ACMMULTIMEDIA 2019
21 - 25 October 2019, Nice, France

P1A-11

MAGUS
MediA recoGnition and UnderStanding

# Hierarchical Visual Relationship Detection

**Xu Sun, Zi Yuan, Tongwei Ren, Jinhui Tang, Gangshan Wu**

## Introduction

**Hierarchical visual relationship detection (HVRD)** encourages predictions with abstract yet compatible relationship triplets when the confidence level of the specific image content is relatively low.

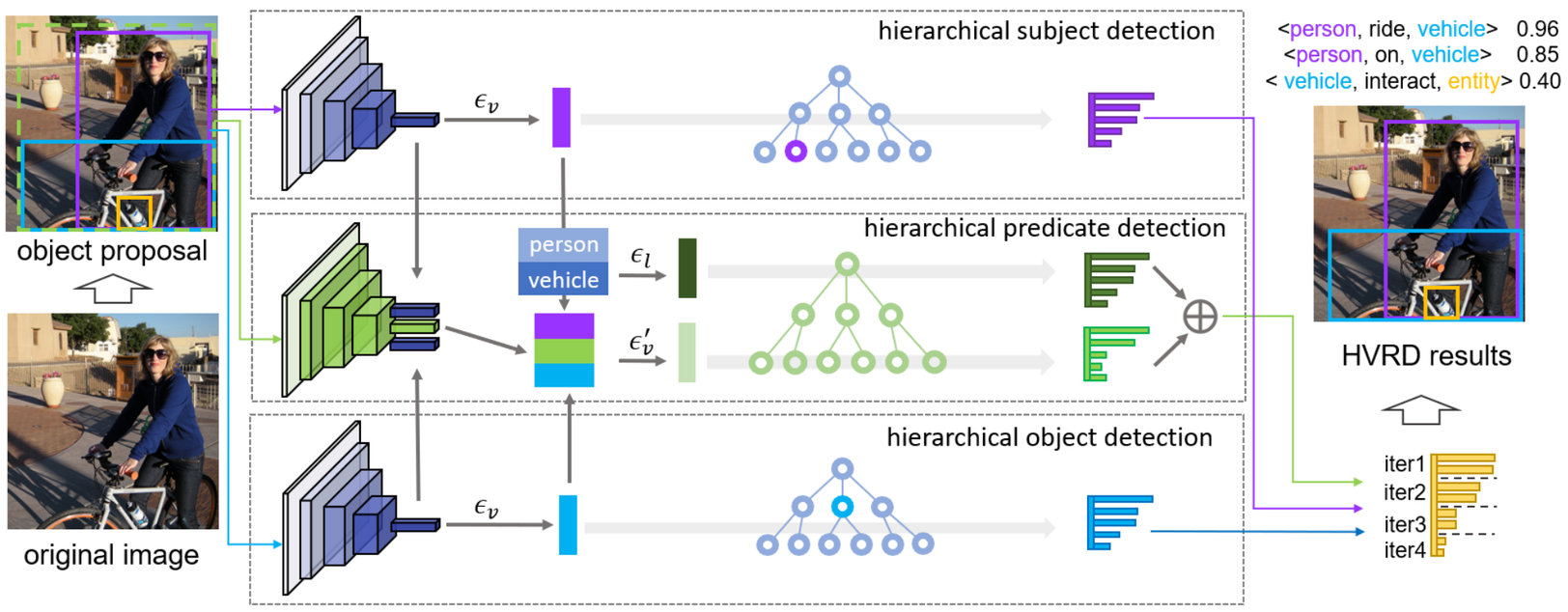
**Contribution**: we **firstly** define the HVRD task, propose a novel HVRD method, and construct two datasets, H-VRD and H-VG, for HVRD evaluation.

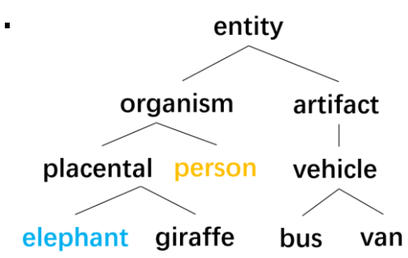

| relationship triplets | VRD | HVRD |
|---|---|---|
| <person, stand by, elephant> | correct | 1.00 |
| <person, near, elephant> | wrong | 0.87 |
| <person, near, animal> | wrong | 0.76 |
| <entity, interact, entity> | wrong | 0.15 |
| <person, watch, elephant> | wrong | 0.00 |

## Method

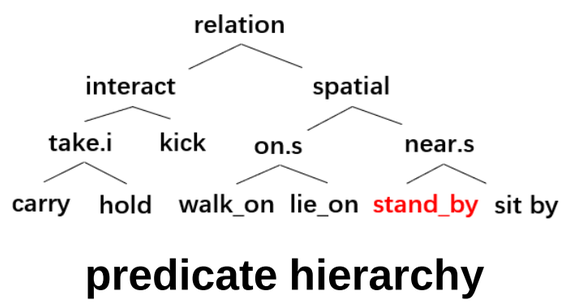
We propose **an order embedding based method** to successfully combine visual information and concept structure knowledge, which consists of three components:

- **Hierarchical concept embedding**: embed concepts in different abstraction levels with order embedding.
- **Hierarchical object detection**: trade off specificity for accuracy with a vision and knowledge joint model.
- **Hierarchical predicate detection**: combining visual feature and context information.




