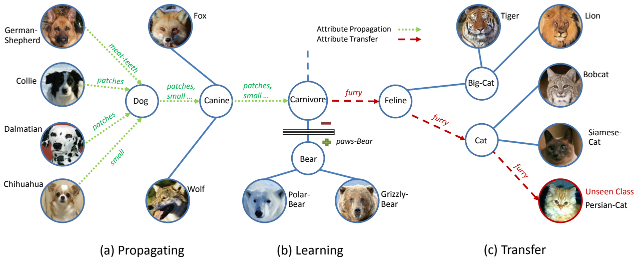


Figure 1: Illustrative figure of our HAT model.

1) Populating the hierarchy with attributes We exploit the object hierarchy $\mathcal{H}$ by transferring the attributes annotation in a bottom-up approach from the seen classes $\mathcal{Q}$ to the root (Figure 1a). The active attributes of node n_j are:

$$a_m^{n_j} = 1 \text{ if } \exists a_m^{n_i} = 1 \text{ and } n_i \in \text{child}(n_j), \tag{1}$$

where $\text{child}(n)$ is the set of nodes of the subtree rooted with n. Consequently, the root node of $\mathcal{H}$ will be described with all attributes of $\mathcal{Q}$.

2) Learning at different levels of abstraction To learn the various attributes classifiers, we first define the support set of an attribute a_m, *i.e.* the set of samples that provide evidence of a_m. An attribute $a_m^{n_j}$ in the hierarchy has the support set $\text{supp}(a_m^{n_j})$. The set contains samples labeled with the attribute of that class ($\text{lbl}(a_m^{n_j}) : n_j \in \mathcal{Q}$), and additionally the samples of its children which share the same attribute with n_j, *i.e.*

$$\text{supp}(a_m^{n_j}) = \bigcup_{n_i \in \text{child}(n_j)} \text{supp}(a_m^{n_i}) \cup \text{lbl}(a_m^{n_j}). \tag{2}$$

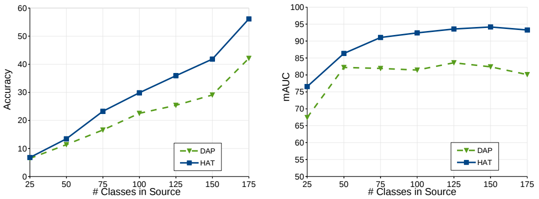
To capture the fine differences that characterize an attribute at node n, we use a child-vs-parent learning scheme Figure 1b. The attribute $a_m^{n_c}$ is learned with the following positive (T_P) and negative (T_N) sets

$$T_P = \text{supp}(a_m^{n_c}) \quad T_N = \text{supp}(a_m^{n_p}) - \text{supp}(a_m^{n_c}), \tag{3}$$

where n_p is the parent node of n_c.

3) Hierarchical transfer We then leverage the hierarchy to guide the knowledge transfer process and find the proper attributes to transfer to novel classes. Accordingly, for a novel class z_l in $\mathcal{H}$, we transfer the attributes of its ancestors across the different levels of abstraction (*e.g.* $a_{furry}^{PersianCat}$ in Figure 1c), such that:

$$s_{z_l}(a_m^{z_l}|\mathbf{x}) = \frac{\sum\limits_{n_i \in \text{anc}(z_l)} [[a_m^{z_l} = a_m^{n_i}]] \; s_{n_i}(a_m^{n_i}|\mathbf{x})}{\sum\limits_{n_i \in \text{anc}(z_l)} [[a_m^{z_l} = a_m^{n_i}]]}, \tag{4}$$



(a) (b)

Figure 2: The (a) accuracy and (b) mean AUC of DAP and HAT in CUB with varying number of classes in the source.

where $[[\cdot]]$ is the Iverson bracket, $s_n(a_m^n|\mathbf{x})$ is the score of the attribute a_m for node n given sample $\mathbf{x}$, and $\mathrm{anc}(n)$ is the set of ancestor nodes of n. Once the attributes are transferred to z_l, the final prediction score $s(z_l|\mathbf{x})$ of the z_l category can be defined by averaging over the attributes of that class as:

$$s(z_l|\mathbf{x}) = \frac{\sum_{m=1}^{M} [[a_m^{z_l} = 1]] \; s(a_m^{z_l}|\mathbf{x})}{\sum_{m=1}^{M} [[a_m^{z_l} = 1]]}. \qquad (5)$$

