LogiCzsl: Exploring Logic-induced Representation for Compositional Zero-shot Learning Supplementary Material

Peng Wu[1], Xiankai Lu[1]*, Hao Hu[1], Yongqin Xian[3], Jianbing Shen[4], Wenguan Wang[2]

[1]Shandong University, [2] ReLER, CCAI, Zhejiang University, [3] Google, [4]SKL-IOTSC, CIS, University of Macau

https://github.com/Pieux0/LOGICZSL

In this document, we provide dataset details, more ablative experimental results, more qualitative visualization, and the detailed introduction of first-order logic. It is organized as follows:

- § A Summary of data split statistics
- § B More ablative experimental results
- § C More qualitative visualization
- § D First-order logic

A. Summary of data split statistics

We follow the common data splits of the three used datasets, *i.e.*, CGQA [3], UT-Zappos50K [5] and MIT-States [2]. Please see Table 1 for details.

Table 1. Summary of data split statistics.

	Composition			Train		Val		Test	
Datasets	$\lvert\mathcal{A}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{A}\rvert \times \lvert\mathcal{O}\rvert$	$\lvert\mathcal{C}_s\rvert$	$\lvert\mathcal{X}\rvert$	$\lvert\mathcal{C}_s\rvert$ / $\lvert\mathcal{C}_u\rvert$	$\lvert\mathcal{X}\rvert$	$\lvert\mathcal{C}_s\rvert$ / $\lvert\mathcal{C}_u\rvert$	$\lvert\mathcal{X}\rvert$
CGQA	413	674	278362	5592	26920	1252 / 1040	7280	888 / 923	5098
UT-Zappos50K	16	12	192	83	22998	15 / 15	3214	18 / 18	2914
MIT-States	115	245	28175	1262	30338	300 / 300	10420	400 / 400	12995

B. More ablative experimental results

We supplement more ablative experimental results with CSP and Troika for comprehensive evaluation of LogiCzsl, including key component analysis (Table 2), aggregation coefficient (Table 3 and Table 5), the importance of logic-induced loss (Table 4 and Table 6), the robustness and efficiency in Table 7. In particular, the results of cross-domain (*i.e.*, training on UT-Zappos and testing on sketch UT-Zappos) and cross-task experiments (*i.e.*, training on CZSL and testing on ZSL) highlight the superior robustness of our method against data distribution shift than baselines.

C. More qualitative visualization

We present more qualitative results from CGQA [3], UT-Zappos50K [5] and MIT-States [2] in Fig. 1. We observe that our proposed LogiCzsl is capable of predicting accurate results where base models (*i.e.*, CSP and Troika)

*Corresponding author: *Xiankai Lu*.

Table 2. Ablative experimental results of logic rules on CGQA [3] with base models, *i.e.*, CSP and Troika.

Method	Logic Rules			CGQA [3]			
	oi-logic	oe-logic	ae-logic	AUC↑	HM↑	Seen↑	Unseen↑
Logic-CSP	✗	✗	✗	6.2	20.5	28.8	26.8
	✓	✗	✗	7.4	22.7	31.1	28.3
	✗	✓	✗	7.1	22.0	30.5	27.8
	✗	✗	✓	7.1	22.0	30.6	27.8
	✓	✓	✓	**7.7**	**23.1**	**31.9**	**28.9**
Logic-Troika	✗	✗	✗	12.4	29.4	41	35.7
	✓	✗	✗	14.3	32.2	42.2	39.0
	✗	✓	✗	15.1	33.1	43.7	**39.5**
	✗	✗	✓	14.2	32.1	42.1	38.8
	✓	✓	✓	**15.3**	**33.3**	**44.4**	39.4

Table 3. Aggregation coefficient q in Eq.3 of manuscript for $\forall$ on CGQA [3] with base models, *i.e.*, CSP and Troika.

Method	Aggregation Coefficient q	CGQA [3]			
		AUC↑	HM↑	Seen↑	Unseen↑
Logic-CSP	baseline (*w/o* logic rules)	6.20	20.50	28.80	26.80
	1.0	7.62	22.99	**32.02**	28.58
	3.0	7.62	22.99	31.99	28.58
	5.0	**7.63**	**23.02**	31.99	**28.67**
	8.0	7.04	21.86	30.55	27.62
Logic-Troika	baseline (*w/o* logic rules)	12.40	29.40	41.00	35.70
	1.0	15.14	33.22	44.26	38.81
	3.0	15.32	**33.43**	44.21	39.51
	5.0	**15.34**	33.30	**44.41**	39.42
	8.0	15.20	33.21	43.75	**39.95**

Table 4. Importance of logic-induced loss in Eq.13 of manuscript on CGQA [3] with base models, *i.e.*, CSP and Troika.

