Semantically Consistent Regularization for Zero-Shot Recognition

Pedro Morgado Nuno Vasconcelos

Statistical Visual Computing Lab, Dept. of Electrical and Computer Engineering, University of California, San Diego.

IEEE 2017 Conference on Computer Vision and Pattern Recognition

Introduction

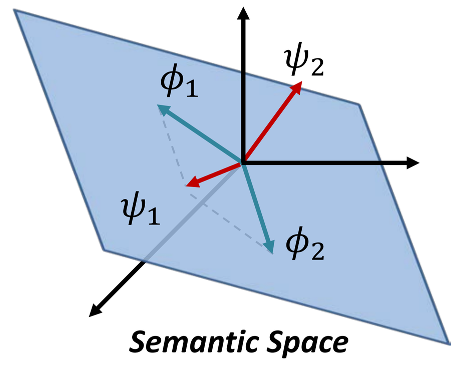
Zero-shot learning: How to classify previously unseen classes?

Our solution: Semantically Consistent Regularizer (SCoRe)

- End-to-end CNN framework.
- Semantics seen as constraints for recognition.
- Extension to multi-state semantics (i.e. non-binary).
- Combines semantic and classification supervision.
- Regularized semantic-to-class mappings for consistency with prior knowledge.

Motivation:

- Recognition by Independent Semantics (RIS)
  - e.g. DAP [Lampert, 2009]
  - 1st order constrains (supervision on single semantics)
  - Cannot learn dependencies between attributes
- Recognition using Semantic Embeddings (RULE)
  - e.g. ALE [Akata, 2013]
  - 2nd order constrains (supervision on linear combinations/class scores)
  - Only supervises the subspace spanned by training codewords
- RIS and RULE introduce complementary learning methodologies

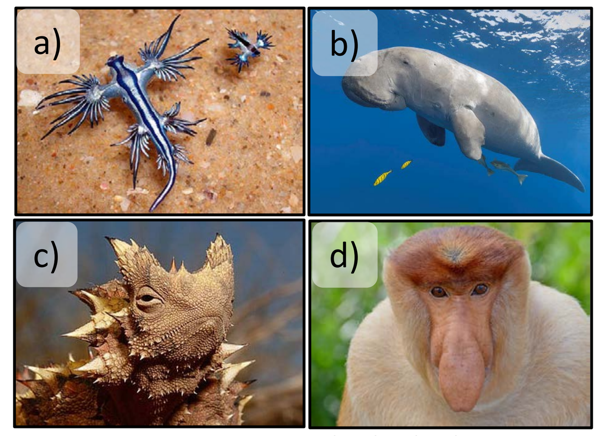


ϕ_i - Training codewords
ψ_i - Zero-shot codewords
Supervised subspace

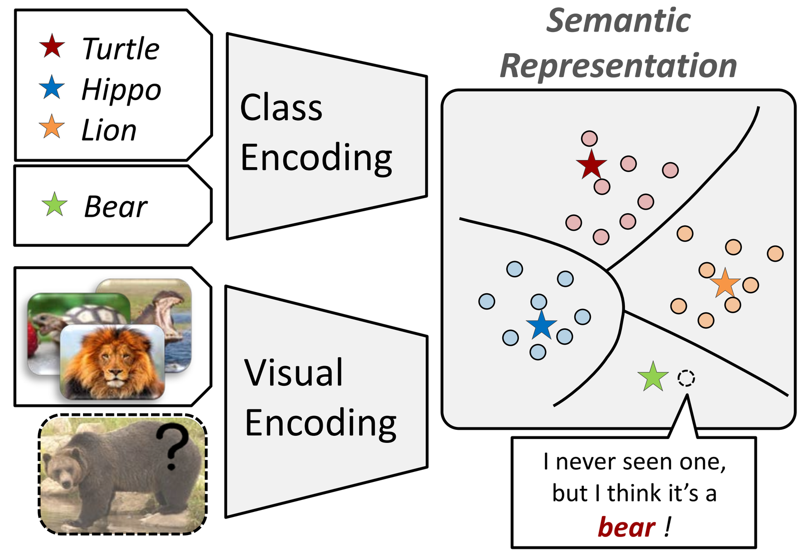
Zero-Shot Learning

Can you name the beast?

1) Thorny devil 3) Proboscis monkey
2) Sea swallow 4) Dugong



Zero-shot learning through semantics.



Why Zero-Shot Learning?

Labeling data is expensive.

Richer image descriptions.
- Object affordances for robotics.
- Attributes for image retrieval.

Humans can do it.

