





## **Overview**

### Goal

Classify samples belonging to unseen categories.

### Motivation

Proper choice of embedding space is essential.

<b>Direction of Mapping</b>	Problem
Visual to Semantic	Hubness
Semantic to Visual	Semantic structure not preserved!

#### Contribution

We propose to devise objective functions which aim to learn a similarity metric and preserve semantic structure in the visual embedding space.

## **Semantic Relations**

Given class embeddings ym and yn, semantic relation between them is defined as:

- Identical if  $s(\mathbf{y_m}, \mathbf{y_n}) = 1$
- **Semantically similar** if  $\tau \leq s(\mathbf{y_m}, \mathbf{y_n}) < 1$
- Semantically dissimilar if  $s(\mathbf{y_m}, \mathbf{y_n}) < \tau$

#### **Preserving Semantic Relations for Zero-Shot Learning** Yashas Annadani Soma Biswas Indian Institute of Science NITK somabiswas@iisc.ac.in

yashas.annadani@gmail.com

## **Technical Details**



- Consider a quadruplet  $(\mathbf{y_r}, \mathbf{x_i}, \mathbf{x_j}, \mathbf{x_k})$  wherein  $\mathbf{y_r}$  is a reference embedding.
- $(x_i, x_j, x_k)$  are such that corresponding  $(y_i, y_j, y_k)$  obey relations in the definition wrt  $\mathbf{y_r}$ .
- Let  $f(\mathbf{y}; \theta_f)$  be the mapping function to be learnt by the encoder. The objective functions to preserve each relation are as follows:

#### **Objective for identical and dissimilar classes**

 $\min_{\theta_f} -s \big( f(\mathbf{y}_{\mathbf{r}}; \theta_f), \, \mathbf{x}_{\mathbf{i}} \big) + \big( \tau - \delta_{kr} \big) \cdot s \big( f(\mathbf{y}_{\mathbf{r}}; \theta_f), \, \mathbf{x}_{\mathbf{k}} \big)$ 

#### **Objective for similar classes**

$$\min_{\boldsymbol{\theta}_{f}} \left[ \boldsymbol{\tau} - s \left( f(\mathbf{y}_{\mathbf{r}}; \boldsymbol{\theta}_{f}), \mathbf{x}_{\mathbf{j}} \right) \right]_{+} + \left[ s \left( f(\mathbf{y}_{\mathbf{r}}; \boldsymbol{\theta}_{f}), \mathbf{x}_{\mathbf{j}} \right) \right]_{+} \right] \\ \left[ z \right]_{+} = \max(0, z)$$

#### **Reconstruction Objective**

$$\min_{\theta_f, \theta_g} ||\mathbf{y_r} - \mathbf{\hat{y}_r}||_2^2$$

## Results

Reconstructed class

Reconstruction

 $(\mathbf{x_j}) - \delta_{jr}|_+$ 





#### **Results on Imagenet**

		Conventional Zero-Shot Learning			Generaliz Zero-Sho Learning	
		2H	<b>3H</b>	All	2H	<b>3H</b>
Тор - 1	SYNC	9.1	2.6	0.9	0.3	0.1
	Proposed	9.4	2.8	1.0	1.2	0.8
Тор - 5	SYNC	25.9	4.9	2.5	8.7	3.8
	Proposed	26.3	4.8	2.7	11.2	4.9

#### **Approximate Semantic Inference**

classes without class embeddings like Word2Vec.







#### **Embedding Space Visualization**



# Cosine similarity scores as output by the network to samples belonging to