Using Scene Graphs for Detecting Visual Relationships

Paper ID: 2794

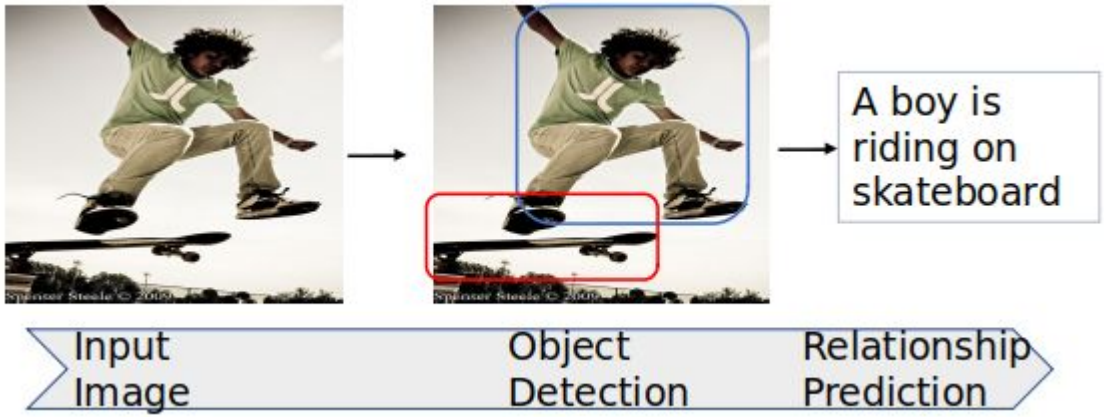
Anurag Tripathi, Siddharth Srivastava, Brejesh Lall, Santanu Chaudhury
Indian Institute of Technology Delhi, India

Objective

The objective is to establish the relationship between the objects in a given image. As we can see in figure 1 this process consist of following steps:-

Identify the object in an image

Establish relationship between the identified objects



Literature Survey

Method	Detector	RoI	Spatial	Language	Jt. Reas.
Iterative Message Passing [13]	VGG(COCO)	Pool	-	-	✓
Multi Level Scene Description Network [14]	VGG(ImageNet)	Pool	-	-	-
Neural Motifs [15]	VGG(ImageNet)	Align	✓	✓	✓
Graph RCNN [7]	VGG(ImageNet)	Align	✓	-	✓
Relationship Detection Network [16]	VGG(COCO)	Align	✓	✓	-
Large Scale Visual Relationship[12]	VGG(COCO)	-	-	✓	-
UVTransE [17]	VGG(ImageNet)	Align	✓	✓	-

