杭州電子科技大学
HANGZHOU DIANZI UNIVERSITY

On Exploring Undetermined Relationships for Visual Relationship Detection

Yibing Zhan[1], Jun Yu[1], Ting Yu[1], Dacheng Tao[2]

[1] Hangzhou Dianzi University

[2] UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney

Outline

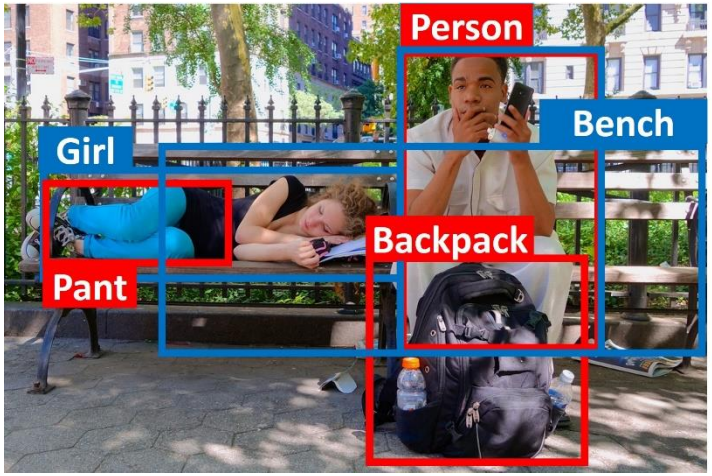
- **1. Visual Relationship Detection**
  - ✓Definition, Challenge, and Solution
- **2. Multi-modal Feature Based Undetermined Relationship Learning Network**
  - ✓Undetermined Relationship Generator
  - ✓Undetermined Relationship Learning Network
- **3. Conclusions**
  - ✓Conclusion and Future work

Visual Relationship Detection

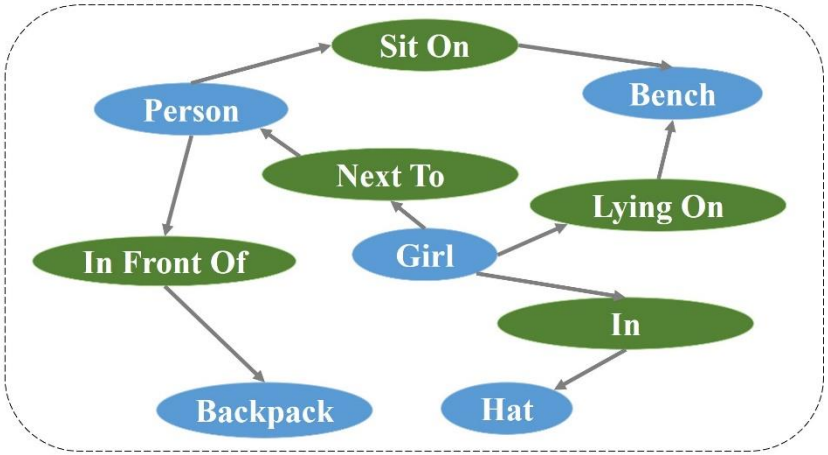
➢ Definition:

- ✓ Detect pairs of correlated objects and predict the object pairs' relationships.
- ✓ Fine-grained image understanding task.

Objects



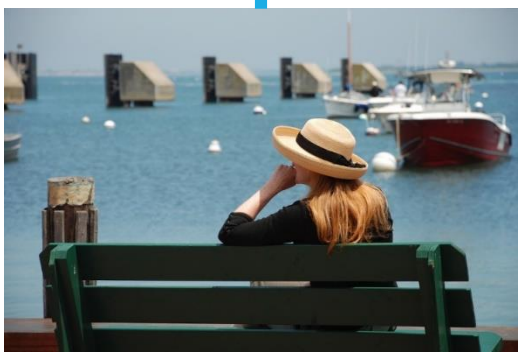
Relationships



Visual Relationship Detection

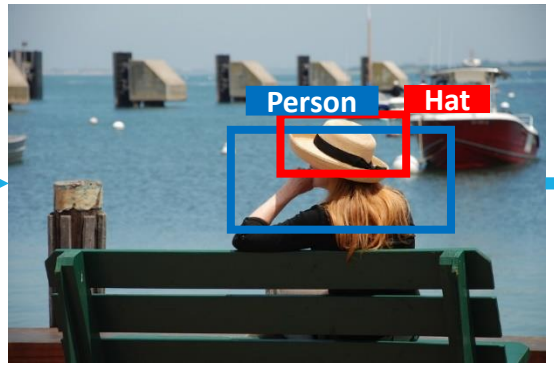
➢ Commonly Strategy:

✓**Relationship detection: Object detection** and **Predicate detection**.



