Zero-shot Learning with Class Description Regularization

Shayan Kousha, Marcus A. Brubaker

LASSONDE SCHOOL OF ENGINEERING | YORK U

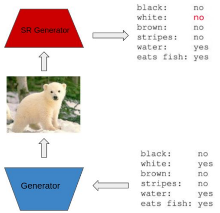
VECTOR INSTITUTE

Project page:

Introduction

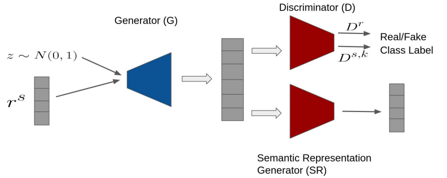
Generative Zero-Shot Learning (GZSL) is a learning methods which aims to leverage information other than the samples from different classes, typically natural language descriptions of classes, to generate samples from classes with little or no directly labelled data available.

- Our model aims to encourage the generative model to pay closer attention to the details of the descriptions.
- The model includes a mapping from the generated visual features back to the original text or attributes of a class which, forces the generator to pay closer attention to these inputs.
- We introduce a loss function which penalizes the generator and regularizer if the generated description is not similar to the input description.



Method

- The base of our generative ZSL model is a generative adversarial network and was introduced in [1].
- We add a semantic representation generator network (SR) and corresponding loss to this model to learn mapping from the visual features of a sample to the semantic representation of the class.



- We explore the unseen space of the generator G by augmenting the training process to include hallucinated semantic representations of new classes.

Semantic representation of two random classes

Hallucinated semantic representation

$$r^h = \alpha r_a^s + (1-\alpha) r_b^s$$

- We introduce a loss function to penalize differences between the output of the SR network and the provided semantic representation of the class

$$\begin{aligned} L_{SR} = & -\mathbb{E}_{x, r^s \sim p_{data}}[sim(r^s, SR(x))] \\ & -\mathbb{E}_{z\sim p_z, r^s \sim p^s_{rep}}[sim(r^s, SR(G(r^s, z)))] \\ & -\mathbb{E}_{z\sim p_z, r^h \sim p^h_{rep}}[sim(r^h, SR(G(r^h, z)))] \end{aligned}$$

Experiments

- Textual-based datasets: *Caltech UCSD Birds-2011* (CUB) and *North America Birds* (NAB)
- Attribute-based datasets: AWA2, SUN, and APY.
- Metrics: Top-1 unseen class acc, Generalized Zero-shot performance with area under Seen-Unseen curve, and Harmonic mean.

The most relevant baselines for our methods are the GAZSL [1] and CIZSL [2] ZSL models on which our approach is built.

Metric	Top-1 Accuracy (%)			Seen-Unseen H (%)		
Dataset	AWA2	SUN	APY	AWA2	SUN	APY
GAZSL [1]	56.33	60.76	27.18	28.36	25.59	14.77
CIZSL [2]	56.13	61.52	30.98	26.66	26.27	16.31
DGRZSL (Our model)	**57.79**	**61.94**	**38.59**	**33.50**	**26.70**	**18.09**

- DGRZSL outperforms the SOTA baseline methods in top-1 accuracy in all cases with an average improvement of 9.4% in the case of the APY dataset.
- DGRZSL displays significantly improved performance in the seen-unseen H metric, improving the accuracy by 29.65%, 1.63%, and 10.91% over state-of-the-art on AWA2, SUN, APY, respectively.

Metric	Top-1 Accuracy (%)				Seen-Unseen AUC (%)			
Dataset	CUB		NAB		CUB		NAB	
Split-mode	Easy	Hard	Easy	Hard	Easy	Hard	Easy	Hard
GAZSL [1]	42.40	18.83	41.0	**9.34**	39.13	**15.16**	28.71	6.43
CIZSL [2]	41.38	**18.99**	40.56	9.29	38.65	14.93	27.86	6.69
DGRZSL (Our model)	**42.44**	17.36	**42.71**	8.76	**39.77**	11.70	**29.81**	**6.87**

- DGRZSL is most successful on easy splits resulting in average improvements of 2.13% and 2.73% for top-1 accuracy and seen-unseen H, respectively.

DGRZSl achieve the highest accuracy in all most all cases when seen-unseen metrics are uses, showing a great generalization capabilities.

References

[1] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. "A generative adversarial approach for zero-shot learning from noisy texts." In CVPR, 2018.

[2] Mohamed Elhoseiny and Mohamed Elfeki. "Creativity inspired zero shot learning." In ICCV, 2019.