



Words and Pictures

David Forsyth

University of Illinois at Urbana-Champaign, USA

Abstract

Many pictures appear with text nearby. The relations between this text and the pictures are complicated and interesting. There are several reasons to study these relations. We could try and predict word annotations from pictures. A good solution to this problem would make it possible to search for unlabelled pictures using keywords - we just label the pictures automatically. Images that have multiple labels are cues to what objects co-occur and what do not; but they are also cues to what image features are stable and what are not. Linguistic phenomena can give strong information about what it means to do object recognition. For example, sentences written about pictures can reveal spatial relations between objects. As another example, adjectives in captions can help focus search for training data. As yet another example, captions sketch the major interesting phenomena in pictures, with reference to a shared discourse between the viewer and the caption writer which reveals information about the picture. A picture titled "This man died in a fire" would show a man, but a picture titled "A man died in this fire" would show a fire. Information of this form might tell us what is worth recognizing and labelling in pictures.

I will review a large set of methods for reasoning about relations between words and pictures, placing a special emphasis on methods that link linguistic phenomena to image observations.

Syllabus: methods for reasoning about relations between words and pictures.