Object Detection

$P(s)\ and\ P(o)$



Predicate Detection

$P(p)$

Person – Wear – Hat

R: Relationship
s: Subject
p: Predicate
o: Object

$$\boldsymbol{P(R) = P(s)P(o)P(p)}$$

Visual Relationship Detection

➢ Challenge:

✓**Hungry for Data**: N object categories and M predicate categories result in N^2M possible relationship category.

VRD Training Dataset:

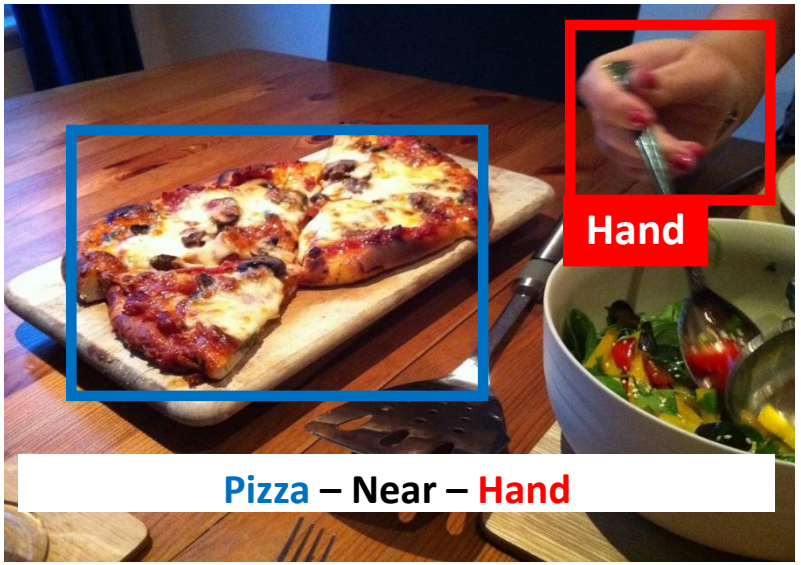
100 object categories

70 predicate categories

37,993 relationships

Zero-shot data:

1169 zero-shot relationships



Visual Relationship Detection

➢ Challenge:

✓**Determinate Detection**: does all detected object pairs contain determinate relationships? K detected objects lead to nearly K^2 potential relationships.

