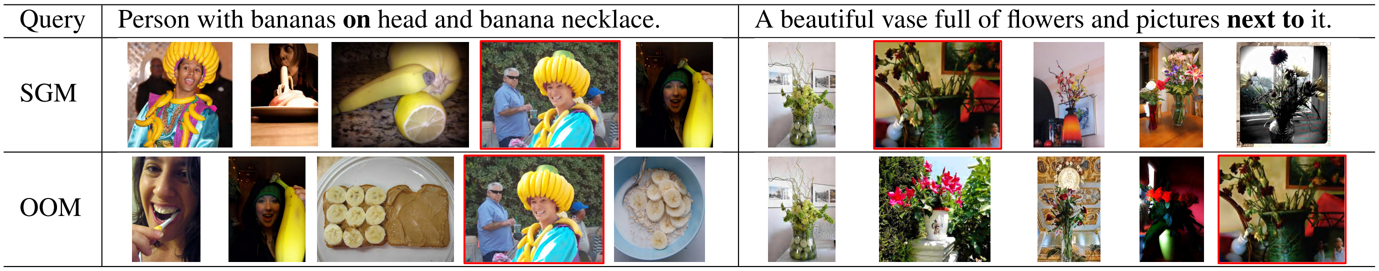
Supplementary Material: Cross-modal Scene Graph Matching for Relationship-aware Image-Text Retrieval

Sijin Wang[1,2], Ruiping Wang[1,2], Ziwei Yao[1,2], Shiguang Shan[1,2], Xilin Chen[1,2]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, 100190, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
{sijin.wang, ziwei.yao}@vipl.ict.ac.cn, {wangruiping, sgshan, xlchen}@ict.ac.cn

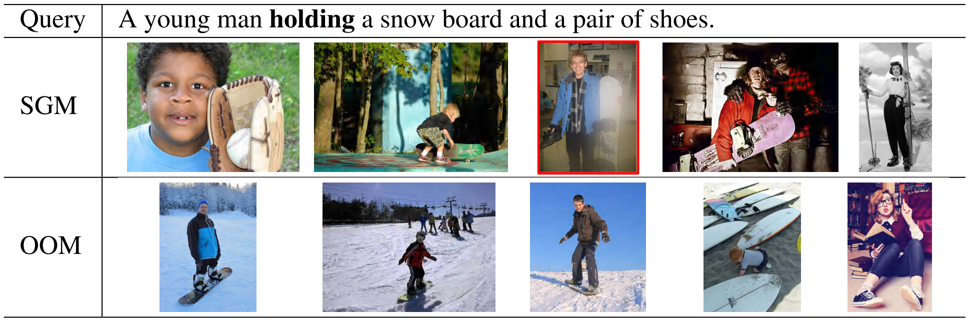
We illustrate more qualitative image retrieval results of SGM vs. OOM on MSCOCO in Fig. 1 and it proves that the relationship-aware matching method is better than the method that only addresses the object-level matching. We show a failure case in Fig. 2. The failure case shows that SGM sometimes focuses too much on relationships, so how to balance the emphasis on objects and relationships will be in our future work. The OOM model also fails in this case, which only retrieves the images with correct objects but wrong relationships.

Then in Fig.3, we show more cases that the SGM has indeed captured the relationships. So when the relationship word in the query is modified, the retrieved results have also changed a lot accordingly.



Query	Person with bananas **on** head and banana necklace.	A beautiful vase full of flowers and pictures **next to** it.
SGM		
OOM		

Figure 1. Qualitative image retrieval results of **SGM vs. OOM** on MSCOCO. Images with red bounding boxes are the ground-truth.



Query	A young man **holding** a snow board and a pair of shoes.
SGM	
OOM	

Figure 2. A failure case of SGM and OOM. Images with red bounding boxes is the ground-truth.

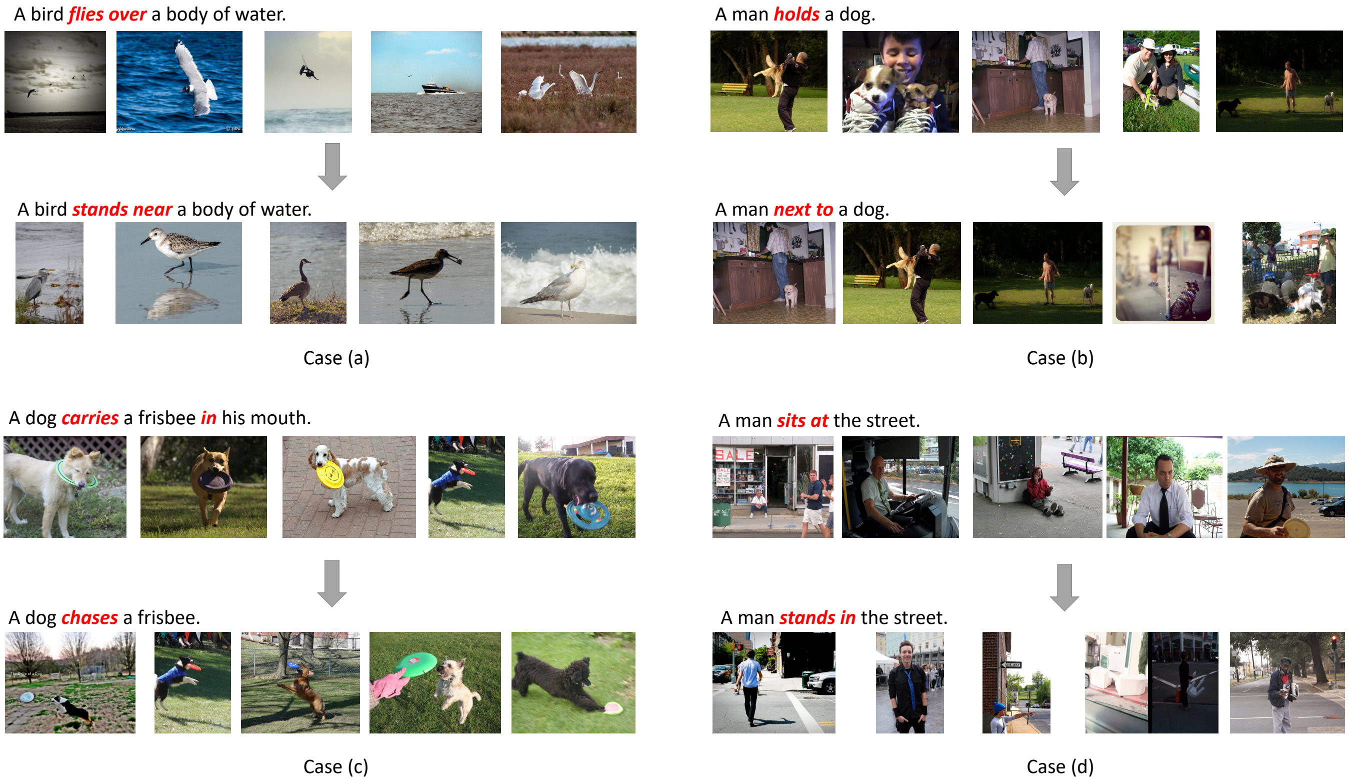

A bird flies over a body of water.
A bird stands near a body of water.
Case (a)
A man holds a dog.
A man next to a dog.
Case (b)
A dog carries a frisbee in his mouth.
A dog chases a frisbee.
Case (c)
A man sits at the street.
A man stands in the street.
Case (d)

Figure 3. Comparison of top-5 retrieved results before and after modifying the relationship words in queries.