Exploring the Limits of Zero-Shot Learning - How Low Can You Go?

Hemanth Dandu[1] Karan Sharma[2] Suchendra M.Bhandarkar[1,2]
Institute for Artificial Intelligence[1] Department of Computer Science[2]
University of Georgia, Athens, Georgia 30602, USA
hemanthreg@gmail.com karan1234@gmail.com suchi@uga.edu

Introduction

Many problem domains are faced with large and growing number of categories, making it difficult to collect and annotate training images for each object category. Zero-Shot Learning (ZSL) frameworks enable the learning of classifiers when no training data is provided. While most ZSL frameworks aim to achieve maximum classification accuracy on standard ***seen*** vs. ***unseen*** splits of data sets, we propose a framework that enables inference of a **larger number of *unseen* categories** using **very few *seen* categories**. In particular, we examine the functional dependence of the classification accuracy of unseen object classes on the number of previously seen classes.

Data Sets

We test our framework on three widely used ZSL data sets namely ***Animals with Attributes-2*** **(AWA2)**, ***Caltech-USCD Birds*** **(CUB)**, and the **Scene Understanding with Attributes (SUN)** database

Table 1: Summary statistics for three data sets

Data set	Detail	Classes	Images	Attributes
AWA2	coarse	50	37,322	85
CUB	fine	200	11,788	312
SUN	fine	717	14,340	102

Figure 1: Sample Images from the AWA2 data set

Proposed Framework

- **Deep Feature Extraction:** Use ***ResNet 101*** pre-trained on ***ImageNet*** to extract deep features from each image
- **Auxiliary Information:** Establish semantic relationships between ***seen*** and ***unseen*** classes. ***Principal Component Analysis*** is used to reduce the auxiliary space to a compact combined semantic space. We use 3 sources of auxiliary information:
  - **Attributes:** All data sets contain attribute information for each of their object categories; e.g. AWA2 data set contains attribute information such as *black, small, walks, smart* etc.
  - **Text Embeddings:** Learned vector representations of object category labels can help to construct semantic relationships between seen and unseen category names. We extract ***FastText*** word representation for each category label.
  - **Hierarchy Embedding:** Creating a hierarchy of categories present in a data set allows us to derive taxonomy-based relationships between the classes to improve ZSL performance.
- **Clustering:** Identify object categories that are good representatives for a large number of similar object categories. We use two clustering techniques, ***Gaussian Mixture Model*** **(GMM)** and ***Affinity Propagation*** **(AF)**
- **Multi-Label Classification:** Training set is filtered for class labels associated with the cluster centers. ***Random Forest*** **(RF)** classifier is trained to distinguish between each cluster center.
- **Predictions & Alternate Hypothesis:** Each test instance is classified as one of the representative cluster centers. Alternative hypotheses are then generated using a similarity measure within the combined semantic space. We use ***cosine similarity*** as the similarity measure.