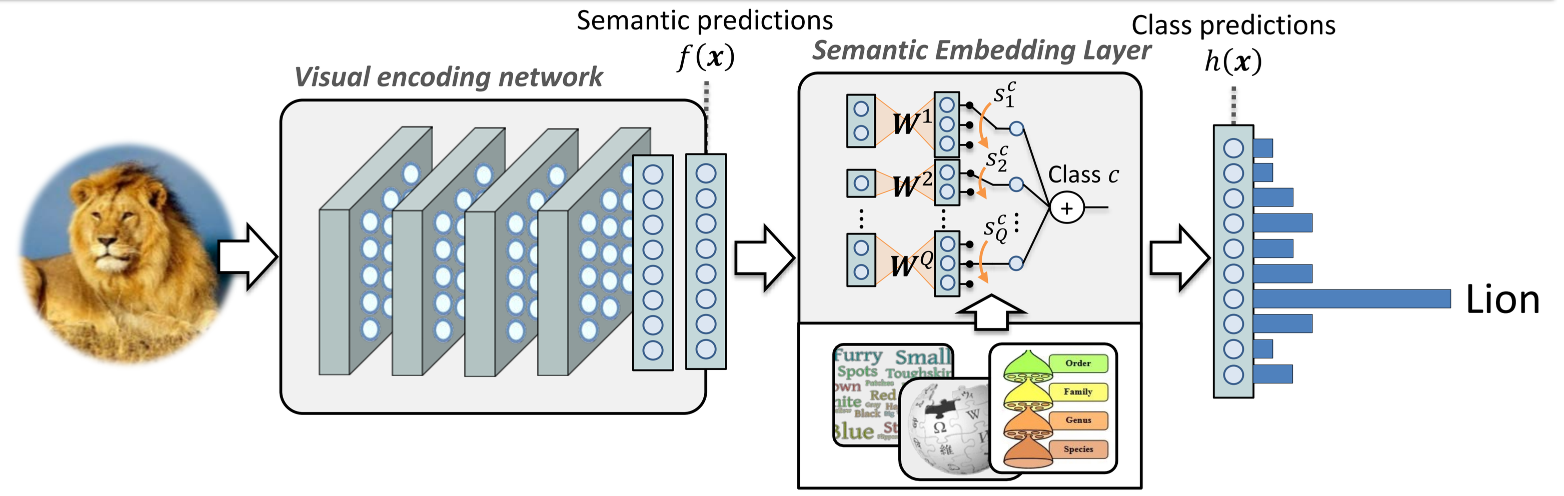
SCoRe – Semantically Consistent Regularizer

Input

Seen classes (source domain):
- Image $\boldsymbol{x}$; semantic description $\boldsymbol{s}$; class y.
- Training data: $\mathcal{D}_\mathcal{Y} = \left\{(\boldsymbol{x}^i, \boldsymbol{s}^i, y^i)_{i=1}^n\right\}$

Unseen classes (target domain):
- Image $\boldsymbol{x}$.
- Goal: predict semantics $\boldsymbol{s}$ and class y.

Model



Loss

$$\underset{\Theta,W}{\text{minimize}} \sum_i \Big(\underbrace{L_c(W^T f(\boldsymbol{x}^i;\Theta), y^i)}_{\text{Classification loss}} + \gamma \sum_k \underbrace{L_s(f(\boldsymbol{x}^i;\Theta), s_k^i)}_{\text{Loss-based regularizer}}\Big) + \beta \sum_y \underbrace{\Omega(\boldsymbol{w}_y, \phi(y))}_{\text{Codeword regularizer}}$$

Model components

Visual encoding network: $f(\boldsymbol{x};\Theta)$	Popular CNNs + FC semantic prediction layer.
Semantic embbeding: $h(\boldsymbol{x})$	$\sum_k \langle \boldsymbol{w}^k_{s_k^c}, f_k(\boldsymbol{x};\Theta)\rangle$
Classification loss: L_c	$-\log \frac{e^{h_y(\boldsymbol{x})}}{\sum_k e^{h_k(\boldsymbol{x})}}$
Loss-based regularizer: L_s	$-\log \sigma(f_k(\boldsymbol{x}))^{s_k}\left(1-\sigma(f_k(\boldsymbol{x}))\right)^{1-s_k}$
Codeword regularizer: Ω	$\sum_k \left\Vert \boldsymbol{w}^k_{s_k^y} - \psi^k(y)\right\Vert^2$
Semantic codewords: ψ^k	Known semantic/class relations

Extended annotations

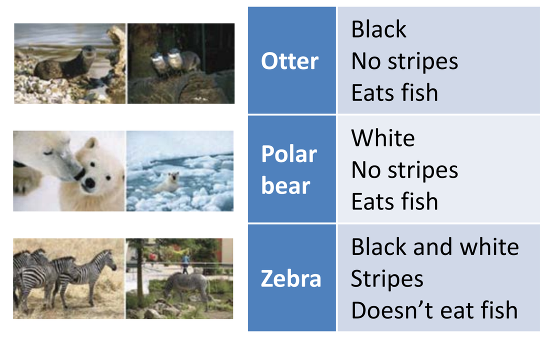
Semantics $s_k^* = \arg\max_s \langle \boldsymbol{w}^k_{s}, f_k(\boldsymbol{x})\rangle$

Classes $y^* = \arg\max_y \sum_k \langle \boldsymbol{w}^k_{s_k^y}, f_k(\boldsymbol{x})\rangle$

Baselines	RIS	RULE
Loss-based regularizer	$\gamma \to \infty$	$\gamma = 0$
Codeword regularizer	$\beta \to \infty$	$\beta \to \infty$

