Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly

Yongqin Xian, Christoph H. Lampert, Bernt Schiele and Zeynep Akata

1 Proposed Splits Version 2.0

In this section, we introduce our Proposed Split Version 2.0 which fixes a mistake in our original data split.

For each dataset, our data split consists of the following four sets: *training set*, *test set of seen classes*, *test set of unseen classes*, *train set* and *val set*. The data split follows the truly zero-shot setting where the *training set* and the *test set of unseen classes* belong to mutually exclusive seen and unseen classes respectively. The *train set* and the *val set* are supposed to be disjoint subsets of the *training set* and are used for cross-validation. However, the *train set* and the *val set* in our released data split accidentally overlap with the *test set of seen classes*, which is flawed. Although this issue does not violate the zero-shot learning rule, it might affect seen class performance in generalized zero-shot learning setting. Therefore, we updated our data splits by removing the test images of seen classes from the *train set* and the *val set*. The rest of the data splits remain to be the same i.e., *training set*, *test set of unseen classes*, *test set of seen classes*. We call this new data splits as Proposed Split version 2.0.

2 Experiments

In this section, we repeat the zero-shot and generalized zero-shot learning experiments on CUB, SUN, AWA1, AWA2 and APY using the updated data splits i.e., Proposed Split Version 2.0 (V2), and compare with the previous published results using the original data split i.e., proposed split version 1.0 (V1). All the 13 zero-shot approaches that are evaluated in the original paper are re-evaluated.

Table 1 shows the zero-shot learning results using V1 and V2. We observe that most of methods are robust to the updates of *train set* and the *val set* i.e., the performance changes from V1 to V2 are almost negligible (absolute difference is less than 0.5%) in 54 out of 65 cases. We find that SYNC and CONSE, are sensitive to hyperparameters and therefore the updates of the validation sets lead to significant performance changes. Generalized zero-shot learning results using V2 are shown in Table 2. Similarly, in most of cases, the result difference between V1 and V2 is negligible. Overall, the proposed split version 2.0 does not lead to significant performance changes for most of approaches. The conclusions made in the original paper still hold.

References

[1] C. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," in *TPAMI*, 2013. 2

[2] M. Norouzi, T. Mikolov, S. Bengio, Y. Singer, J. Shlens, A. Frome, G. Corrado, and J. Dean, "Zero-shot learning by convex combination of semantic embeddings," in *ICLR*, 2014. 2

[3] R. Socher, M. Ganjoo, C. D. Manning, and A. Ng, "Zero-shot learning through cross-modal transfer," in *NIPS*, 2013. 2

[4] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *ICCV*, 2015. 2

[5] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele, "Latent embeddings for zero-shot classification," in *CVPR*, 2016. 2

[6] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Label-embedding for image classification," *TPAMI*, 2016. 2

[7] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, M. A. Ranzato, and T. Mikolov, "Devise: A deep visual-semantic embedding model," in *NIPS*, 2013, pp. 2121–2129. 2

[8] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele, "Evaluation of output embeddings for fine-grained image classification," in *CVPR*, 2015. 2

[9] B. Romera-Paredes and P. H. Torr, "An embarrassingly simple approach to zero-shot learning," *ICML*, 2015. 2

[10] S. Changpinyo, W.-L. Chao, B. Gong, and F. Sha, "Synthesized classifiers for zero-shot learning," in *CVPR*, 2016. 2

[11] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *CVPR*, 2017. 2

[12] V. K. Verm and P. Rai, "A simple exponential family framework for zero-shot learning," in *ECML*, 2017, pp. 792–808. 2

