Augmentation Network for Generalised Zero-Shot Learning Supplementary Material

Rafael Felix[1,2,3][0000−0002−6186−9426], Michele Sasdelli[1,2,3][0000−0003−1021−6369], Ian Reid[1,2,3][0000−0001−7790−6423], and Gustavo Carneiro[1,2,3][0000−0002−5571−6220]

[1] The University of Adelaide, Australia
[2] Australian Institute for Machine Learning (AIML),
[3] Australian Centre for Robotic Vision (ACRV),
{rafael.felixalves,michele.sasdelli,ian.reid,gustavo.carneiro}@adelaide.edu.au

1 Architecture

In Fig. 1, *we depict the proposed AN-GZSL, which represents the first method to perform inference with multiple modalities without the use of external domain classifiers for modulating the inference between seen and unseen classes.*

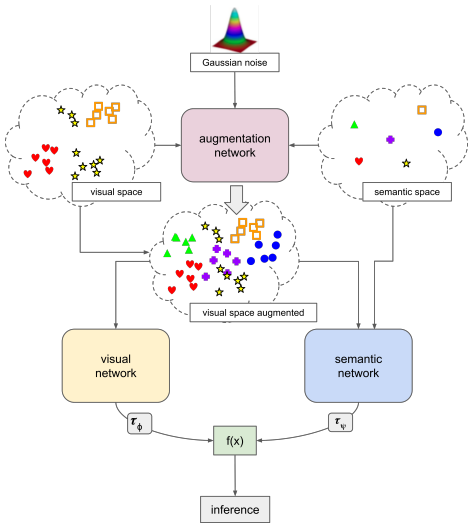

Fig. 1. Our proposed model Augmentation Network for multi-modal and multi-domain Generalised Zero-Shot Learning (AN-GZSL). AN-GZSL is composed of the augmentation network to generates visual samples, the visual and the semantic networks, a classification calibration (represented by τ_ψ and τ_ϕ in (2)) that enables multi-domain classification, and the multi-modal classification that combines the visual and semantic modules.

2 Data set additional information

In Table 1, we report additional information for the benchmark data sets, they are CUB, FLO, SUN and AWA.

Table 1. Information about CUB[1], FLO[2], SUN [3], AWA[4], and ImageNet [5, 6]. Column (1) shows the number of seen classes, denoted by $|\mathcal{Y}^S|$, split into the number of training and validation classes (train+val), (2) presents the number of unseen classes $|\mathcal{Y}^U|$, (3) displays the number of samples available for training $|\mathcal{D}^{Tr}|$ and (4) shows number of testing samples that belong to the unseen classes $|\mathcal{D}_U^{Te}|$ and number of testing samples that belong to the seen classes $|\mathcal{D}_S^{Te}|$ from [7].

Name	$\lvert\mathcal{Y}^S\rvert$ (train+val)	$\lvert\mathcal{Y}^U\rvert$	$\lvert\mathcal{D}^{Tr}\rvert$	$\lvert\mathcal{D}_U^{Te}\rvert + \lvert\mathcal{D}_S^{Te}\rvert$
CUB	150 (100+50)	50	7057	1764+2967
FLO	82 (62+20)	20	1640	1155+5394
SUN	745 (580+65)	72	14340	2580+1440
AWA	40 (27+13)	10	19832	4958+5685
ImageNet	1000 (1000 + 0)	100	1.2kk	5200+50k

3 Run-Time and Performance

The training of AN-GZSL models takes an average of 4 hours to train per data set on a *Nvidia Titan XP*, which is comparable to similar generative approaches [2,43,44]. The algorithm to perform grid search is based on random search for each parameter τ_ϕ and τ_ψ. For both parameters, the search is performed from N values randomly selected between $[1 \times 10^{-5}$,1]. We select the value (for each parameter) that produces the best confidence performance on the validation set. This parameter search takes less then two minutes using a *Nvidia Titan XP*.

4 AUSUC

Using the graph in Fig. 2, we compute the AUSUC on each data set for AN-GZSL – results are shown in the Table 2. Moreover, we added the results reported by the previous methods EZSL [8], fCLSWGAN [9], cycle-WGAN [7] and DAZSL [10] in Table 2. We were able to compute the AUSUC results for AN-GZSL and cycle-WGAN, but the other AUSUC results were extracted from [10].

