Introduction

Zero Shot Learning is the ability to detect classes not part of the training procedure.
30,000 human-distinguishable basic object classes - major barrier is thus collecting training data for many classes.
Model human’s ability to identify unseen objects.

General Methodology

An unseen class related to a seen class by representing both in a semantic embedding space.
Test image is projected to the semantic space (using regression or classification), and a similarity measure with each unseen class is used for prediction.
This method is prone to the Projection Domain Shift (PDS) problem which arises because the projection function learnt from source domain is applied to target domain without any adaptation.

ConSE

Semantic embedding of an unseen image is being obtained using a weighted combination of the most likely seen classes.
\[ f(x) = \frac{1}{2} \sum \rho(y_0(x, t), s(y_0(x, t))) \]
Prediction:
\[ y_0(x, l) = \arg \max_{y \in Y} cos(f(x), s(y)) \]

HirseF

Builds upon ConSE to obtain better semantic embeddings by extracting hierarchical structure defined in the WordNet.
Ensures that labels with low/no occurrence in the vocabulary get reliable embedding vectors.
\[ f(x) = \frac{1}{2} \sum \rho_s(y_0(x, t), s(y_0(x, t))) \]
Here \( s(y_0) = \frac{1}{2} \sum \rho_s(y', y) \)

Novelty Detection using Cross Modal Transfer

All above models fail if the test set contains both seen and unseen classes.
Simultaneously operates on both seen and unseen classes during test time using a novelty detection approach.
Outlier detection at test time is based on the property that an image from an unseen class won’t be very close to the existing training images but will be roughly in the same semantic region.

Transductive Multi-View Embedding

Counts PDS issue by aligning different semantic views using CCA to a shared embedding space.
Unlike other methods, exploits information from multiple semantic representations.
A graph is constructed from the projection of each view in the embedding space.
Label prediction is performed using semi-supervised label propagation from the prototypes to the target data points within and across the graphs.

ZSL using Semantic Graph

Models semantic relationships between both seen and unseen classes in the form of a graph.
Avoids PDS by learning an n-way classifier in the visual feature space, and using the embedding space only to compute semantic relatedness between seen and unseen classes.
An absorbing Markov chain process is designed in which unseen classes are treated as absorbing states and for a test image, ZSL classification is achieved by finding the class label with highest absorbing probability.

Unsupervised Domain Adaptation

Identifies visual feature projection as an unsupervised domain adaptation problem and proposes a dictionary learning approach to solve this problem.
Semantic embeddings are obtained as sparse coding coefficients and Dictionary learning is performed separately for source and target domains.
\[ D_s = \min_{D_t} \|X_s - D_h Y_t\|^2 + \lambda \|D_t\|^2 \]
\[ s.t.\|d_t\|^2 \leq 1 \]

Methods

Datasets used is Animals With Attributes(AwA) with the semantic word space created using wikipedia articles.

<table>
<thead>
<tr>
<th>Method</th>
<th>A</th>
<th>W</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConSE</td>
<td>✓</td>
<td>-</td>
<td>35.1</td>
</tr>
<tr>
<td>Semantic Gr.</td>
<td>-</td>
<td>✓</td>
<td>43.1</td>
</tr>
<tr>
<td>Semantic Gr.</td>
<td>-</td>
<td>✓</td>
<td>49.5</td>
</tr>
<tr>
<td>TMV-BLP</td>
<td>✓</td>
<td>✓</td>
<td>47.1</td>
</tr>
<tr>
<td>UDA</td>
<td>✓</td>
<td>-</td>
<td>47.5</td>
</tr>
<tr>
<td>UDA</td>
<td>✓</td>
<td>✓</td>
<td>49.7</td>
</tr>
</tbody>
</table>

Table 1: Summary of Results

Dataset used is ImageNet 2011 with semantic space for ConSE created using GloVe vectors while for HierSE using flickr tags.

<table>
<thead>
<tr>
<th>Method</th>
<th>hit@1</th>
<th>hit@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConSE</td>
<td>14.5</td>
<td>31.5</td>
</tr>
<tr>
<td>HierSE</td>
<td>17.8</td>
<td>50.9</td>
</tr>
</tbody>
</table>

Table 2: Comparison between ConSE and HierSE

References