Supplementary Material

Do Huu Dat
VinUniversity
22dat.dh@vinuni.edu.vn

Po-Yuan Mao
Academia Sinica

Tien Hoang Nguyen
VNU-UET

Wray Buntine
VinUniversity

Mohammed Bennamoun
University of Western Australia

A. Impact of Hyperparameter on Accuracy & Convergence:

Figure 1 shows that despite using different hyperparameter configurations, the accuracy on both unseen and seen data consistently converges to a similar value. The primary difference is in the speed of this convergence, with a slight performance drop observed when α and β are significantly larger than γ.

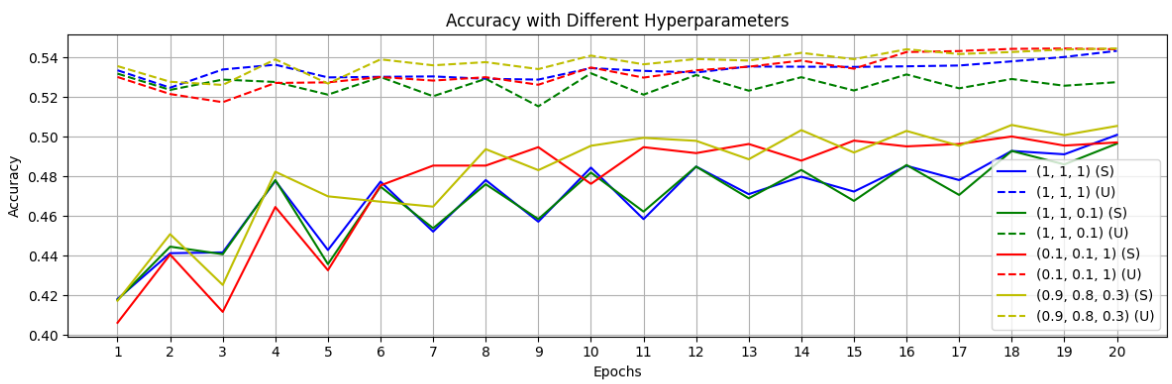

Figure 1. Seen (S) and Unseen (U) accuracy in different set of (α, β, γ)

B. Evaluations

We provide experimental comparisons in Tables 1 and 2 against all previously established compositional zero-shot learning methods, including AoP [9], LE+ [8], TMN [11], SymNet [4], CompCos [6], CGE [8], Co-CGE [7], SCEN [2], KG-SP [1], CSP [10], and DFSP [5]. Performance is assessed in both closed-world and open-world scenarios.

Method	MIT-States				UT-Zappos				C-GQA			
	S	U	H	AUC	S	U	H	AUC	S	U	H	AUC
AoP [9]	14.3	17.4	9.9	1.6	59.8	54.2	40.8	25.9	17.0	5.6	5.9	0.7
LE+ [8]	15.0	20.1	10.7	2.0	53.0	61.9	41.0	25.7	18.1	5.6	6.1	0.8
TMN [11]	20.2	20.1	13.0	2.9	58.7	60.0	45.0	29.3	23.1	6.5	7.5	1.1
SymNet [4]	24.2	25.2	16.1	3.0	49.8	57.4	40.4	23.4	26.8	10.3	11.0	2.1
CompCos [6]	25.3	24.6	16.4	4.5	59.8	62.5	43.1	28.1	28.1	11.2	12.4	2.6
CGE [8]	31.1	5.8	6.4	1.1	62.0	44.3	40.3	23.1	32.1	2.0	3.4	0.5
Co-CGE [7]	31.1	5.8	6.4	1.1	62.0	44.3	40.3	23.1	32.1	2.0	3.4	0.5
SCEN [2]	29.9	25.2	18.4	5.3	63.5	63.1	47.8	32.0	28.9	25.4	17.5	5.5
CLIP [12]	30.2	45.9	26.1	11.1	15.8	49.2	15.6	5.0	7.7	24.8	8.4	1.3
CSP [10]	46.6	49.9	36.3	19.4	64.2	66.2	46.6	33.0	28.8	26.8	20.5	6.2
CSP [10]	46.6	49.9	36.3	19.4	64.2	66.2	46.6	33.0	28.8	26.8	20.5	6.2
DFSP [5]	46.9	52.0	37.3	20.6	66.7	71.7	47.2	36.0	**38.2**	**32.0**	**27.1**	**10.5**
HOMOE	**50.5**	**54.6**	**39.9**	**23.3**	**68.4**	**73.9**	**49.1**	**37.5**	35.8	30.8	24.5	9.1

