Inferring Analogous Attributes:
Large-Scale Transfer of Category-Specific Attribute Classifiers

Chao-Yeh Chen and Kristen Grauman

Abstract

The appearance of an attribute can vary considerably from class to class, causing standard class-independent attribute models to break down. Yet, training object-specific models for each attribute is impractical, and defeats the purpose of using attributes to bridge category boundaries. We propose a novel form of transfer learning that addresses this dilemma. Given a sparse set of class-specific attribute classifiers, our tensor factorization approach can infer new ones for object-attribute pairs unobserved during training. We apply our idea to learn over 25,000 analogous attribute classifiers on SUN and ImageNet.

1. Introduction

Attributes are visual properties that describe objects or scenes, such as “fluffy” or “formal”. A major appeal of attributes is the fact that they appear across category boundaries. But are attributes really category-independent? Does fluffiness on a dog look the same as fluffiness on a towel? While the linguistic semantics are preserved across categories, the visual appearance of the property may be transformed to some degree. This suggests that the standard approach—pooling training images from any category and learning a discriminative classifier—will weaken the learned model to account for the “least common denominator” of the attribute’s appearance.

Taking the other extreme, one might attempt to learn category-sensitive attribute classifiers, by gathering positive exemplar images for each category+attribute combination (e.g., separate sets of fluffy dog images, fluffy towel images). However, learning attributes in this manner is quite costly in terms of annotations. In fact, even in the era of Big Vision, the long-tailed distribution of object/scene/attribute occurrences in the real world means that some object-attribute pairs will have inadequate exemplars to build a statistically sound model. Furthermore, naively training each attribute in an object-specific manner would fail to leverage the common semantics of attributes.

1Per the call for papers, we are submitting single-blind because this work appears in the main conference at CVPR 2014.

2We use “category” to refer to either an object or scene class.

Figure 1. Having learned a sparse set of object-specific attribute classifiers, our approach infers analogous attribute classifiers. The inferred models are object-sensitive, despite having no object-specific labeled images of that attribute during training.

To resolve this problem, we propose a novel form of transfer learning to infer analogous category-sensitive attribute models. Intuitively, even though an attribute’s appearance may be specialized for a particular object, there likely are latent variables connecting it to other objects’ manifestations of the property. So, having learned some category-sensitive attributes, we aim to predict how the attribute might look on a new object, even without labeled examples depicting that object with the attribute. See Fig.1

2. Approach: Main Idea

Given training images labeled by their category and attributes, our method produces as output a series of category-sensitive attribute classifiers. Some of those classifiers are explicitly trained with the labeled data, while the rest are “analogous” attributes inferred by our method using no additional training images.

For each attribute for which we do have category-specific labeled examples, we train an importance-weighted support vector machine (SVM). It uses images from all categories, but places a higher penalty on violating attribute label constraints for the specified category. Only a fraction of possible attributes can be explicitly trained in this manner—for example, only ~25% for ImageNet or SUN.

Next we define a tensor to capture the structure underlying many such category-sensitive models. Let $m$ index the
This is where our method comes in. We infer datasets.

For 18K of the 76% of the cases, with average increases of 0.15 in AP. Possible category-sensitive models surpass universal ones in this regard as the “upper bound” category-sensitive results, yet use no category-specific labeled images. Critically, our inferred models are more accurate than the status quo universal approach. We infer models for all missing attributes; whereas the category-sensitive method would require 20 labeled examples per classifier—about 384K additional labeled images—to train those models, our method uses zero.

Table 1 (col 4) shows this key result, with comparisons to standard transfer methods where applicable (cols 5 and 6). Our inferred analogous attributes are nearly as accurate as the “upper bound” category-sensitive results, yet use no category-specific labeled images. Critically, our inferred models are more accurate than the status quo universal approach. We infer models for all missing attributes; whereas the category-sensitive method would require 20 labeled examples per classifier—about 384K additional labeled images—to train those models, our method uses zero.

Fig. 2 illustrates how analogous attributes enable transfer. We take a category \( j \) and identify its neighboring categories in the latent feature space. Then, for each neighbor \( i \), we sort its attribute classifiers \( \mathbf{w}(i,:), \) real or inferred) by their maximal cosine similarity to any of category \( j \)'s attributes \( \mathbf{w}(j,:). \) The resulting shortlist shows which attribute+category pairs our method expects to transfer to category \( j \). We show 4 examples, with one representative image for each category. Neighboring categories in the latent space are often semantically related (e.g., syrup/bottle) or visually similar (e.g., airplane cabin/conference center). Our method receives no explicit side information on semantic distance, yet it discovers such ties via the observed attribute classifiers. Some semantically more distant neighbors (e.g., platypus/rorqual, courtroom/cardroom) are also amenable to transfer.