Improving Semantic Embedding Consistency by Metric Learning for Zero-Shot Classification

Maxime Bucher[1,2], Stéphane Herbin[1], Frédéric Jurie[2]

ONERA - The French Aerospace Lab[1], GREYC[2]

Motivation

- **Image classification** goal is to predict the class of an object present in image.
  Traditional classification pipeline design → train a classifier from annotated image database.
  However, in many applications having access to image data is often difficult.
- **Zero-Shot learning** : recognize object never seen during training using **attribute descriptions** as an intermediate knowledge based semantic image representation.
  → **Problem :** Attribute **embedding space** can be **redundant and noisy**.

- **Our approach :** use a **Metric Learning** framework with dimension reduction and space transformation
  → **improve semantic embedding.**

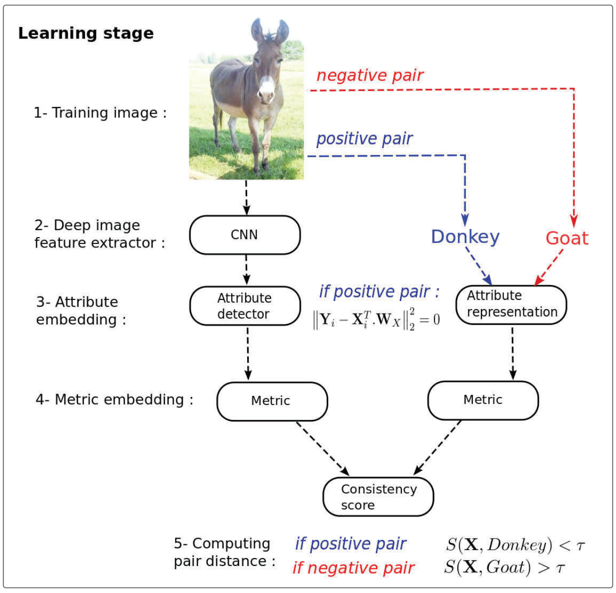
Main idea

- **Attribute embedding :** intermediate representation level, **understandable by human** designers and sufficiently formal to be the support of algorithmic inferences.
- **Attribute representation weaknesses** : may not be the ideal embedding space, can be too redundant and noisy to support reliable inferences.

Our goal → improve semantic embedding by learning a **consistency score** $S(\mathbf{X},\mathbf{Y})$ between image $\mathbf{X}$ and attribute vector $\mathbf{Y}$.

Our idea → optimize jointly the attribute embedding and the classification metric, in a multi-objective framework :

- **Attribute detection :** improve image embedding capacity
- **Metric Learning** : optimization based on asymmetric positive/negative pairs of attributes/images, acts as a **discriminating space transformation and dimension reduction.**



Consistency score :

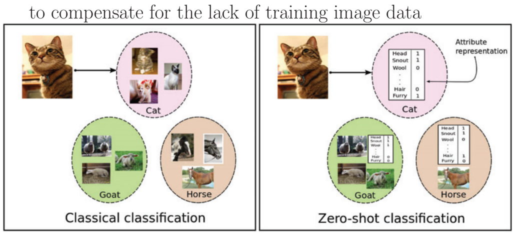
$$S(\mathbf{X},\mathbf{Y}) = \left\|(\mathbf{X}_i^T.\mathbf{W}_X - \mathbf{Y}_i)^T.\mathbf{W}_A\right\|_2$$

Optimization criterion :

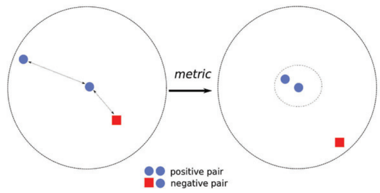
$$\mathcal{L}(\mathbf{W}_A,\mathbf{W}_X) = \sum_i \max\left(0, 1 - Z_i(\tau - \left\|(\mathbf{X}_i^T.\mathbf{W}_X - \mathbf{Y}_i)^T.\mathbf{W}_A\right\|_2^2\right) + \lambda \sum_i \left\|\mathbf{Y}_i - \mathbf{X}_i^T.\mathbf{W}_X\right\|_2^2 + \texttt{Regularization}$$