Method	Importance of LG	CGQA [3]			
		AUC↑	HM↑	Seen↑	Unseen↑
Logic-CSP	baseline (*w/o* logic rules)	6.20	20.50	28.80	26.80
	s_1	7.64	22.91	31.79	**28.93**
	s_2	7.63	23.02	**31.99**	28.67
	s_3	**7.65**	**23.05**	31.92	28.85
Logic-Troika	baseline (*w/o* logic rules)	12.40	29.40	41.00	35.70
	s_1	14.25	32.21	42.03	38.90
	s_2	**15.34**	**33.30**	**44.41**	**39.42**
	s_3	15.24	33.23	44.31	39.16

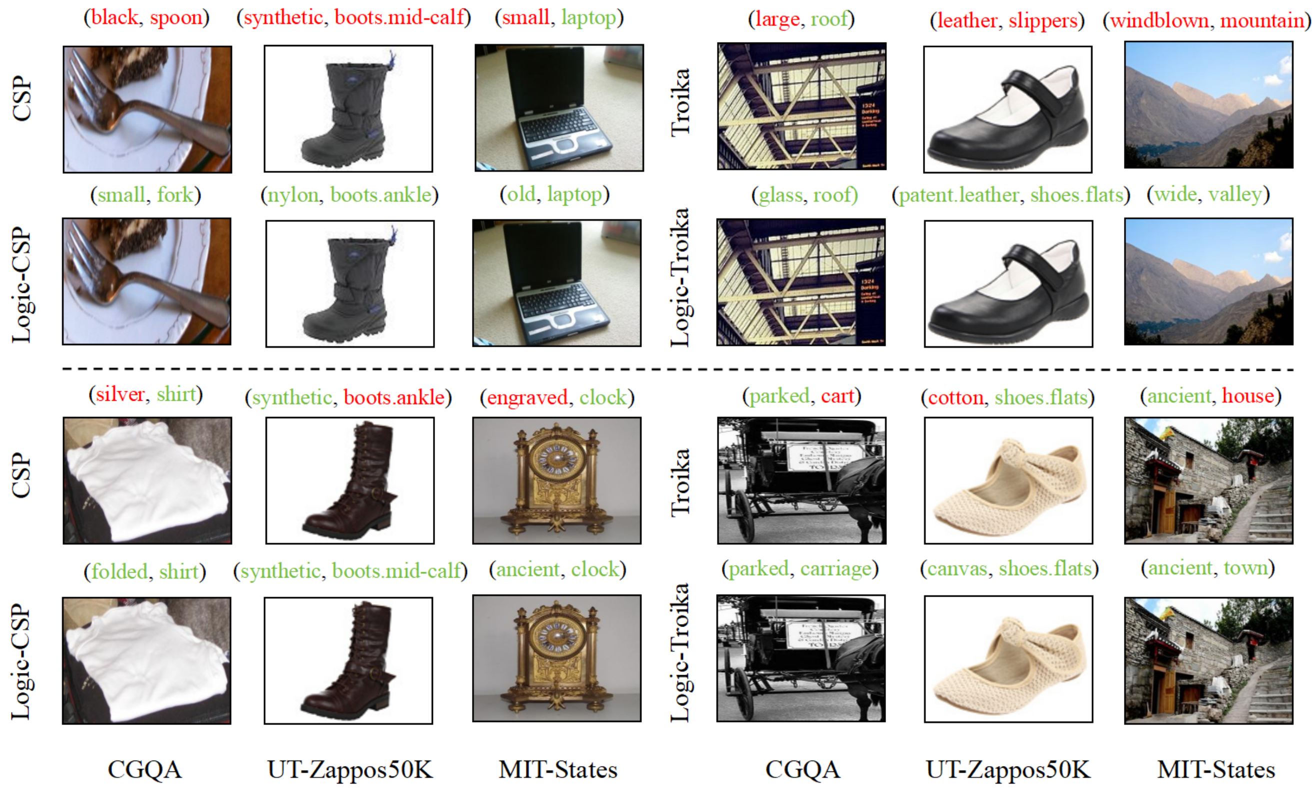

Figure 1. More visual comparisons between CSP [4], Troika [1] and our method LOGICZSL on CGQA [3], UT-Zappos50K [5], MIT-States [2]. Green denotes the right prediction and red denotes the wrong prediction.

Figure 2. Failure cases of prediction results

Table 5. Aggregation coefficient q in Eq.3 of manuscript for $\forall$ on MIT-States [2] with base models, *i.e.*, CSP and Troika.

