CS229 PROJECT PROPOSAL

Zero Shot Learning
Project Category : Computer Vision

Shreyash Pandey (shreyash@stanford.edu)
Abhijeet Phatak (aphatak@stanford.edu)

1 Motivation

There are around 30,000 human-distinguishable basic object classes and many more fine grained ones. A major barrier to progress in visual recognition is thus collecting training data for these many classes. To counter this problem, a technique known as Zero Shot Learning (ZSL) has recently been introduced through which a model is able to detect classes which were not part of the training set. The idea is to design algorithms that simulate how humans identify unseen objects - by drawing information about that object from a different source (like text) and then using that to identify the object.

We plan to implement some state of the art methods that perform ZSL and survey their strengths and weaknesses. While the task at hand might be theoretical, there are numerous applications of ZSL - one very exciting one being assigning trending Instagram tags to an image. We plan to design a web interface for the same.

2 Method

A ZSL model typically utilizes information from text corpora, images and their labels and maps them to a common semantic space. Such a semantic space could either be a word space or an attribute space. Attribute space is defined using attributes(usually binary) such as 'hasFur', 'hasTail', 'isBrown' etc. and are usually not preferred since manually tagging images with such attributes is not scalable and is inefficient. Word space is defined by word2vec operations. General methods involve learning a mapping from the data to project images into the semantic space. Now during test time, the input image is mapped to the semantic space and a nearest neighbour search or some other similarity metric is used to select the closest unseen class.

This process would potentially involve training a convolutional neural network to map images to the semantic space. We plan to use unsupervised techniques (such as PCA, t-SNE) to cluster labels and images in the semantic space to gain further insight into the problem. There are some domain adaptation issues that arise due to disparity in the training data and test data (of unseen classes). We plan to present a survey of such issues and how to overcome them.

3 Intended Experiments

We plan to test our approach on some standard visual recognition datasets such as Imagenet, Places and MSCOCO. This would involve withholding some classes during training, which would later be used for testing our model. Top-1 and Top-5 hit accuracies are reasonable metrics to test our model.

One very cool experiment would be to predict Instagram tags for an input image. This would involve keeping a list of (parsed) trending Instagram tags (which are unseen classes for our model), and then using ZSL to map an image to multiple hashtags. This could potentially be useful for people who want to gain popularity on social media. By the end of this course, we plan to have a web interface where we post the image URL and obtain image tags that were not present in the training set.