Feature Descriptors for Object Detection

David Wu, Hyunghoon Cho

Using the Deformable Part Model with Autoencoded Image Descriptors

Methodology

We consider two different training set sizes: small (20 positive : 200 negative examples) and medium (50 - 500). First, we train a sparse, single-layered autoencoder with sigmoid activations on a set of whitened and standardized training images taken from the PASCAL VOC Challenge. Using this new set of features, we construct a histogram of feature activations analogous to that of HOG. Using this histogram of autoencoded features, we train the full Deformable Part Model. We evaluate the resultant model on a test set taken from PASCAL and compute the corresponding precision-recall curves.

Introduction

Given its performance in recent years' PASCAL Visual Object Classes (VOC) Challenge, the Deformable Part Model (DPM) is widely regarded to be one of the state-of-the-art object detection and localization algorithms. The DPM as described by Felzenszwalb, et. al. uses Histogram of Oriented Gradients (HOG) descriptors as the underlying feature representation for an object. In this paper, we consider using features learned by a single-layered, sparse autoencoder as a substitute for HOG descriptors in the DPM. The rationale for this is that these learned feature descriptors may capture additional details present in the image that are not reflected in a human-designed set of features such as HOG. Using this more descriptive set of features may in turn yield a better object detector. Additionally, while it is true that deep learning algorithms such as the sparse autoencoder alone generally do not perform as well compared to other vision algorithms, it may be possible to integrate these learned feature descriptors into an existing image classification system such as the DPM.

To further evaluate the performance of such an integration, we also consider the effect of factors such as block normalization and colored features on the performance of our autoencoder-backed DPM. Finally, we examine the performance of the autoencoder-backed DPM across several different object classes from the PASCAL VOC Challenge.

Conclusions

- Current results indicate that while autoencoded features do not perform quite as well as HOG features, they still appear to be a viable low-level feature representation for use in object detection systems.
- Factors that improved the average precision of the detector included additional training examples and increasing the number of features (hidden nodes) up to a certain extent.
- Block normalization and incorporating colors did not yield observable improvements in model performance.
- Future considerations would include fine tuning the latent SVM used to train the DPM, using stacked autoencoders to learn more complex feature representations, and optimize runtime of training algorithm to allow for larger training sets.