Learning the right thing with visual attributes*

Kristen Grauman

Abstract

Visual attributes are human-nameable semantic properties. They are the adjectives of the visual recognition world, capturing anything from material properties (”metallic”, ”furry”), shapes (”flat”, ”boxy”), expressions (”smiling”, ”surprised”), to functions (”sittable”, ”drinkable”). An attribute may be a binary predicate (”shiny”) or a relative comparison (”shinier than”). Many promising applications of visual attributes—including zero-shot learning and image search—demand that the vision system model the correct concept precisely. However, existing methods are prone to learning the wrong thing. In particular, the standard discriminative learning pipeline tends to learn correlated properties, fails to account for differences in human perception, and is inadequate to capture fine-grained attribute differences.

I will present our work investigating how to ”learn the right thing” when training attribute models. First, to reduce confusions from correlated attributes, we introduce a novel multi-task learning approach that encourages feature competition among unrelated attributes. Then, turning to the fine-grained attribute problem, we develop lazy local approaches that generate prediction functions on the fly for each novel test case. They make it possible to detect subtle differences between very similar images, an essential capability for sophisticated image search applications. Finally, we question the status quo of learning purely object-independent attributes. Rather than train a single classifier for each attribute, we explore a new form of large-scale transfer that infers ”analogous” class-sensitive attribute models. This allows, for example, predicting what spottedness will look like on a dog, when during training we have only observed spottedness on a disjoint set of objects. We demonstrate our idea to learn over 25,000 object-sensitive attributes for SUN and ImageNet.

*Machine Learning External Seminar, Gatsby Unit, July 29, 2015.