(a): Person-Wear-Hat

Determinate relationships

(b): Person-On-Street

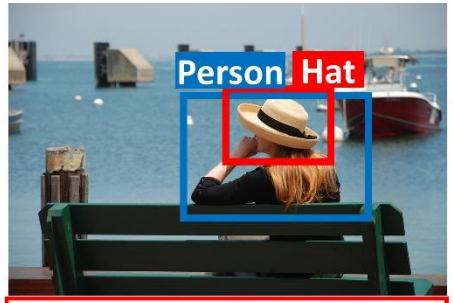
Less significant relationships

(c): Bike-Car

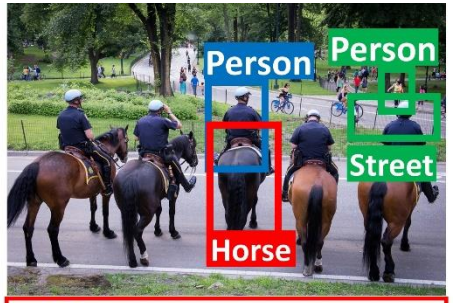
Irrelevant Objects

(d): Monitor-Lamp

Falsely Detected Objects

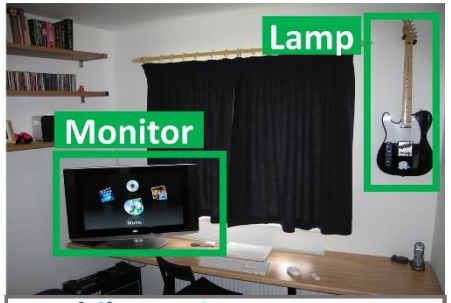


(a) Person – Wear – Hat



(b) Person – On – Horse

(c) Bike – Car
Irrelevant Objects

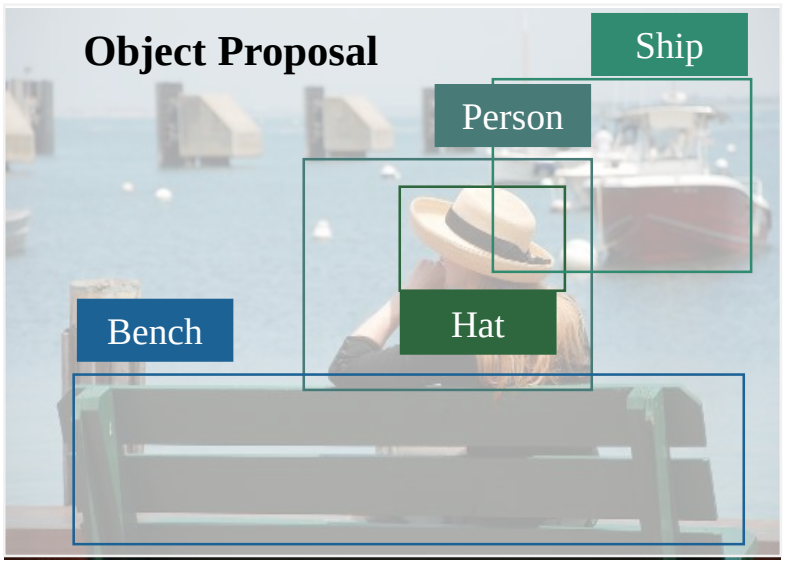


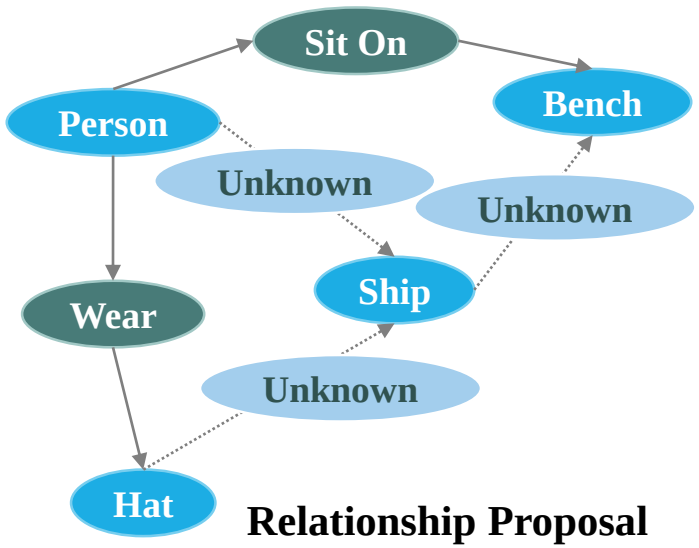
(d) Monitor – Lamp
Falsely Detected Lamp

Visual Relationship Detection

➢ Solution:

- ✓**Undetermined relationships:** unlabeled data; unknown predicate; less noticeable relationships.
- ✓**Determinate relationships:** positive data; determinate predicate; noticeable relationships.

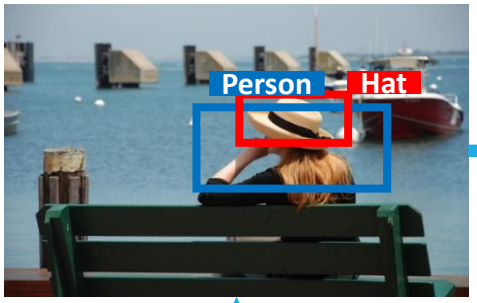




Undetermined Relationship Learning Network

➢ MF-URLN:

✓**Relationship detection: Object detection, Predicate detection**, and **Determinate detection.**



