

Image Change Captioning – Report

Since setting `batch_size` to 128 will reach the limit of CUDA memory on my machine, I reduce it a bit (to 124) when training the model, and this might have a impact on performance. Other hyperparams remain the same as `transformer_quick.yaml`.

1 Results

Evaluation results on validation set and test set are shown in table 1 and table 2, respectively. Table 3 shows more detailed results on test set.

	Bleu1	Bleu2	Bleu3	Bleu4	METEOR	ROUGE_L	CIDEr	SPICE
semantic	0.55	0.40	0.28	0.20	0.16	0.37	0.32	0.08
non-semantic	0.61	0.52	0.46	0.41	0.39	0.68	1.00	0.26
total	0.59	0.46	0.35	0.28	0.22	0.51	0.74	0.17

Table 1: Best results on validation set. Best performance on semantic and non-semantic changes come from different models. The model with best total performance is used for test.

	Bleu1	Bleu2	Bleu3	Bleu4	METEOR	ROUGE_L	CIDEr	SPICE
semantic	0.63	0.50	0.39	0.30	0.18	0.44	0.26	0.12
non-semantic	0.73	0.67	0.60	0.56	0.49	0.75	1.14	0.34
total	0.66	0.54	0.45	0.37	0.25	0.59	0.76	0.23

Table 2: Results on test set.

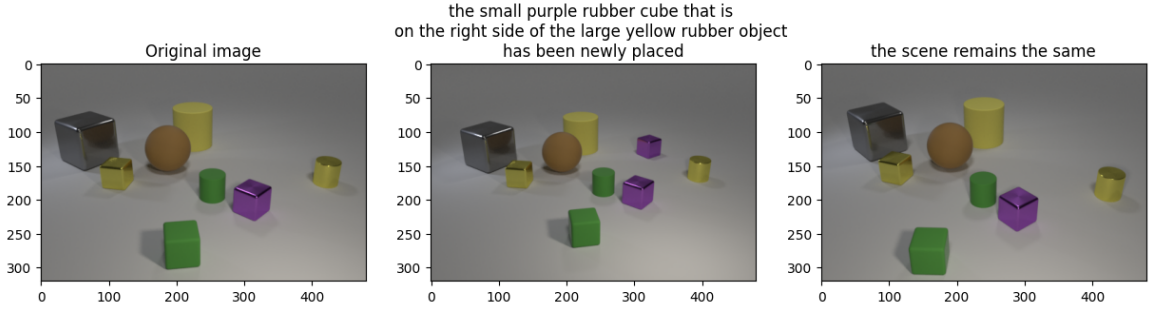
	Bleu1	Bleu2	Bleu3	Bleu4	METEOR	ROUGE_L	CIDEr	SPICE
COLOR	0.39	0.15	0.09	0.06	0.05	0.25	0.01	0.01
MATERIAL	0.40	0.18	0.11	0.08	0.05	0.26	0.01	0.01
MOVE	0.63	0.51	0.39	0.31	0.23	0.49	0.22	0.17
DROP	0.69	0.58	0.45	0.34	0.28	0.58	0.31	0.21
ADD	0.69	0.57	0.46	0.36	0.28	0.60	0.31	0.20

Table 3: Results of different types of semantic changes on test set. Performance on spacial attributes (MOVE, DROP and ADD) are significantly better than on appearance attributes (COLOR and MATERIAL).

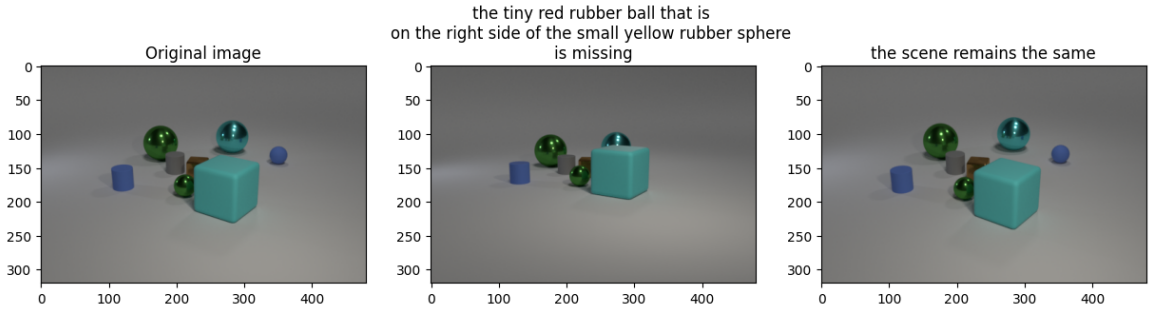
2 Visualization

Fig. 1 and fig. 2 shows some success cases and failure cases respectively. The model only performs well on cases where apparent motion occurs, but has pool performance on color or

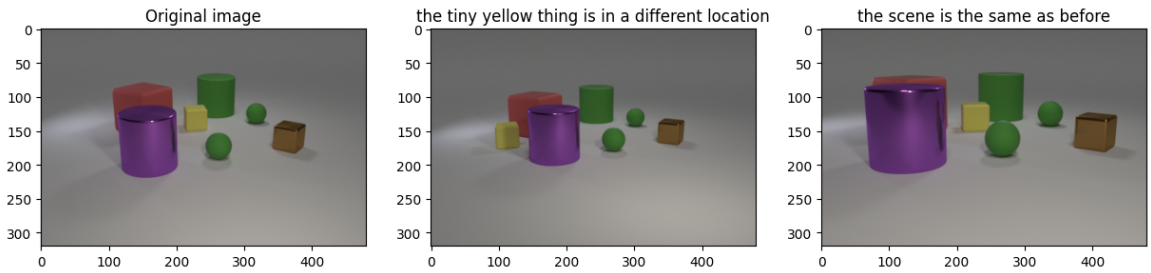
material change. The reasons are (1) the iteration number is relatively small and the model does not well converge; (2) only a fractional of the whole data ($\sim 25\%$) are used for training so the model does not learn some features related to color and material.



