# Zero-Shot Classification with Discriminative Semantic Representation Learning


Meng Ye[1] and Yuhong Guo[2]

[1]Temple University, [2]Carleton University

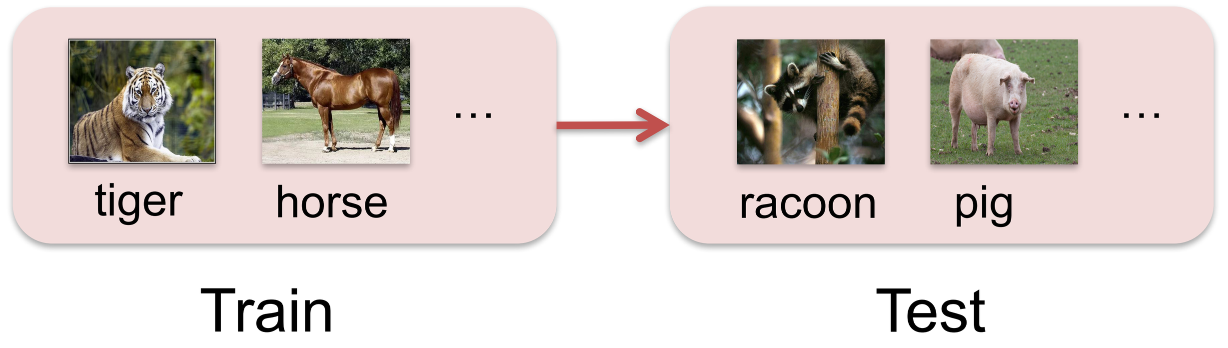

College of Science and Technology TEMPLE UNIVERSITY

Carleton UNIVERSITY

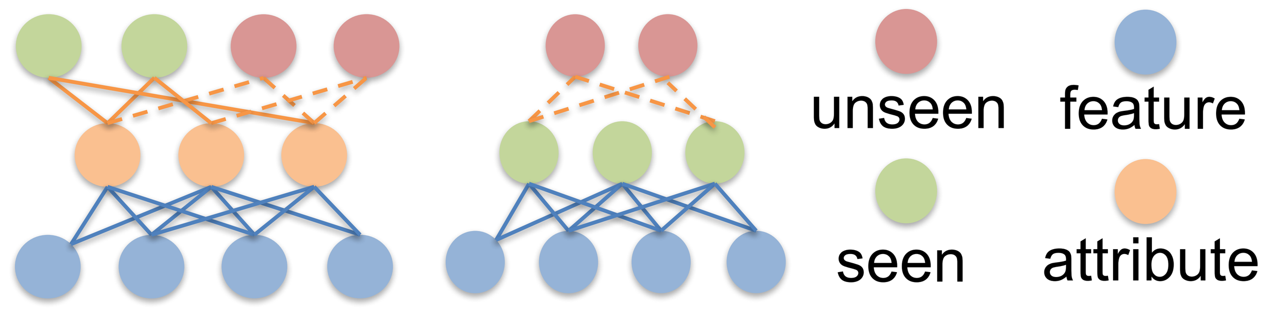
IEEE 2017 Conference on Computer Vision and Pattern Recognition

