Tensor Composition Net for Visual Relationship Prediction (Supplemental Material)

BMVC 2021 Submission # 376

1 Datasets and Low Rank Assumption

Detailed statistics of the VRD and VG datasets are reported in Table 1. Our TCN-VRP relies on the assumption that relation tensors can be approximated by a set of low-rank factors. To validate this assumption, we employ HOSVD to decompose the average global tensor $\bar{\mathbf{T}}$ with different reconstruction errors. The results in Table 2 show that 2% reconstruction error can be achieved using less than half the original input dimensions, which corresponds to compressing the tensor to less than 7% of original size.

Dataset	#Train img.	#Test img.	#Avg Rels	#Objs	#Preds
VRD	4,000	1,000	7.6	100	70
VG200	73,801	25,857	11.8	200	100

Table 1: Statistics of different datasets. The number of train images, test images, relationships per image (on average), object categories and predicate categories are shown.

Dataset (Sub,Obj,Pred) Dimensions		$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.10$
VRD	Tucker Rank:	27, 26, 17	12, 10, 10	6, 4, 6
(100, 100, 70)	Compression:	0.026	0.006	0.002
VG200	Tucker Rank:	94, 73, 33	55, 36, 14	23, 16, 5
(200, 200, 100)	Compression:	0.066	0.012	0.003

Table 2: Representing relation tensor with Tucker composition. Input tensor dimension (Subj,Obj,Pred), Tucker rank and compression ratio, at various reconstruction error thresholds.

2 Qualitative results of Relation-Based Image-retrieval(RBIR)

To qualitatively evaluate our model for relation-based image-retrieval, we use four different triplets (i.e. “clock-on-tower", “person-play-Frisbee", “bird-on-branch" and “boat-in-water") as image retrieval queries, as shown in Figure 1. For “clock-on-tower" and “bird-on-branch", our top five returned images match the query exactly. As for “ person-play-Frisbee", our model retrieved a wrong image (the third of second row) since the person is not “playing" Frisbee although “person" and “Frisbee" objects exist. Another wrong retrieval result is the third ranked result for “boat-in-water". The image is tagged instead with “boat-on-water", but this should also be regarded as a correct retrieval given “boat-in-water”.

Figure 1: Qualitative examples of relation-based image-retrieval. The four rows (from top to bottom) show Top 5 results for: *clock-on-tower*, *person-play-frisbee*, *bird-on-branch* and *boat-in-water*, respectively. Red frames are false positives. The image in last row is tagged with *boat-on-water* rather than *boat-in-water*, but it should be regard as correct.