Semantics

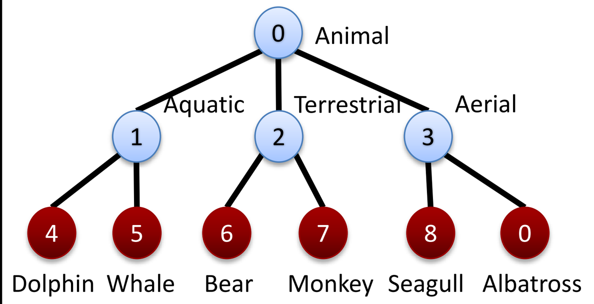
Attributes



- Human annotated
- Binary semantics

$$\psi^k(y) = \begin{cases} 1 & \text{if attribute } k \text{ in class } y \\ -1 & \text{otherwise} \end{cases}$$

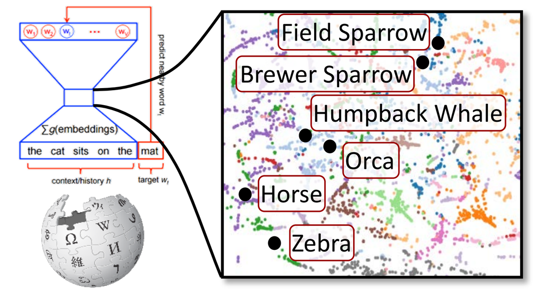
Taxonomy



- Each node treated as a small classification problem between child nodes plus reject option.
- Taxonomy: WordNet

$\psi^k(y)$: Maximally separated codewords

Word2Vec

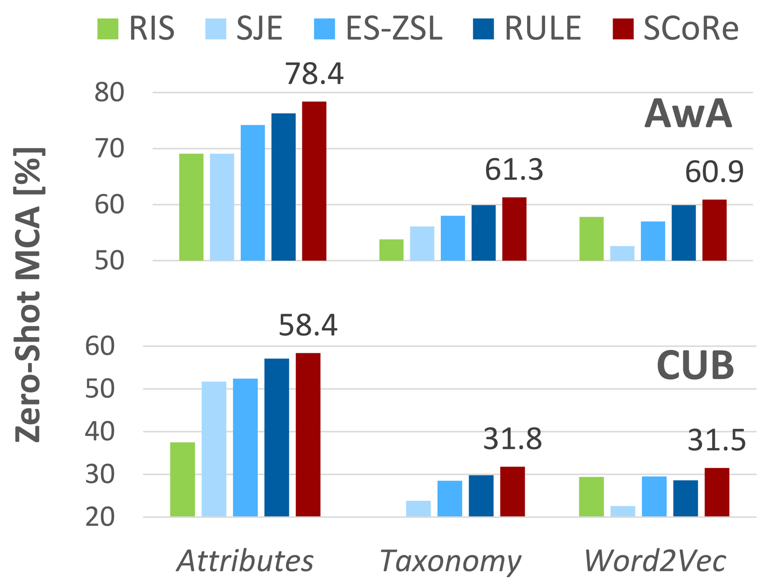


- Multiple Word2Vec embeddings
- Skip-gram model
- Corpora: Wikipedia (June, 2016)

$\psi^k(y)$: Word2Vec k^{th} model activation to class string (e.g. "lion")

Zero-Shot Experiments

Quantitative Results



	AwA A	AwA G	AwA V	CUB A	CUB G	CUB V
DAP [32]	45.3†	59.5‡	-	16.9†	36.6‡	-
SJE [4]	61.9	66.7	-	40.3	50.1	-
ES-ZSL§ [49]	53.0	74.2	74.4	40.6	53.1	49.0
Changpinyo *et al.* [8]	-	72.9	-	-	54.7	-
Xian *et al.* [60]	-	72.5	-	-	45.6	-
Zhang *et al.* [61]	-	-	76.3	-	-	30.4
Gan *et al.* [21]	-	-	73.8	-	-	43.7
Deep-RIS	56.6	68.9	66.4	24.3	37.5	39.1
Deep-RULE	65.3	76.3	78.0	46.0	57.1	57.9
Deep-SCoRe	**66.7**	**78.3**	**82.8**	**48.5**	**58.4**	**59.5**

Attribute encodings. G: GoogLeNet; V: VGG19.

Predictions

AwA CUB Failures

Extended Annotations

Classes	Hippo	✓	Leopard	✓	White-breast Kingfisher	✓	Acadian Flycatcher	✗
Attributes	No tail	✓	Spots	✓	Beak size: same as head	✓	No black beak	✓
	Brown	✗	Yellow	✓	Solid tail pattern	✓	Multi-color belly	✗
	Big	✓	No horns	✓	No stripped wings	✓	Small	✓
Taxonomy	Mammal	✓	Mammal	✓	Kingfisher	✓	Passerine	✓
	Ungulate	✓	Carnivore	✓	River kingfisher	✗	Tyrannid	✓
	Even-toed	✓	Feline	✓			Flycatcher	✓
	Hippotamus	✓	Lynx rufus	✗				

Supported by:

FCT Fundação para a Ciência e a Tecnologia
NSF

Reproducible research:

Download this poster:

SVCL