Classical VS Zero/Few-shot Learning

- **Classical classification :** class inference decision based on training images
- **Zero/Few-shot learning :** use an attribute representation to compensate for the lack of training image data

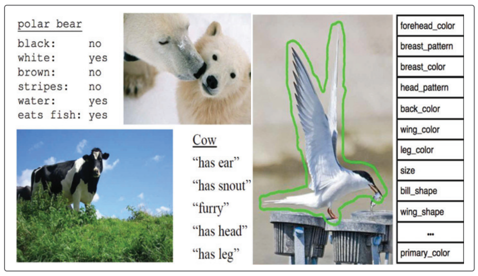


Metric Learning

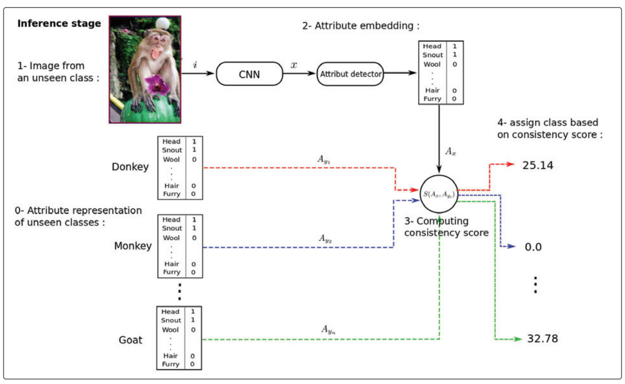


Basic idea : learn a **distance function** that assigns **small** (resp. large) **distance** to pairs of examples that are **semantically similar** (resp. dissimilar).

Image - attribute samples



Experiments



At test time the learned model can predict the consistency of a test image with a given set of attributes :

- every unseen class is described by an attribute vector
- the consistency score between the tested image and all class vectors is computed
- the label with maximal score is the predicted class

State of the art performances in terms of accuracy on 4 Zero-shot datasets.

Feat.	Method	aP&Y	AwA	CUB	SUN
VGG-VeryDeep [3]	Lampert *et al.* [1]	38.16	57.23	-	72.00
	Romera-Paredes *et al.* [2]	24.22±2.89	75.32±2.28	-	82.10±0.32
	Zhang *et al.* [5]	46.23±0.53	76.33±0.83	30.41±0.20	82.50±1.32
	Zhang *et al.* [4]	50.35±2.97	**80.46±0.53**	42.11±0.55	83.83±0.29
	Ours w/o ML	47.25±0.48	73.81±0.13	33.87±0.98	74.91±0.12
	Ours w/o constraint	48.47±1.24	75.69±0.56	38.35±0.49	79.21±0.87
	Ours	**53.15±0.88**	77.32±1.03	**43.29±0.38**	**84.41±0.71**

→ **Few-shot Learning and image retrieval** also possible based on the metric learning based consistency score (see article).

Références

[1] Christoph H Lampert, Hannes Nickisch et Stefan Harmeling. "Attribute-Based Classification for Zero-Shot Visual Object Categorization." In : *IEEE Trans. on Pattern Analysis and Machine Intelligence* 36.3 (2014), p. 453–465.

[2] Bernardino Romera-Paredes et Philip HS Torr. "An embarrassingly simple approach to zero-shot learning". In : *Proceedings of the International Conference on Machine learning*. 2015, p. 2152–2161.

[3] Karen Simonyan et Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In : *ICLR*. 2014.

[4] Ziming Zhang et Venkatesh Saligrama. "Zero-Shot Learning via Joint Latent Similarity Embedding". In : *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, p. 6034–6042.

[5] Ziming Zhang et Venkatesh Saligrama. "Zero-Shot Learning via Semantic Similarity Embedding." In : *IEEE International Conference on Computer Vision (ICCV)*. 2015.