Predicting Visual Exemplars of Unseen Classes For Zero-Shot Learning

Soravit Changpinyo, Wei-Lun Chao, and Fei Sha
University of Southern California

ICCV'17

Highlights

- ***Simple and efficient*** zero-shot learning (ZSL) methods for image classification that are ***highly effective.***
- ***Predicting*** **visual "exemplars"** from semantic representations via kernel-based regression.
- ***Explicitly*** **taking advantage of clustering structures** in the semantic embedding space.
- ***Improving*** upon several recent state-of-the-art ZSL approaches.

Zero-Shot Learning

Classes in training & testing phrases are **disjoint.**

Training
Seen Classes

horse bear tiger

***Semantic descriptions* for "all" classes**

- **Visual attributes** (e.g., striped, black, four-legged)
- **Word vectors** of class names
- **WordNet hierarchy**
- …

Zero-shot learning model

Testing
Unseen Classes

zebra or panda or cat ?

Approach

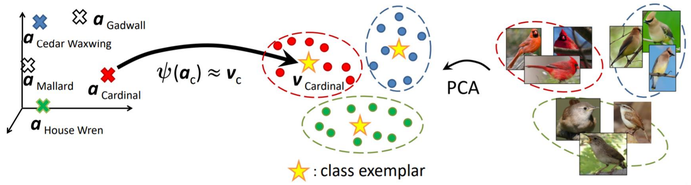
Main Idea

- Assume the **existence of clustering structures** in the semantic embedding space.
- [Step 1] Represent each cluster with a visual exemplar & **learn to predict exemplars** from their corresponding semantic representations.
- [Step 2] Predict exemplars of "unseen" classes for **ZSL**.

Step 1: Predicting visual exemplars

Notations

- a_c := the **semantic representation** of class c
- v_c := the **target visual exemplar** of class c
- ψ := a mapping from a_c ***to*** v_c (to be learned)



In this work

- v_c is the PCA projection of the average of visual features of class c.
- ψ is a support vector regressor with RBF kernel.

Step 2: ZSL with Exemplars

What can we do with predicted visual exemplars of "unseen" classes?

$v_u = \psi(a_u)$ for each unseen class u

[EXEM only]
1-nearest neighbor classification in the exemplar space

[EXEM + ZSL algorithm]
Use predicted exemplars as **"ideal" semantic representations** in ***any*** ZSL algorithm.

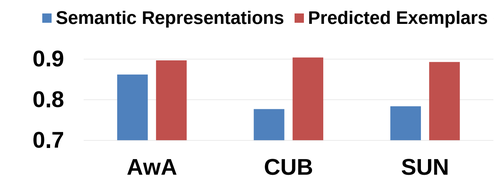
Experiments

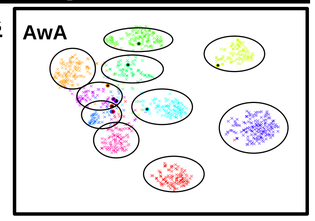
Datasets

	AwA (animals)	CUB (birds)	SUN (scenes)	ImageNet
# of seen classes	40	150	645/646	1,000
# of unseen classes	10	50	72/71	20,842
Total # of images	30,475	11,788	14,340	14,197,122
Semantic representations	Attributes	Attributes	Attributes	Word vectors (wv) WordNet hierarchy (hie)

Quality of Predicted Exemplars

Correlation Coeff. To "Real" Exemplar Distances





Zero-Shot Classification Accuracy

	Method	Per-class Accuracy AwA	CUB	SUN	Flat Hit@5 ImageNet (wv)	ImageNet (hie)
Baselines	**ConSE** [Norouzi et al. 14]	63.3	36.2	51.9	3.8	
	BiDiLEL [Wang & Chen 17]	72.4	49.7			
	LatEm [Xian et al. 16]	72.1	48.0	64.5		
	CCA [Lu 16]					5.2
	SynC$_{ovo}$ [Changpinyo et al. 16]	69.7	53.4	62.8	4.5	6.0
	SynC$_{struct}$ [Changpinyo et al. 16]	72.9	54.5	62.7	4.4	
EXEM + ZSL algo	**EXEM (ConSE)**	70.5	46.2	60.0		
	EXEM (LatEm)	72.9	56.2	67.4		
	EXEM (SynC$_{ovo}$)	73.8	56.2	66.5	5.0	6.1
	EXEM (SynC$_{struct}$)	**77.2**	**59.8**	66.1		
EXEM only	**EXEM (1NN)**	**76.2**	56.3	**69.6**	5.2	**6.3**
	EXEM (1NNs)	**76.5**	**58.5**	67.3	**5.3**	6.2

More Results & Analysis in the paper

- ZSL results when **"peeking" into labeled data of some unseen classes** (from zero-shot to few-shot learning)
- Improved results on **Generalized Zero-Shot Learning** [Chao et al. ECCV16]
- Analysis on
  - **PCA**
  - **Support vector regression vs. Multi-layer perceptron** (for predicting visual exemplars)