Contrastive Embedding for Generalized Zero-Shot Learning
Supplementary Material

Zongyan Han[1], Zhenyong Fu[1*], Shuo Chen[2,1] and Jian Yang[1*]
[1] PCALab, Nanjing University of Science and Technology, China
[2] RIKEN Center for Advanced Intelligence Project, Japan
{hanzy, z.fu, csjyang}@njust.edu.cn shuo.chen.ya@riken.jp

A. Dimension of the Embedding Space

In Table A1, we report the results of our hybrid GZSL with contrastive embedding (CE-GZSL) with respect to different dimensions of the embedding h. On each of the datasets, as the dimension of the embedding h grows, the performance of our CE-GZSL improves significantly. However, a high dimensional embedding space will inevitably increase the computational burden. Thus, in our experiments, we set the dimension of the embedding h to 2,048 in order to achieve a trade-off between performance and computational cost.

Table A1: Our CE-GZSL results with respect to different dimensions of the embedding h. U and S are the Top-1 accuracies tested on unseen classes and seen classes, respectively, in GZSL. H is the harmonic mean of U and S.

Dimension	AWA1			AWA2			CUB			FLO			SUN		
	U	S	H	U	S	H	U	S	H	U	S	H	U	S	H
256	52.9	53.9	53.3	56.6	71.8	63.3	62.5	51.5	62.0	63.0	47.2	53.9	46.7	28.9	35.7
512	65.0	55.7	60.0	60.7	75.4	67.3	62.2	66.1	64.1	68.1	58.0	62.6	**50.1**	33.6	40.2
1,024	**65.7**	65.1	65.4	61.9	77.9	69.0	62.2	66.5	64.3	63.2	78.0	69.8	48.7	36.7	41.9
2,048	64.9	73.9	**69.0**	63.1	**78.6**	**70.0**	**63.9**	66.8	**65.3**	**69.0**	78.7	73.5	48.8	38.6	**43.1**
4,096	62.9	**76.4**	**69.0**	**66.0**	74.5	**70.0**	63.0	**67.6**	65.2	68.5	**80.3**	**73.9**	45.6	**39.0**	42.1

B. Results on aPY

We further evaluate our CE-GZSL on Attribute Pascal and Yahoo (aPY) [1], which contain 15,339 images from 32 categories and each category is annotated by 64 attributes. We only find four methods that report their GZSL results on aPY and we compare our method with them. The results are shown in Table A2. Our method can achieve the second-best U and H on aPY.

Table A2: We compare our CE-GZSL with the state-of-the-art method on aPY. U and S are the Top-1 accuracies tested on unseen classes and seen classes, respectively, in GZSL. H is the harmonic mean of U and S.

Method	aPY		
	U	S	H
TCN [2]	24.1	64.0	35.1
Li *et al.* [4]	26.5	**74.0**	39.0
LisGAN [3]	34.3	<u>68.2</u>	45.7
IZF [5]	**42.3**	60.5	**49.8**
Our CE-GZSL	<u>35.5</u>	65.4	<u>46.0</u>

*Corresponding authors.

References

[1] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. Describing objects by their attributes. In *CVPR*, 2009. 1

[2] Huajie Jiang, Ruiping Wang, Shiguang Shan, and Xilin Chen. Transferable contrastive network for generalized zero-shot learning. In *ICCV*, 2019. 1

[3] Jingjing Li, Mengmeng Jing, Ke Lu, Zhengming Ding, Lei Zhu, and Zi Huang. Leveraging the invariant side of generative zero-shot learning. In *CVPR*, 2019. 1

[4] Kai Li, Martin Renqiang Min, and Yun Fu. Rethinking zero-shot learning: A conditional visual classification perspective. In *ICCV*, 2019. 1

[5] Yuming Shen, Jie Qin, Lei Huang, Fan Zhu, and Lin Shao. Invertible zero-shot recognition flows. In *ECCV*, 2020. 1