(a) The case where an object appears.

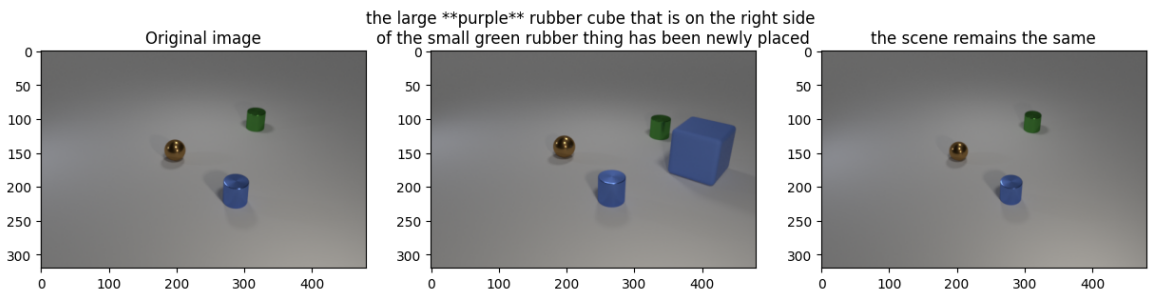


(b) The case where an object disappears. Though the colors described by the model are totally wrong, we still view it as a success case on spacial change of objects.

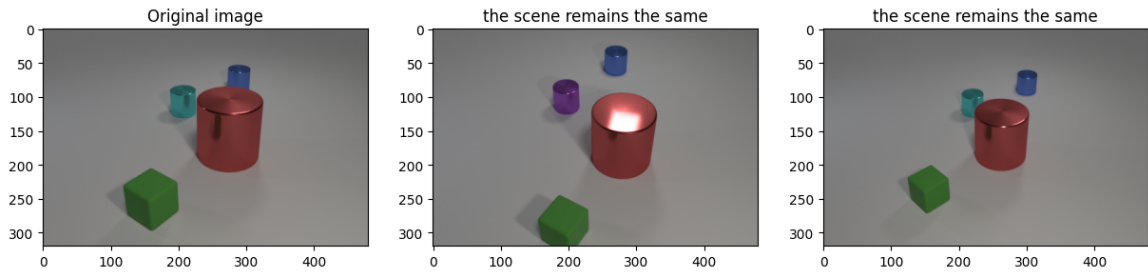


(c) The case where an object moves.

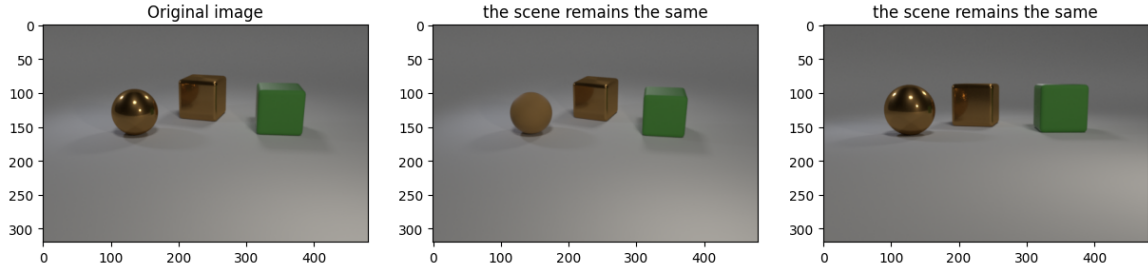
Figure 1: Success cases on spacial attribute changes. Appearances, vanishments and movements are more likely to be inferred correctly by the model.



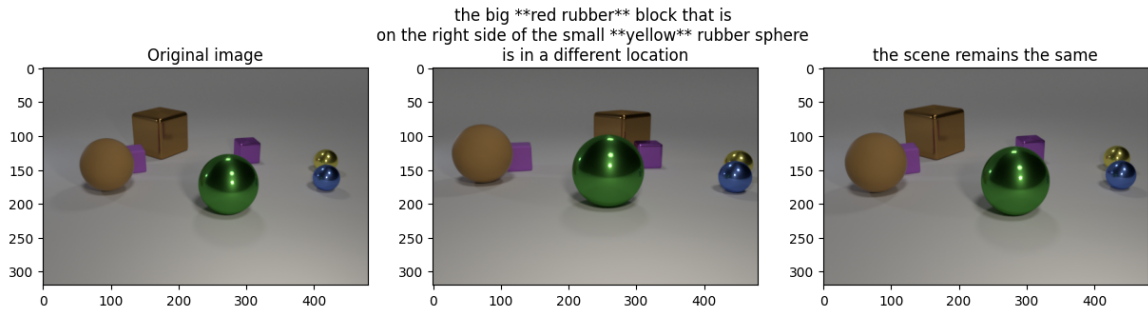
(a) Failure case where the model outputs wrong color of an object.



(b) Failure case where the model fails to discover the color change.



(c) Failure case where the model fails to discover the material change.



(d) Failure case where the model incorrectly describes color and material simultaneously.

Figure 2: Failure cases. Material and color are more likely to be wrongly inferred or ignored.