Method	Aggregation Coefficient q	MIT-States [2]			
		AUC↑	HM↑	Seen↑	Unseen↑
Logic-CSP	baseline (*w/o* logic rules)	19.40	36.30	46.60	49.90
	1.0	20.56	37.42	48.45	50.99
	3.0	20.55	37.40	48.40	51.02
	5.0	**20.59**	**37.42**	**48.49**	51.00
	8.0	20.31	37.38	47.10	**51.74**
Logic-Troika	baseline (*w/o* logic rules)	22.10	39.30	49.00	53.00
	1.0	**23.35**	**40.49**	**50.84**	53.87
	3.0	23.31	40.33	50.76	53.92
	5.0	23.34	40.44	50.71	**53.97**
	8.0	22.74	39.49	50.34	53.39

Table 6. Importance of logic-induced loss in Eq.13 of manuscript on MIT-States [2] with base models, *i.e.*, CSP and Troika.

Method	Importance of LG	MIT-States [2]			
		AUC↑	HM↑	Seen↑	Unseen↑
Logic-CSP	baseline (*w/o* logic rules)	19.40	36.30	46.60	49.90
	s_1	**20.60**	37.38	**48.49**	**51.03**
	s_2	20.59	**37.42**	48.49	51.00
	s_3	20.52	37.38	48.28	51.02
Logic-Troika	baseline (*w/o* logic rules)	22.10	39.30	49.00	53.00
	s_1	23.22	40.35	50.55	53.84
	s_2	**23.35**	**40.49**	**50.84**	53.87
	s_3	23.18	40.31	50.55	**53.89**

Table 7. Robustness and efficiency of proposed method.

Method	Cross-domain		Cross-task	Training Time	Memory
	Seen↑	Unseen↑	Accuracy ↑	min/epoch	G
Troika	29.6	32.5	54.6	10.0	61.8
Logic-Troika	30.7	33.2	57.6	10.5	65.0

make mistakes. For instance, base models struggle to distinguish between similar objects like "house" and "town", and have difficulty in recognizing the accurate attribute, such as "folded", "ancient". By exploring the logic-induced representation which resolves the semantic relationship limitation, LOGICZSL can correct the wrong predictions of base models. These results demonstrate the effectiveness of our proposed logic-induced learning framework. In addition, we also illustrate some failure cases in Fig. 2, which can be attributed to the complex visual distractions and extremely small targets.

D. First-order logic

First-order logic, also known as first order predicate logic, is a formal language for expressing knowledge. There are two basic elements for first-order logic: *predicate* and *variable*. A *predicate* v (*e.g.*, in Eq.4 of manuscript) denotes the name of a property or relation verification function $v(\boldsymbol{x})$, also known as *atom*, which returns true or false, where $\boldsymbol{x}$ is a *variable*. For example, $\mathrm{dog}(\boldsymbol{x})$ is used to verify the relation: "$\boldsymbol{x}$ is a dog". An *atom* (*e.g.*, $\mathrm{dog}(\boldsymbol{x})$) is called a *ground atom* (*e.g.*, $\mathrm{dog}(\boldsymbol{x}_1)$) if all the *variables* of it are instantiated with *constants* (*e.g.*, $\boldsymbol{x}_1$). For instance, if "$\boldsymbol{x}_1$ is a dog", then $\mathrm{dog}(\boldsymbol{x}_1)$ returns true, otherwise the result would be false. A first-order rule is defined by *predicate*, *variable* and a set of *logical connectives*, including connectives (*e.g.*, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, implication $\Rightarrow$) and quantifiers (*i.e.*, 'for all' $\forall$ or 'there exists' $\exists$). The defined first-order logic rule can be grounded on data if all the *variables* in the logic rule are instantiated with *constants*.

References

[1] Siteng Huang, Biao Gong, Yutong Feng, Yiliang Lv, and Donglin Wang. Troika: Multi-path cross-modal traction for compositional zero-shot learning. In *CVPR*, 2024. 2

[2] Phillip Isola, Joseph J Lim, and Edward H Adelson. Discovering states and transformations in image collections. In *CVPR*, 2015. 1, 2

[3] Muhammad Ferjad Naeem, Yongqin Xian, Federico Tombari, and Zeynep Akata. Learning graph embeddings for compositional zero-shot learning. In *CVPR*, 2021. 1, 2

[4] Nihal V Nayak, Peilin Yu, and Stephen H Bach. Learning to compose soft prompts for compositional zero-shot learning. In *ICLR*, 2023. 2

[5] Aron Yu and Kristen Grauman. Fine-grained visual comparisons with local learning. In *CVPR*, 2014. 1, 2