	SUN		CUB		AWA1		AWA2		aPY	
Method	**V1**	**V2**	**V1**	**V2**	**V1**	**V2**	**V1**	**V2**	**V1**	**V2**
DAP [1]	39.9	39.9	40.0	40.0	44.1	44.1	46.1	46.1	33.8	33.8
IAP [1]	19.4	19.4	24.0	24.0	35.9	35.9	35.9	35.9	36.6	36.6
CONSE [2]	38.8	38.0	34.3	33.6	45.6	46.3	44.5	44.6	26.9	26.4
CMT [3]	39.9	40.1	34.6	34.6	39.5	39.5	37.9	37.9	28.0	28.0
SSE [4]	51.5	51.5	43.9	43.9	60.1	60.1	61.0	61.0	34.0	35.0
LATEM [5]	55.3	55.6	49.3	49.6	55.1	55.1	55.8	55.8	35.2	36.8
ALE [6]	58.1	58.1	54.9	54.9	59.9	59.9	62.5	62.5	39.7	**39.7**
DEVISE [7]	56.5	56.5	52.0	52.0	54.2	54.2	59.7	59.7	**39.8**	37.0
SJE [8]	53.7	52.7	53.9	53.9	65.6	65.6	61.9	61.9	32.9	31.7
ESZSL [9]	54.5	54.5	53.9	51.9	58.2	58.2	58.6	58.6	38.3	38.3
SYNC [10]	56.3	56.2	**55.6**	**56.0**	54.0	51.8	46.6	49.3	23.9	23.9
SAE [11]	40.3	40.3	33.3	33.3	53.0	53.0	54.1	54.1	8.3	8.3
GFZSL [12]	**60.6**	**60.8**	49.3	49.3	**68.3**	**68.2**	**63.8**	**63.8**	38.4	38.4

TABLE 1: Zero-shot learning results on SUN, CUB, AWA1, AWA2 and aPY using V1 = Proposed Splits Version 1.0, V2 = Proposed Splits Version 2.0 with ResNet features. The results report top-1 accuracy in %. We highlight the numbers in red if the difference between V1 and V2 is larger than 0.5%.

	SUN			CUB			AWA1			AWA2			aPY		
Method	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**	**ts**	**tr**	**H**
DAP [1]	4.2	25.1	7.2	1.7	67.9	3.3	0.0	88.7	0.0	0.0	84.7	0.0	4.8	78.3	9.0
IAP [1]	1.0	37.8	1.8	0.2	72.8	0.4	2.1	78.2	4.1	0.9	87.6	1.8	5.7	65.6	10.4
CONSE [2]	6.8	35.9	11.4	2.0	70.6	3.9	0.4	**89.6**	0.8	0.5	**90.6**	1.0	0.0	**91.2**	0.0
CMT [3]	8.1	21.8	11.8	7.2	49.8	12.6	0.9	87.6	1.8	0.5	90.0	1.0	1.4	85.2	2.8
CMT* [3]	8.7	28.0	13.3	4.7	60.1	8.7	8.4	86.9	15.3	8.7	89.0	15.9	**10.9**	74.2	**19.0**
SSE [4]	2.1	**36.4**	4.0	8.5	46.9	14.4	7.0	80.5	12.9	8.1	82.5	14.8	0.3	78.4	0.6
LATEM [5]	14.7	28.8	19.5	15.2	57.3	24.0	7.3	71.7	13.3	11.5	77.3	20.0	1.3	71.4	2.6
ALE [6]	**21.8**	33.1	**26.3**	23.7	62.8	**34.4**	**16.8**	76.1	**27.5**	14.0	81.8	23.9	4.6	73.7	8.7
DEVISE [7]	16.9	27.4	20.9	**23.8**	53.0	32.8	13.4	68.7	22.4	**17.1**	74.7	**27.8**	3.5	78.4	6.7
SJE [8]	14.4	29.7	19.4	23.5	59.2	33.6	11.3	74.6	19.6	8.0	73.9	14.4	1.3	71.4	2.6
ESZSL [9]	11.0	27.9	15.8	14.7	56.5	23.3	6.6	75.6	12.1	5.9	77.8	11.0	2.4	70.1	4.6
SYNC [10]	7.9	43.3	13.4	11.5	**70.9**	19.8	9.0	88.9	16.3	9.7	89.7	17.5	7.4	66.3	13.3
SAE [11]	8.8	18.0	11.8	7.8	54.0	13.6	1.8	77.1	3.5	1.1	82.2	2.2	0.4	80.9	0.9
GFZSL [12]	0.0	39.6	0.0	0.0	45.7	0.0	1.8	80.3	3.5	2.5	80.1	4.8	0.0	83.3	0.0

TABLE 2: Generalized Zero-Shot Learning on Proposed Split Version 2.0 (PS) measuring ts = Top-1 accuracy on $\mathcal{Y}^{ts}$, tr=Top-1 accuracy on $\mathcal{Y}^{tr}$, H = harmonic mean (CMT*: CMT with novelty detection). We measure top-1 accuracy in %. We highlight the numbers in red if the difference between V1 and V2 is larger than 0.5%.