

Visual relations direct the otter understanding of object interactions, which provide futher semantic information for applications such as captioning and QA

Dataset

Dataset	Image	Object Class	Predicate Class	Unique Relation	Relation/ Object	
VRD[1]	5,000	100	70	6,672	24.25	
VG[2]	99,658	200	100	19,237	57	

[1] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei. Visual relationship detection with language priors. In ECCV, 2016. [2] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. IJCV, 2016

VTransE: Visual Translation Embedding Network for Visual Relation Detection

Hanwang Zhang¹, Zawlin Kyaw², Shih-Fu Chang¹, and Tat-Seng Chua² ¹Columbia University ²National University of Singapore



Codes





Faster R-CNN (%mAP) before and after VTransE
--------------------	----------------------------

VR	D	VG		
Before	After	Before	After	
13.32	13.98	6.21	6.58	

VR			VG						
Relation Det.		Retrieval		Phrase Det.		Relation Det.		Retrieval	
50	R@100	Rr@5	Med r	R@50	R@100	R@50	R@100	Rr@5	Med r
58	1.85	2.82	211	-	_	_	_	-	_
86	14.70	8.75	137	-	-	_	_	Ι	_
.07	15.20	7.89	41	9.46	10.45	5.52	6.04	14.65	7