Object Detection

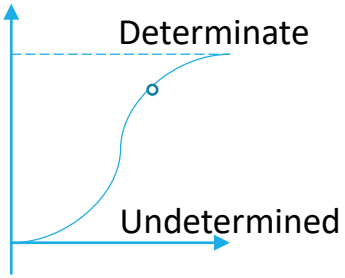
$P(s)$ and $P(o)$

Predicate Detection

$P(p)$

Person – Wear – Hat

$P(d)$



Determinate Detection

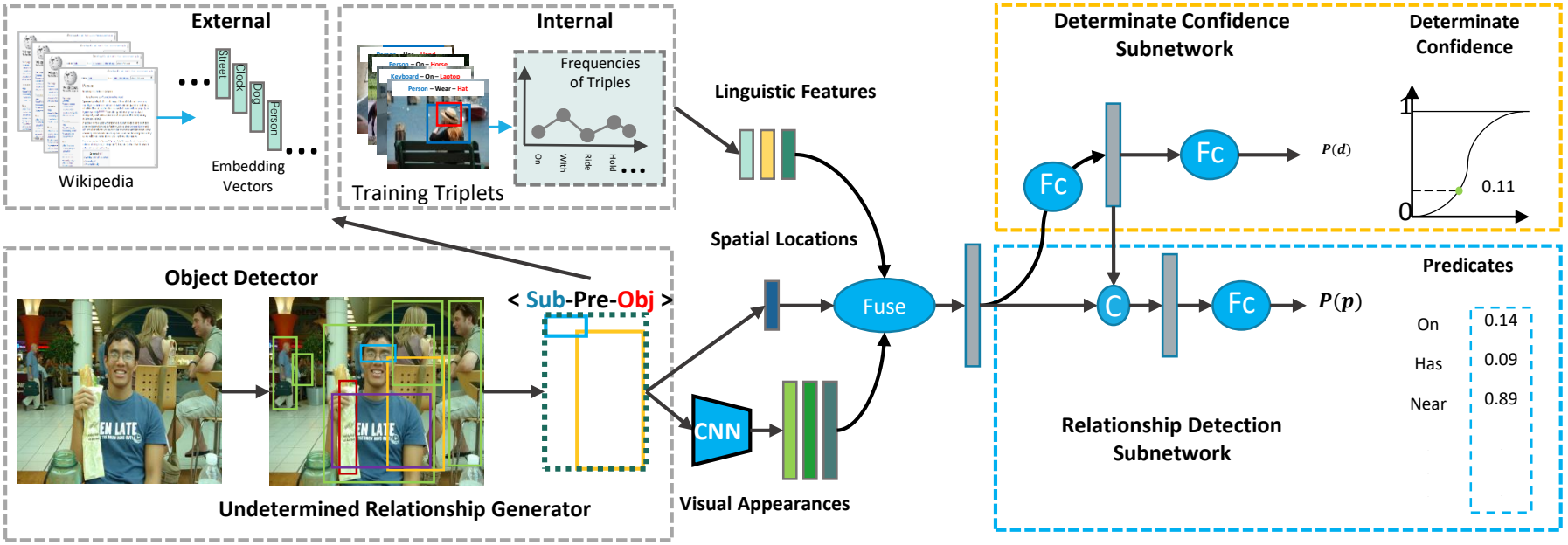
R: Relationship
s: Subject
p: Predicate
o: Object
d: determinate confidence

$$\boldsymbol{P(R) = P(s)P(o)P(p)P(d)}$$

Note*: d* reflects *the* probability of one object pair being manually selected and labeled

Undetermined Relationship Learning Network

➢ Framework:



(a): An object detector to detect objects

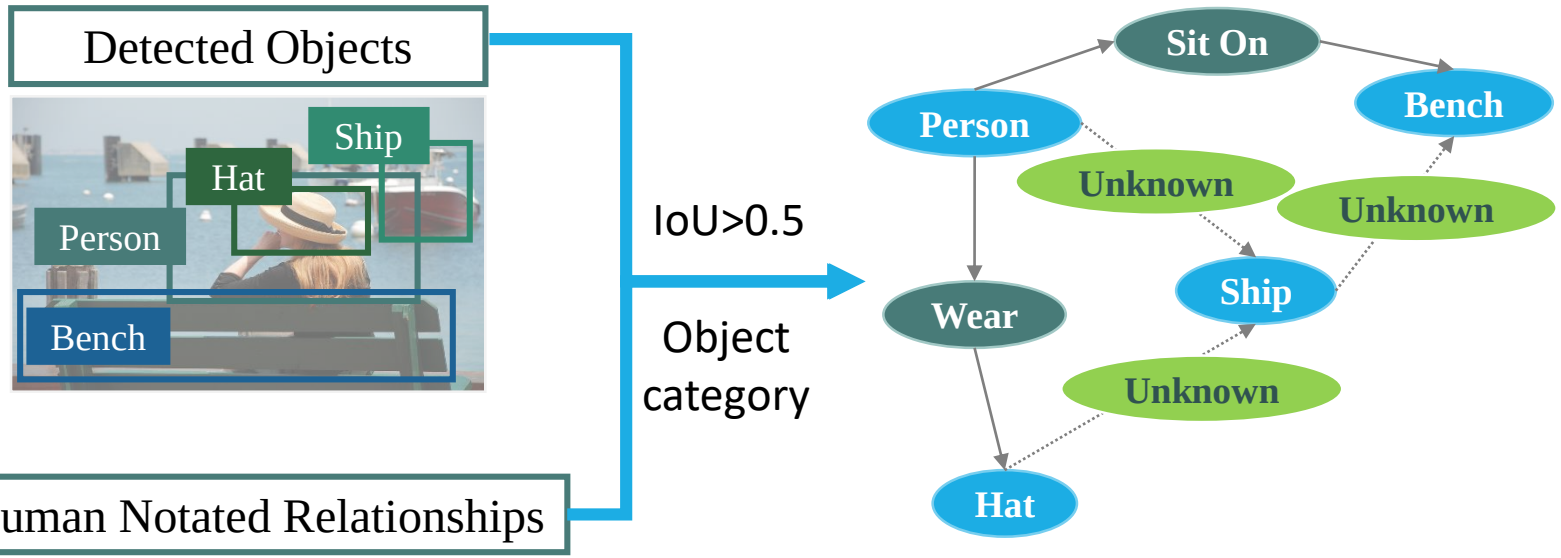
(b): An **undetermined relationship generator** to generate undetermined relationships