## Introduction

- Zero-shot learning
  - Seen and unseen classes are disjoint

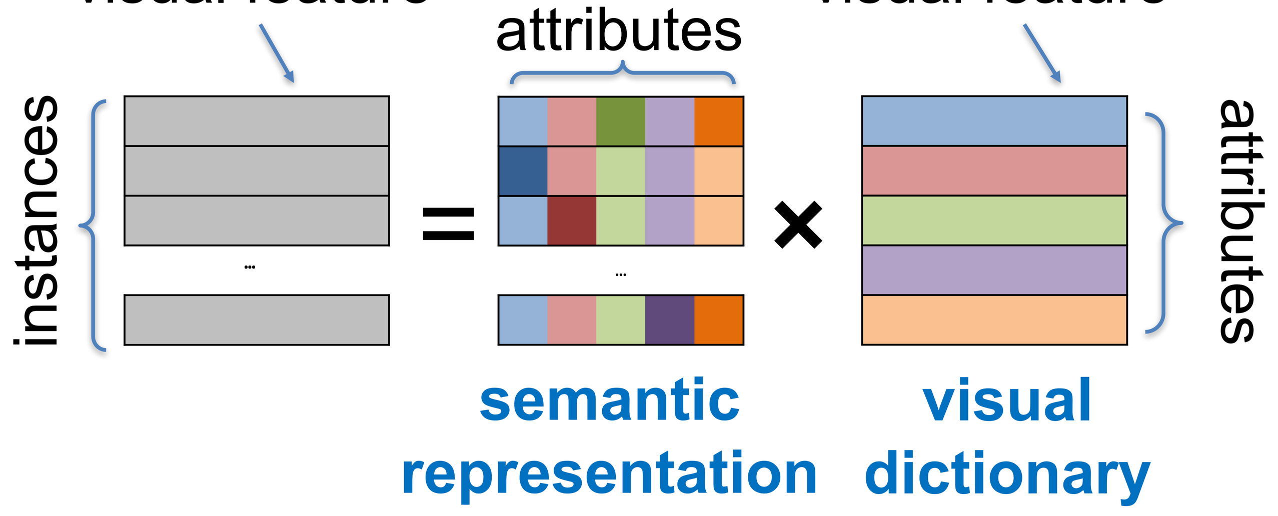



- Semantic Information
  - Attributes, e.g. tiger(stripes, orange, …)
  - An intermediate layer between visual feature and class labels
  - shared by seen/unseen classes.

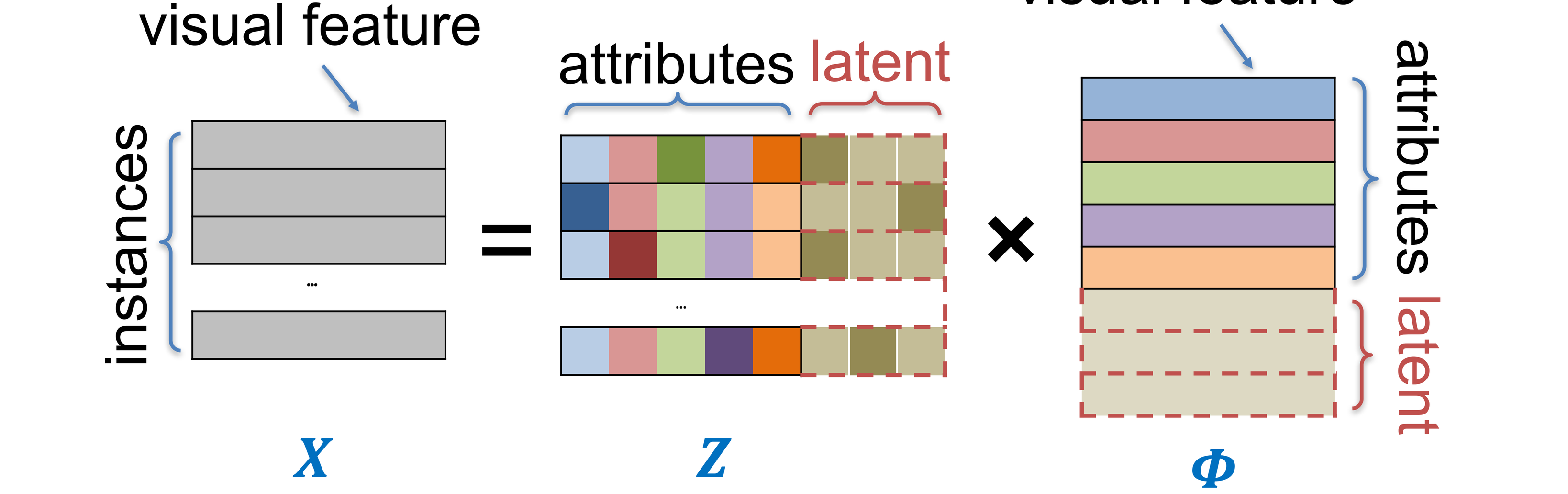



## Motivation

- Non-negative matrix factorization




- Latent components




## Methodology

- Overall objective

NMF reconstruction error — alignment hinge loss — sparsity constraints

$$\min_{Z\geq 0,\,\Phi\geq 0} \frac{1}{2}\|X - Z\Phi\|_F^2 + \gamma\sum_{i=1}^{n_s}\xi_i + \mu\|\Phi\|_1 + \rho\|Z\|_1$$