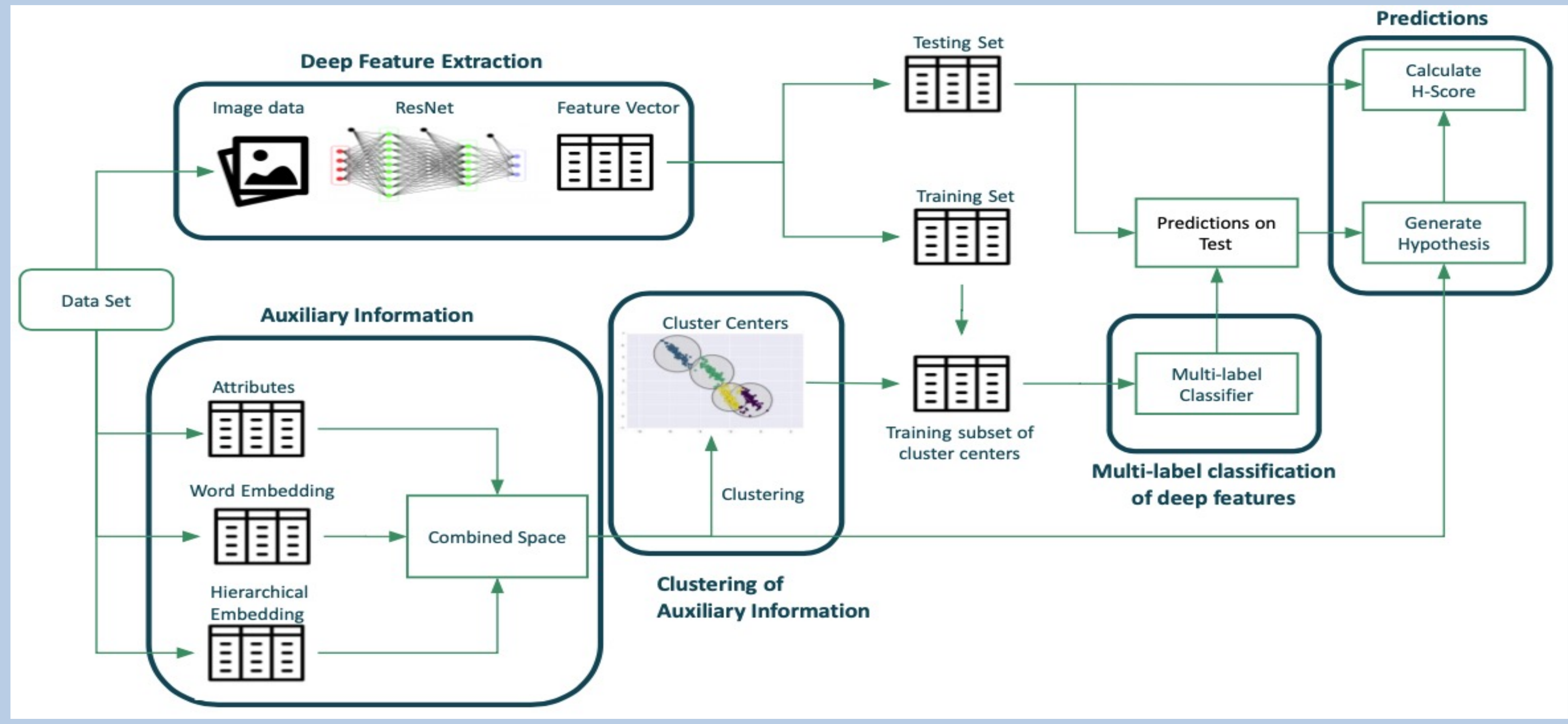

Figure 2: Overview of the proposed framework

Results & Comparison

- H-Score (HS) is used as the evaluation metric which is the harmonic mean of the ***seen class accuracy*** (*SCA*) and ***unseen class accuracy*** (*UCA*).

$$HS = \frac{2 \times (SCA \times UCA)}{(SCA + UCA)}$$

Equation 1: Harmonic mean between SCA and UCA

- We compare our results with ***Attribute Label Embedding*** (***ALE***) in our experiment set up.

Table 2: Comparison of average classification accuracy across *k* values between the proposed model and ALE when using GMM-based clustering.

	Proposed Model			ALE		
Data Set	Avg. Seen	Avg. Unseen	Avg. H-Score	Avg. Seen	Avg. Unseen	Avg. H-Score
AWA2	94%	32%	**45%**	90%	14%	24%
CUB	78%	22.86%	**33%**	70%	17.50%	27.19%
SUN	58.20%	15%	21.60%	41.50%	17.80%	**25%**

Table 3: Comparison of average classification accuracy between the proposed model and ALE when using AP-based clustering.

	Proposed Model			ALE		
Data Set	Seen	Unseen	H-Score	Seen	Unseen	H-Score
AWA2	96.44%	23.43%	**37.7%**	83.40%	10%	17.50%
CUB	91%	9.70%	**17.50%**	55%	8.33%	14.40%
SUN	83.10%	4.30%	8.20%	24.40%	8.20%	**12.30%**

Conclusion:

- We introduce a framework for ***generalized Zero-Shot Learning*** (**GZSL**) that is simple yet effective when *unseen* classes significantly outnumber *seen* classes
- The proposed framework achieves 21% higher accuracy on the AWA2 data set and 6% higher accuracy on the CUB data set when compared to the ALE scheme in the context of GZSL
- Current drawback of the proposed framework is its inability to infer unseen classes that are very distant from the representative classes in the semantic space which presents scope for future improvement