(c): An **undetermined relationship learning network** to predict predicates and decides the determinate confidence scores

Undetermined Relationship Learning Network

➢ Undetermined relationship Generator:

✓Detected object pairs are automatically classified into determinate relationships and undetermined relationships.



Undetermined: no human-notated relationships

Determinate: have human-notated relationships

Undetermined Relationship Learning Network

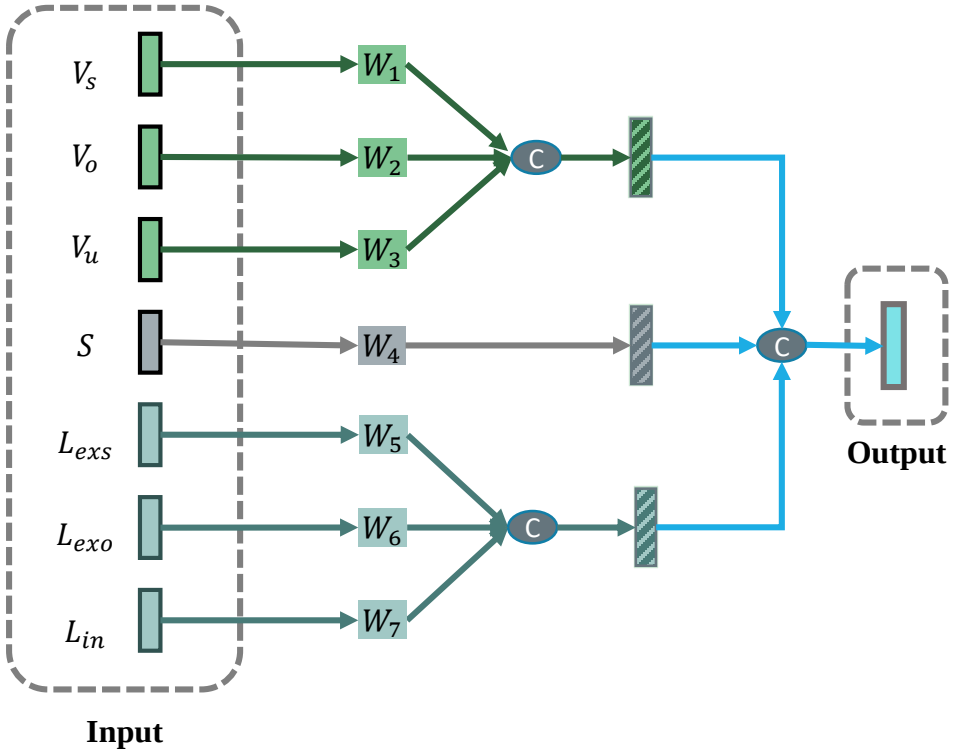
➢ Undetermined Relationship Learning Network:

✓Three types of features from visual, spatial, and linguistic modalities are extracted and fused.