3. Experiments

We evaluate using three datasets: (1) aPascal/aYahoo (aPaY); (2) Animals with Attributes (AwA); (3) CUB-200-2011 Birds (CUB). Each provides different characteristics regarding the granularity of classes. This give us the chance to see how the performance of the proposed HAT model varies with regards to the complexity of the embedded knowledge in the source set. We learn the object hierarchies using the WordNet ontology. As image features[1], we use the output of the $7^t h$ layer of the CNN-M2K deep model from [3]. We train linear SVMs for the attribute classifiers.

Zero-shot Classification In Table 1, we report the normalized multi-class accuracy on the three test sets. Our model outperforms the state-of-the-art on the three datasets with a wide margin. Even when compared to our strong baselines (DAP- and ENS-deep) which use the same features and classifiers, HAT still performs the best. Furthermore, we find that normalizing the prediction scores of the novel classes (Eq. 5) makes the scores more comparable. This improves the accuracy of both the baseline (ENS-n) and our model (HAT-n) with the latter surpassing the former. The improvement in accuracy of HAT relative to the baseline is higher on AwA and CUB (19% and 30%) compared to aPaY (7%). This is expected since the classes in aPaY are visually farther apart from each other compared to the classes in AwA and CUB. Thus, it is harder for the baseline models (DAP & ENS) to distinguish such fine grained objects using the abstract global attributes.

Unknown attribute associations of the novel class Although this evaluation setup is not possible with the global attribute model, HAT enables us to carry out zero-shot recognition even if the attribute description of the novel class is unknown. To do that, we again leverage the hierarchy and transfer the attribute description of the parent node to the novel class. Using this setup, HAT achieves an accuracy of 21.1% (aPaY), 52.6% (AwA) and 25.9% (CUB). This drop in performance is reasonable since we are transferring the more generic attributes of the parent. Hence, confusion can arise when multiple test classes share the same parent in the hierarchy. Nonetheless, HAT makes it possible to perform attribute-based zero-shot classification when only the novel class label is available.

Model	Features	aPaY	AwA	CUB
DAP [5]	shallow	19.1	41.4	-
IAP [5]	shallow	16.9	42.2	-
AHLE [1]	shallow	-	43.5	17.0
HEX [4]	deep	-	44.2	-
DAP	deep	31.9	54.0	33.7
ENS	deep	31.7	57.4	29.0
HAT (ours)	deep	**38.3**	**63.1**	**44.4**
ENS-n	deep	43.1	57.7	37.3
HAT-n (ours)	deep	**46.3**	**68.8**	**48.6**

Table 1: Zero-shot multi-class accuracy.

Source set complexity We use the CUB dataset and start with a random set of 25 classes to be in the source. We gradually increase the source set with additional 25 random classes. At each step, the rest of the 200 classes is used as the target set to conduct zero-shot classification. This helps to have a better understanding of the characteristics of the different models as the richness of the embedded information in the source changes compared to the target. In Figure 2 we see that when the source is relatively poor and contains less structured knowledge, both DAP and HAT performs at the same level. However, as the source get bigger and more complex HAT consistently outperforms DAP with an increasingly wider margin. Unlike DAP that uses a single layer of global attributes, HAT is able to take advantage of the complexity of information available in the source. HAT captures the commonality among the categories and exploits it to learn and transfer more discriminative attributes to distinguish the unseen categories.

References

[1] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid. Label-Embedding for Attribute-Based Classification. In *CVPR*, 2013. 2

[2] Z. Al-Halah and R. Stiefelhagen. How to Transfer? Zero-Shot Object Recognition via Hierarchical Transfer of Semantic Attributes. In *WACV*, 2015. 1

[3] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the Devil in the Details: Delving Deep into Convolutional Nets. In *BMVC*, 2014. 2

[4] J. Deng and et al. Large-Scale Object Classification using Label Relation Graphs. In *ECCV*, 2014. 2

[5] C. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *T-PAMI*, 2013. 2

[1] The deep features used in this work are available on: https://cvhci.anthropomatik.kit.edu/~zalhalah