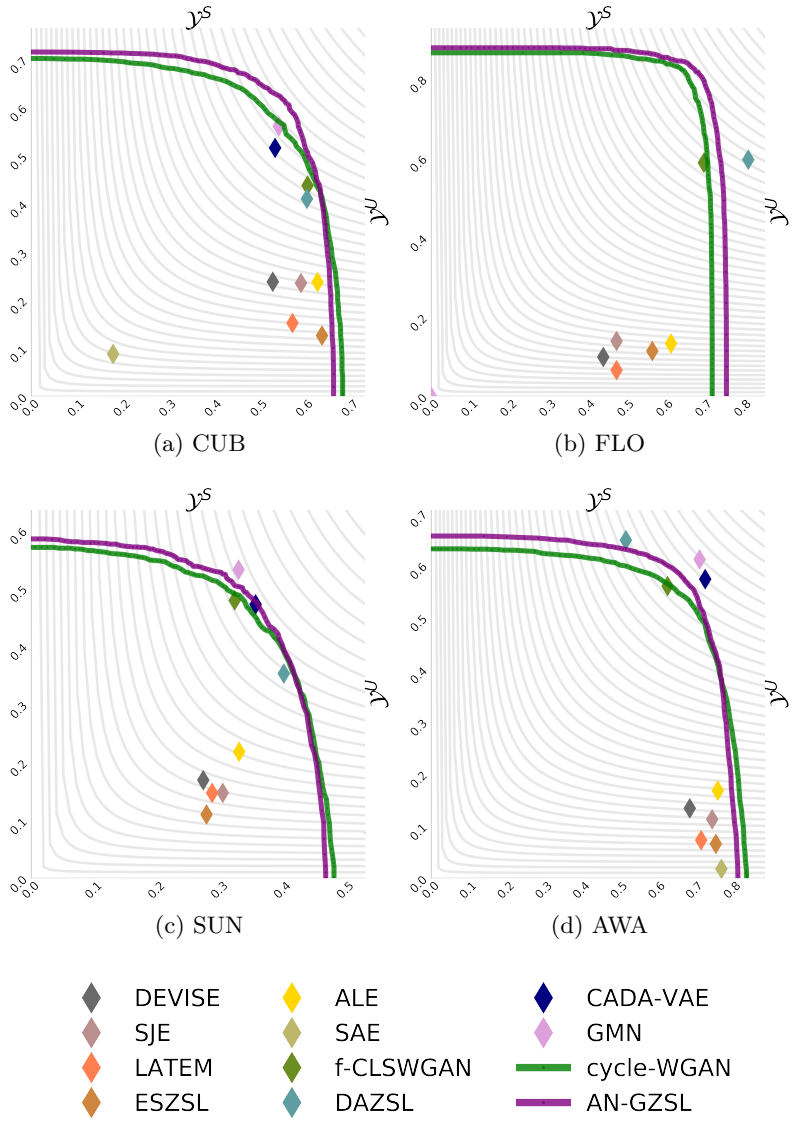

Fig. 2. ROC curves for the proposed method AN-GZSL, and several baseline and state-of-the-art methods (best seen on the digital format with colors).

Table 2. Area under the curve of seen and unseen accuracy (AUSUC). The highlighted values per column represent the best results in each dataset.

Classifier	CUB	FLO	SUN	AWA
ESZSL [8]	30.2	25.7	12.8	39.8
fCLSWGAN [9]	34.5	53.1	22.0	46.1
cycle-WGAN [7]	42.6	60.8	23.2	47.4
DAZSL [10]	35.6	58.1	21.0	**55.9**
AN − GZSL	**43.7**	**64.6**	**23.6**	47.9

References

1. Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., Perona, P.: Caltech-ucsd birds 200. (2010)
2. Nilsback, M.E., Zisserman, A.: Automated flower classification over a large number of classes. In: Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on, IEEE (2008) 722–729
3. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo. In: Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, IEEE (2010) 3485–3492
4. Xian, Y., Lampert, C.H., Schiele, B., Akata, Z.: Zero-shot learning - A comprehensive evaluation of the good, the bad and the ugly. CoRR **abs/1707.00600** (2017)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 248–255
6. Wang, P., Liu, L., Shen, C., Huang, Z., van den Hengel, A., Shen, H.T.: Multi-attention network for one shot learning. In: 2017 IEEE conference on computer vision and pattern recognition, CVPR. (2017) 22–25
7. Felix, R., Kumar, B.V., Reid, I., Carneiro, G.: Multi-modal cycle-consistent generalized zero-shot learning. In: European Conference on Computer Vision, Springer (2018) 21–37
8. Romera-Paredes, B., Torr, P.: An embarrassingly simple approach to zero-shot learning. In: International Conference on Machine Learning. (2015) 2152–2161
9. Xian, Y., Lorenz, T., Schiele, B., Akata, Z.: Feature generating networks for zero-shot learning. arXiv (2017)
10. Atzmon, Y., Chechik, G.: Adaptive confidence smoothing for generalized zero-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 11671–11680