P1A-11

Hierarchical Visual Relationship Detection

Xu Sun[1,3], Yuan Zi[1], Tongwei Ren[1,3], Jinhui Tang[2], Gangshan Wu[1]

[1] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[2] School of Computer Science, Nanjing University of Science and Technology, Nanjing, China
[3] Shenzhen Research Institute of Nanjing University, Shenzhen, China

MAGUS
MediA recoGnition
and UnderStanding

Motivation and Solution

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

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

<person, stand by, elephant>

- Our solution
  - **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
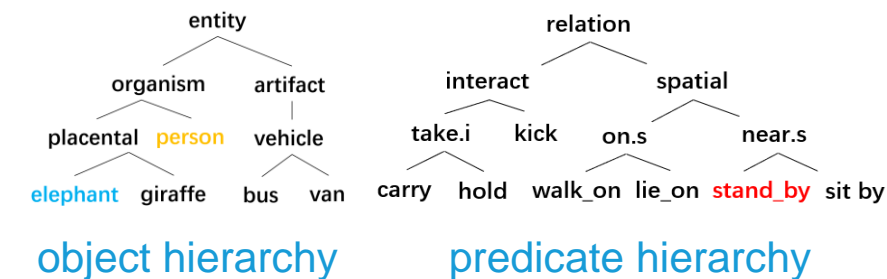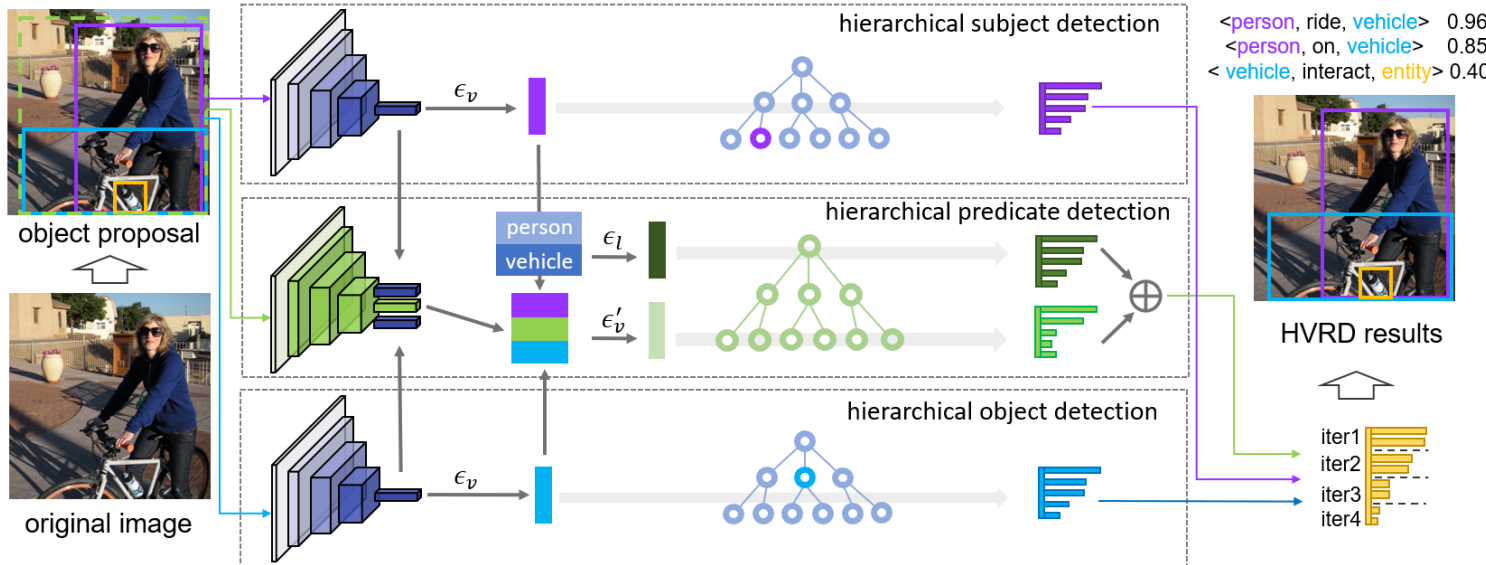
Experiments

- Datasets: **H-VRD** and **H-VG**
  - Construct two datasets for HVRD by extending VRD and VG datasets
- Evaluation criterion: **recall@*N* (*k*=α)**

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

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

- Comparison
  - Task: HPD and HVRD
  - Result: our method is **superior to** the state-of-the-art baselines on **all the criteria**

H-VRD dataset

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

H-VG dataset

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