**object hierarchy**



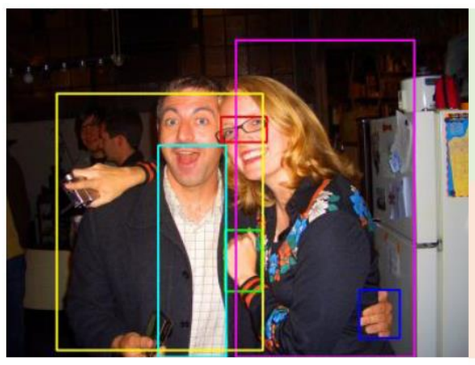

**predicate hierarchy**

## Experiments

### Datasets

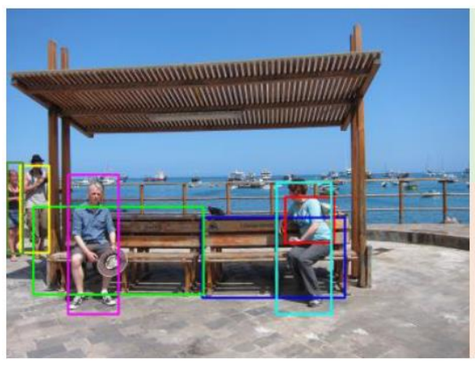
We construct two datasets for HVRD evaluation based on VRD and VG datasets, named **H-VRD** and **H-VG**.

### Qualitative results



<shirt, on, person>
<person, near, person>
<person, has, hand>
<person, wear, glasses>

<shirt, on, person> (1.00)
<person, near, person> (1.00)
<person, has, hand> (1.00)
<person,wear,spectacles> (0.96)



<shirt, on, person>
<person, next to, person>
<person, right of, person>
<person, on, bench>

<shirt, on, person> (1.00)
<person, near, person> (0.92)
<person, near, person> (0.87)
<person, on, seat>(0.96)

**the groundtruth relationship triplets**

**our predicted hierarchical relationship triplets**

### Evaluation criterion

**Recall@N(k=α)** adopts a soft judgment strategy to calculate the hit score $s(g_i)$ of each groundtruth relationship instance $g_i$:

**component hit score**
$$\varphi^S(g,r)=\begin{cases}\frac{d_{r^S}}{d_{g^S}}, & r^S \in T_{g^S},\\ 0, & \text{otherwise},\end{cases},$$

**triplet hit score**
$$\varphi(g,r)=\begin{cases}0, & \varphi^S(g,r)\cdot\varphi^P(g,r)\cdot\varphi^O(g,r)=0\\ \frac{1}{3}(\varphi^S(g,r)+\varphi^P(g,r)+\varphi^O(g,r)), & otherwise\end{cases}$$

**instance hit score**
$$s(g_i)=\max_{r_k\in R_{\alpha_i}}\varphi(g_i,r_k),$$

### Comparison with the state-of-the-arts

| Method | HPD | | | | HVRD | | | |
|---|---|---|---|---|---|---|---|---|
| | HR@50 | HR@100 | BR@50 | BR@100 | HR@50 | HR@100 | BR@50 | BR@100 |
| Lu's | 50.32 | 50.32 | 50.75 | 50.75 | 13.81 | 14.92 | 13.84 | 15.26 |
| VTS | 50.08 | 50.08 | 50.59 | 50.59 | 11.84 | 13.95 | 12.04 | 15.15 |
| DR-net | 53.62 | 53.62 | 54.02 | 54.02 | 14.80 | 16.90 | 14.84 | 17.50 |
| DSR | 54.19 | 54.23 | 54.71 | 54.79 | 14.64 | 16.82 | 14.68 | 17.46 |
| Ours | **60.28** | **60.28** | **66.20** | **66.20** | **15.94** | **18.66** | **17.03** | **19.94** |

| Method | HPD | | | | HVRD | | | |
|---|---|---|---|---|---|---|---|---|
| | HR@50 | HR@100 | BR@50 | BR@100 | HR@50 | HR@100 | BR@50 | BR@100 |
| VTS | 64.44 | 64.66 | 65.24 | 65.47 | 6.19 | 8.17 | 6.21 | 8.63 |
| DSR | 64.27 | 68.56 | 65.12 | 69.47 | 0.31 | 0.57 | 0.32 | 0.57 |
| Ours | **73.89** | **73.99** | **76.11** | **76.25** | **9.40** | **11.29** | **9.77** | **11.74** |

Xu Sun: sunx@smail.nju.edu.cn Tongwei Ren: rentw@nju.edu.cn