- $Z$: semantics of both seen/unseen data
- $\Phi$: transferable visual representations
- $\xi_i$: discriminative ability of semantic vectors

$$\xi_i = \max_{k\in\mathcal{Y}^s}(\Delta(Y_i\mathbf{1}_k = 0) - \|Z_iB - M_k\|^2 + \|Z_iB - Y_i^sM\|^2)_+$$

- $B$: select the first $m$ columns（attributes）of $Z$
- **latent components**: handle background noise or contents not covered by pre-defined attributes
- Sparsity: better part-based representations.
- Learning: Iterative alternating optimization

- Assign class labels
  - $Z^uB$ is used to determine class labels:

$$Y_i^u = \mathbf{1}_{k^*}^T, \quad \text{where} \quad k^* = \min_{k\in\mathcal{Y}^u}\|Z_i^uB - M_k\|^2$$

semantic representation — class prototype

- Prediction with Label propagation
  - Combine visual feature and semantic representation to compute similarities:

$S = [\hat{X}, \widehat{Z^uB}]$ — Symmetric k-NN graph

$$W_{ij} = \begin{cases} \exp\left(-\frac{d(S_i, S_j)}{2\sigma^2}\right), & i \in KNN(j) \text{ or } j \in KNN(i) \\ 0, & \text{otherwise} \end{cases}$$

  - This k-NN graph is then used to perform standard regularized label propagation:

$$Y^* = (I_{n^u} - \alpha L)^{-1} \times Y$$

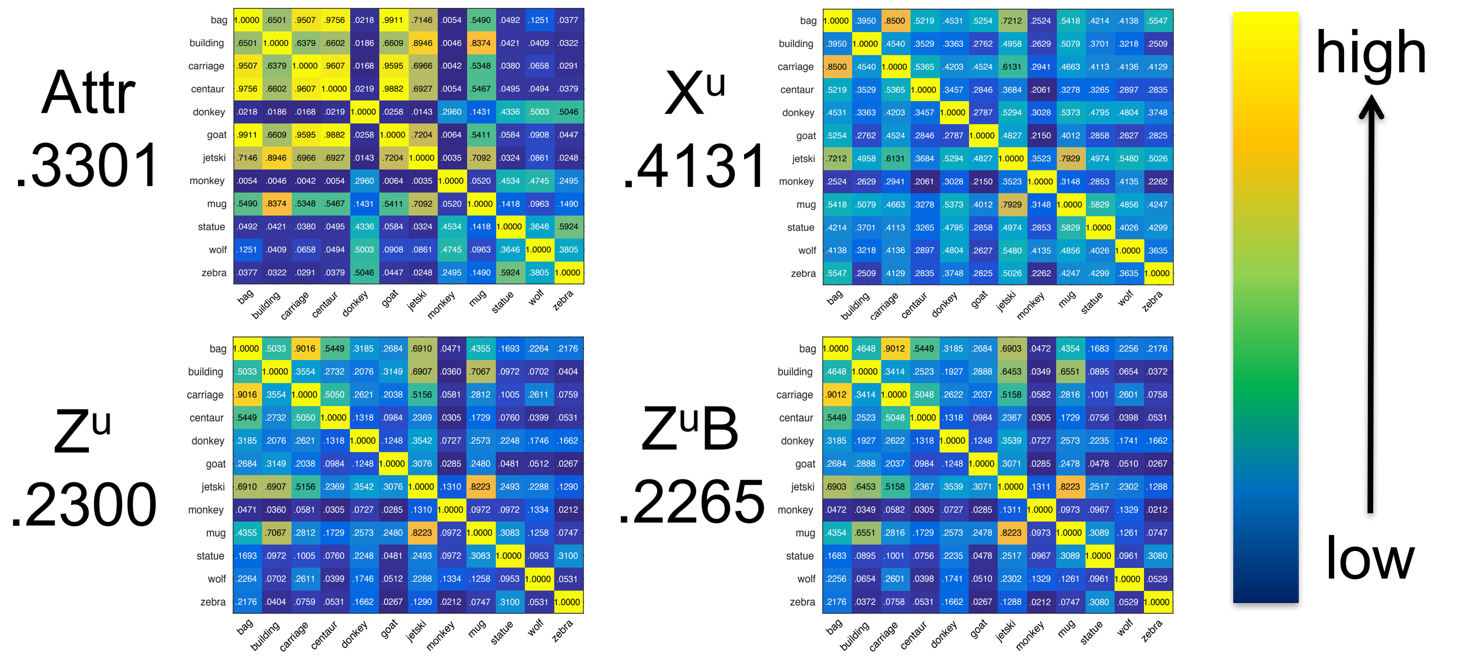
Normalized Lapalacian matrix: $L = Q^{-1/2}WQ^{1/2}$

