

Computer Vision based Compositional Zero-shot Learning

Dipti Prasad Mukherjee
Indian Statistical Institute, Kolkata Centre

Abstract

Computer vision based machine learning (CVML) for zero-shot compositions recognizes composite concepts. For example, CVML tries to recognize a rotten banana when the CVML system has seen ripe banana and rotten apple during training. It is zero-shot learning as rotten banana was not seen by the CVML during training. It is learning of compositions as state rotten and object banana describe a new concept unknown during training while individual parts (like, state rotten and object banana) are already seen by the learning model. Naturally, the challenge during training of compositions is to disentangle features of state from features of objects. The better disentanglement of features of state and object during training results in better description of previously unseen compositions during inference. The state features also influence the definition of compositions. For example, in the compositions peeled apple and peeled orange, the state peeled has different visual connotations. This context dependency of state features poses additional challenge to feature disentanglement. Therefore, one of the key research questions in CVML is to measure the quality and extent of disentanglement of state features from features of object. We propose to discuss the CVML architecture and possible disentanglement measures.