Table 1. Closed World Evaluation. Comparison to state-of-the-art models

Method	MIT-States				UT-Zappos				C-GQA			
	S	U	H	AUC	S	U	H	AUC	S	U	H	AUC
AoP [9]	16.6	5.7	4.7	0.7	50.9	34.2	29.4	13.7	-	-	-	-
LE+ [8]	14.2	2.5	2.7	0.3	60.4	36.5	30.5	16.3	19.2	0.7	1.0	0.08
TMN [10]	12.6	0.9	1.2	0.1	55.9	18.1	21.7	8.4	-	-	-	-
SymNet [4]	21.4	7.0	5.8	0.8	53.3	44.6	34.5	18.5	26.7	2.2	3.3	0.43
CompCos [6]	25.4	10.0	8.9	1.6	59.3	46.8	36.9	21.3	-	-	-	-
CGE [8]	32.4	5.1	6.0	1.0	61.7	47.7	39.0	23.1	32.7	1.8	2.9	0.47
Co-CGEˆClosed [7]	31.1	5.8	6.4	1.1	62.0	44.3	40.3	23.1	32.1	2.0	3.4	0.53
Co-CGEˆOpen [7]	30.3	11.2	10.7	2.3	61.2	45.8	40.8	23.3	32.1	3.0	4.8	0.78
KG-SP [1]	28.4	7.5	7.4	1.3	61.8	52.1	42.3	26.5	31.5	2.9	4.7	0.78
DRANet [3]	29.8	7.8	7.9	1.5	65.1	54.3	44.0	28.8	31.3	3.9	6.0	1.05
CLIP [12]	30.1	14.3	12.8	3.0	15.6	20.5	11.3	2.2	7.5	4.4	4.0	0.28
CSP [10]	46.3	15.7	17.4	5.7	64.1	44.1	38.9	22.7	28.7	5.2	6.9	1.2
DFSP [5]	47.5	18.5	19.3	5.8	66.8	60.0	44.0	30.3	**38.3**	**7.2**	**10.4**	**2.4**
HOMOE	**50.4**	**19.7**	**20.7**	**7.9**	**68.4**	**61.9**	**45.1**	**31.1**	35.7	6.6	9.0	2.0

Table 2. Open World Evaluation. Comparison to state-of-the-art models

References

[1] Shyamgopal Karthik, Massimiliano Mancini, and Zeynep Akata. Kg-sp: Knowledge guided simple primitives for open world compositional zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9336–9345, 2022. 2

[2] Xiangyu Li, Xu Yang, Kun Wei, Cheng Deng, and Muli Yang. Siamese contrastive embedding network for compositional zero-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9326–9335, 2022. 2

[3] Yun Li, Zhe Liu, Saurav Jha, and Lina Yao. Distilled reverse attention network for open-world compositional zero-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1782–1791, 2023. 2

[4] Yong-Lu Li, Yue Xu, Xiaohan Mao, and Cewu Lu. Symmetry and group in attribute-object compositions. 2020 ieee. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11313–11322, 2020. 2

[5] Xiaocheng Lu, Song Guo, Ziming Liu, and Jingcai Guo. Decomposed soft prompt guided fusion enhancing for compositional zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23560–23569, 2023. 2

[6] Massimiliano Mancini, Muhammad Ferjad Naeem, Yongqin Xian, and Zeynep Akata. Open world compositional zero-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5222–5230, 2021. 2

[7] Massimiliano Mancini, Muhammad Ferjad Naeem, Yongqin Xian, and Zeynep Akata. Learning graph embeddings for open world compositional zero-shot learning. *IEEE Transactions on pattern analysis and machine intelligence*, 2022. 2

[8] Muhammad Ferjad Naeem, Yongqin Xian, Federico Tombari, and Zeynep Akata. Learning graph embeddings for compositional zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 953–962, 2021. 2

[9] Tushar Nagarajan and Kristen Grauman. Attributes as operators: factorizing unseen attribute-object compositions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 169–185, 2018. 2

[10] Nihal V Nayak, Peilin Yu, and Stephen Bach. Learning to compose soft prompts for compositional zero-shot learning. In *The Eleventh International Conference on Learning Representations*, 2022. 2

[11] Senthil Purushwalkam, Maximilian Nickel, Abhinav Gupta, and Marc'Aurelio Ranzato. Task-driven modular networks for zero-shot compositional learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3593–3602, 2019. 2

[12] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 2