## Experiments

- Zero-shot test results (multi-class accuracy)

| Method | aPY | AwA | CUB | SUN | Avg. |
|---|---|---|---|---|---|
| [16] | 21.14 | 49.16 | 25.43 | 48.50 | 36.06 |
| [14] | - | 73.2 | 39.5 | - | - |
| [34] | 46.23 | 76.33 | 30.41 | 82.50 | 58.87 |
| [35] | 50.35 | 79.12 | 41.78 | 83.83 | 63.77 |
| DSRL+LP | 51.29 | 87.22 | 57.14 | 85.40 | 70.26 |

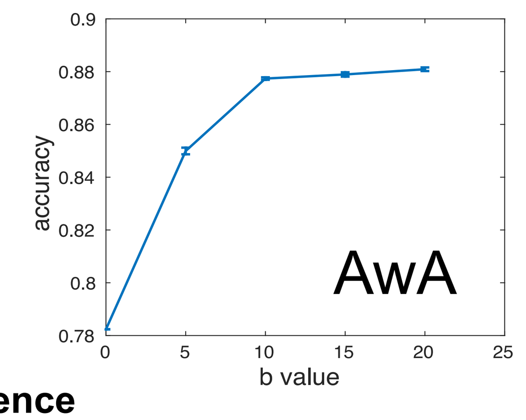
- Inter-unseen class similarities with learned instance representation(aPY)

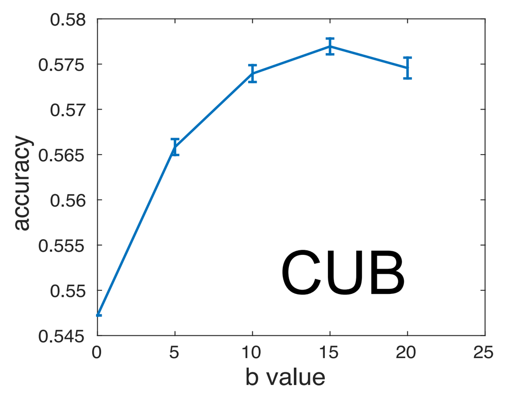



- Similar pairs w.r.t. visual dictionary

| | |
|---|---|
| aPY | ⟨face, nose⟩, ⟨face, hair⟩<br>⟨pot, leaf⟩, ⟨Jet Engine, Propeller⟩ |
| AwA | ⟨ocean, plankton⟩, ⟨black, white⟩ |
| CUB | ⟨wing_color_blue, upperpart_color_blue⟩<br>⟨wing_color_rufous, upperpart_color_rufous⟩<br>⟨underpart_color_rufous, breast_color_rufous⟩<br>⟨underpart_color_green, breast_color_green⟩ |
| SUN | ⟨open_area, natural_light⟩, ⟨sailing, diving⟩<br>⟨hiking, rugged_scene⟩, ⟨diving, scary⟩ |

- Impact of latent components





**Reference**

[14] E. Kodirov *et al*. Unsupervised domain adaptation for zero-shot learning. In *ICCV*, 2015

[16] C. H. Lampert *et al*. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009

[34] Z. Zhang *et al*. Zero-shot learning via semantic similarity embedding. In *ICCV*, 2015.

[35] Z. Zhang *et al*. Zero-shot learning via joint la- tent similarity embedding. In *CVPR*, 2016