Transfer Learning in Multimodal Compositional and Relational Reasoning: Supplementary Material

Anonymous CVPR submission

Paper ID 7813

1. Datasets: COG & CLEVR-CoGenT

Variant	number of frames	maximum memory duration	maximum number of distractors	size of training set	size of validation set	size of test set
Canonical	4	3	1	10000320	500016	500016
Hard	8	7	10	10000320	500016	500016

Table 1. Details of the Canonical and Hard variants of the COG dataset.

Dataset	Cubes	Cylinders	Spheres
CoGenT-A	Gray / Blue / Brown / Yellow	Red / Green / Purple / Cyan	Any color
CoGenT-B	Red / Green / Purple / Cyan	Gray / Blue / Brown / Yellow	Any color

Table 2. Colors & shapes combinations present in CoGenT-A & -B datasets.

The CoGenT dataset [2] contains:

- Training set of 70,000 images and 699,960 questions in Condition A,
- Validation set of 15,000 images and 149,991 questions in Condition A,
- Test set of 15,000 images and 149,980 questions in Condition A (without answers),
- Validation set of 15,000 images and 150,000 questions in Condition B,
- Test set of 15,000 images and 149,992 questions in Condition B (without answers),
- Scene graphs and functional programs for all training/validation images/questions.

2. Complete COG results

Table 3. COG test set accuracies for SAMNet & COG models. Below 'paper' denotes results from [3] while 'code' denotes results of our experiments using their implementation [1].

Model	SAMNet				COG Model			
					paper	ours	ours	paper
Trained on	canonical	canonical	canonical	hard	canonical	canonical	canonical	hard
Fine tuned on	-	-	hard	-	-	-	hard	-
Tested on	canonical	hard	hard	hard	canonical	hard	hard	hard
Overall accuracy	98.0	91.6	96.5	96.1	97.6	65.9	78.1	80.1
AndCompareColor	93.5	82.7	89.2	80.6	81.9	57.1	60.7	51.4
AndCompareShape	93.2	83.7	89.7	80.1	80.0	53.1	50.3	50.7
AndSimpleCompareColor	99.2	85.3	97.6	99.4	99.7	53.4	77.1	78.2
AndSimpleCompareShape	99.2	85.8	97.6	99.2	100.0	56.7	79.3	77.9
CompareColor	98.1	89.3	95.9	99.7	99.2	56.1	67.9	50.1
CompareShape	98.0	89.7	95.9	99.2	99.4	66.8	65.4	50.5
Exist	100.0	99.7	99.8	99.8	100.0	63.5	96.1	99.3
ExistColor	100.0	99.6	99.9	99.9	99.0	70.9	99	89.8
ExistColorOf	99.9	95.5	99.7	99.8	99.7	51.5	76.1	73.1
ExistColorSpace	94.1	88.8	91.0	90.8	98.9	72.8	77.3	89.2
ExistLastColorSameShape	99.5	99.4	99.4	98.0	100.0	65.0	62.5	50.4
ExistLastObjectSameObject	97.3	97.5	97.7	97.5	98.0	77.5	61.7	60.2
ExistLastShapeSameColor	98.2	98.5	98.8	97.5	100.0	87.8	60.4	50.3
ExistShape	100.0	99.5	100.0	100.0	100.0	77.1	98.2	92.5
ExistShapeOf	99.4	95.9	99.2	99.2	100.0	52.7	74.7	72.70
ExistShapeSpace	93.4	87.5	91.1	90.5	97.7	70	89.8	89.80
ExistSpace	95.3	89.7	93.2	93.3	98.9	71.1	88.1	92.8
GetColor	100.0	95.8	99.9	100.0	100.0	71.4	83.1	97.9
GetColorSpace	98.0	90.0	95.0	95.4	98.2	71.8	73.	92.3
GetShape	100.0	97.3	99.9	99.9	100.0	83.5	89.2	97.1
GetShapeSpace	97.5	89.4	93.9	94.3	98.1	78.7	77.3	90.3
SimpleCompareShape	99.9	91.4	99.7	99.9	100.0	67.7	96.7	99.3
SimpleCompareColor	100.0	91.6	99.80	99.9	100.0	64.2	90.4	99.3

3. Complete CLEVR-CoGenT results

Table 4. Complete set of results for SAMNet on CLEVR-CoGenT.

Experiments	Test Accuracy (%) on `valA` – 150,000 samples					
	Overall	**Exist**	**Count**	**CompareInteger**	**CompareAttribute**	**QueryAttribute**
Exist only	26.07	*74.20*	0.0	59.79	59.15	0.0
Count only	14.89	0.0	*62.78*	0.0	0.0	0.0
CompareInteger only	23.44	48.15	0.0	*77.98*	54.08	0.0
CompareAttribute only	23.50	51.93	0.0	59.06	*61.73*	0.0
QueryAttribute only	34.64	0.0	0.0	0.0	0.0	*97.08*
All tasks	95.32	98.4	86.75	96.0	98.65	98.02
All tasks but Exist	90.33	*60.42*	86.12	96.18	98.52	98.60
All tasks but Count	74.59	97.51	*0.0*	94.37	98.42	98.47
All tasks but CompareInteger	91.53	98.22	86.09	*56.35*	98.78	98.40
All tasks but CompareAttribute	81.92	98.32	86.36	95.86	*23.18*	98.44
All tasks but QueryAttribute	42.17	76.79	54.64	79.87	64.13	*0.0*
Trained on all tasks – Finetune on t						
Exist	94.16	*98.04*	82.96	95.02	98.12	97.97
Count	95.10	96.46	*88.28*	94.78	97.01	98.26
CompareInteger	95.31	98.39	86.56	*96.07*	98.70	98.09
CompareAttribute	95.31	98.40	86.78	96.00	*98.66*	98.04
QueryAttribute	93.49	97.45	82.09	92.23	97.85	*97.76*

References

[1] Igor Ganichev. Cog implementation. https://github.com/google/cog, 2018. 2

[2] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2910, 2017. 1

[3] Guangyu Robert Yang, Igor Ganichev, Xiao-Jing Wang, Jonathon Shlens, and David Sussillo. A dataset and architecture for visual reasoning with a working memory. In *European Conference on Computer Vision*, pages 729–745. Springer, 2018. 2