Literature Survey

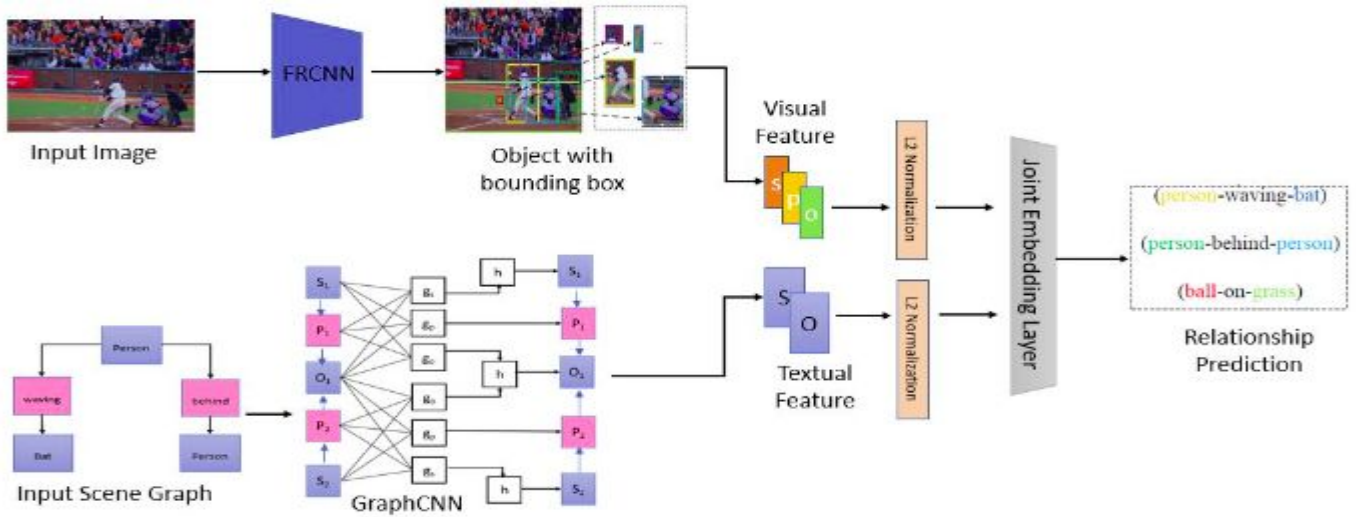
- Authors in [1] [2] use a holistic view where in each and every relation is considered as a class and a detector is learnt for each class. Shortcoming of this approach is that its not scalable and such approaches are suitable for small vocabulary database
- In [3], authors propose a region-object relevance-guided framework that looks for regions related to the labels of object pairs in images.
- In [4], authors proposed a Graph R-CNN framework for localized objects using object models and subsequently establish an edge between all the pairs of localized objects.
- In [5], the authors proposed a framework combining the strengths of deep learning and statistical models

Motivation

- The spatial organization of objects is meaningful i.e. not random
- The closer the objects are, the easier it is to identify the semantic meaning for the interaction between the objects
- The relationships between such objects can be visualized as a scene graph

Approach

- Visual appearance embedding
- Spatial embedding
- Semantic embedding



Dataset and Metrics

- Visual GENOME Dataset
  - It contains 99,658 images with 200 object categories and 100 predicates. There are totally 1,174,692 relation instances among 19,237 unique triplets. The default split contains 73,801 images for training and 25,857 images for testing.
- Metrics:
  - To evaluate the proposed method, we report results on predicate detection and relationship detection. Predicate detection refers to predicting the correct relationship given a subject-object pair, while relationship detection refers to localization of object as well as detection of predicate between the localized object pair. To compare with the prior work, we use Recall@50 and Recall@100 as the evaluation metrics where R@K refers to fraction of true positive relationships out of top K confidence score

Dataset and Metrics

	Predicate Detection				Relationship Detection			
	Recall		ZS_R		Recall		ZS_R	
Method	@50	@100	@50	@100	@50	@100	@50	@100
VTransE [23]	62.63	62.87	-	-	5.52	6.04	-	-
VRDS [24]	69.06	74.37	-	-	8.28	8.28	-	-
MR-Net [11]	65.27	66.45	-	-	**12.64**	**14.32**	-	-
Ours (S *or* O)	77.69	82.05	**23.31**	**37.01**	11.60	12.60	**7.22**	**7.55**
Ours (S *and* O)	**86.56**	**87.48**	18.18	27.27	10.53	11.81	6.06	6.56

Table 1. Summarization of state-of-the-art methods for relationship detection. Last three columns represent utilization of Spatial Features, LanguageFeatures and Joint Reasoning respectively.

References

1. J.-F. Hu, W.-S. Zheng, J. Lai, S. Gong, and T. Xiang,“Recognising human-object interaction via exemplar based modelling,” inProceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3144–3151.
2. A. Prest, C. Schmid, and V. Ferrari, “Weakly supervised learning of interactions between humans and objects,”IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 34, no. 3, pp. 601–614, 2011
3. Y. Goutsu, “Region-object relevance-guided visual relationship detection.” BMVC, 2018.
4. J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, “Graph r-cnn for scene graph generation,” inProceedings of the European Conference on Computer Vision (ECCV), 2018,pp. 670–685
5. B. Dai, Y. Zhang, and D. Lin, “Detecting visual relationships with deep relational networks,” inProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3076–3086.

THANK YOU