Visual features: V_s, V_o, V_u

Spatial features: S

Linguistic features: L_{exs}, L_{exo}, L_{in}



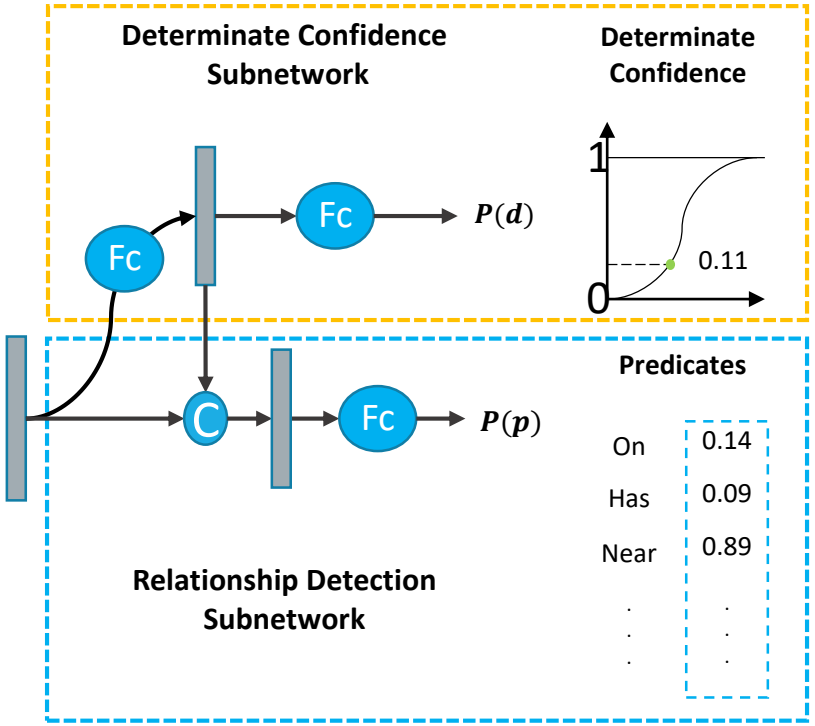
Undetermined Relationship Learning Network

➢ Undetermined Relationship Learning Network:

✓Two subnetworks are proposed to predict predicates and produce determinate confidence scores.

Determinate Confidence Subnetwork produces determinate confidence scores

Relationship Detection Subnetwork predicts predicates



Undetermined Relationship Learning Network

➢ Experimental Results:

✓Performance Comparison

- VG and VRD using Recall

Table 1. Performance comparison of visual relationship detection methods on the VRD dataset. Pre., Phr., and Rel. represent predication detection, phrase detection, and relation detection, respectively. "-" denotes that the result is unavailable.

	Pre.	Phr.		Rel.	
	$R_{50/100}$	R_{50}	R_{100}	R_{50}	R_{100}
VRD-Full [23]	47.9	16.2	17.0	13.9	14.7
VTransE [37]	44.8	19.4	22.4	14.1	15.2
VIP-CNN [20]	-	22.8	27.9	17.3	20.0
Weak-S [26]	52.6	17.9	19.5	15.8	17.1
PPRFCN [38]	47.4	19.6	23.2	14.4	15.7
LKD:S [34]	47.5	19.2	20.0	16.6	17.7
LKD:T [34]	54.1	22.5	23.6	18.6	20.6
LKD:S+T [34]	55.2	23.1	24.0	19.2	21.3
DVSRL [22]	-	21.4	22.6	18.2	20.8
TFR [15]	52.3	17.4	19.1	15.2	16.8
DSL [41]	-	22.7	24.0	17.4	18.3
STA [32]	48.0	-	-	-	-
Zoom-Net [33]	50.7	24.8	28.1	18.9	21.4
CAI+SCA-M [33]	56.0	25.2	28.9	19.5	22.4
VSA [12]	49.2	19.1	21.7	16.0	17.7
MF-URLN	**58.2**	**31.5**	**36.1**	**23.9**	**26.8**

Table 2. Performance comparison of six methods on the VG dataset. "-" denotes that the result is unavailable.

	Pre.		Phr.		Rel.	
	R_{50}	R_{100}	R_{50}	R_{100}	R_{50}	R_{100}
VTransE [37]	62.6	62.9	9.5	10.5	5.5	6.0
PPRFCN [38]	64.2	64.9	10.6	11.1	6.0	6.9
DSL [41]	-	-	13.1	15.6	6.8	8.0
STA [32]	62.7	62.9	-	-	-	-
VSA [12]	64.4	64.5	9.7	10.0	6.0	6.3
MF-URLN	**71.9**	**72.2**	**26.6**	**32.1**	**14.4**	**16.5**

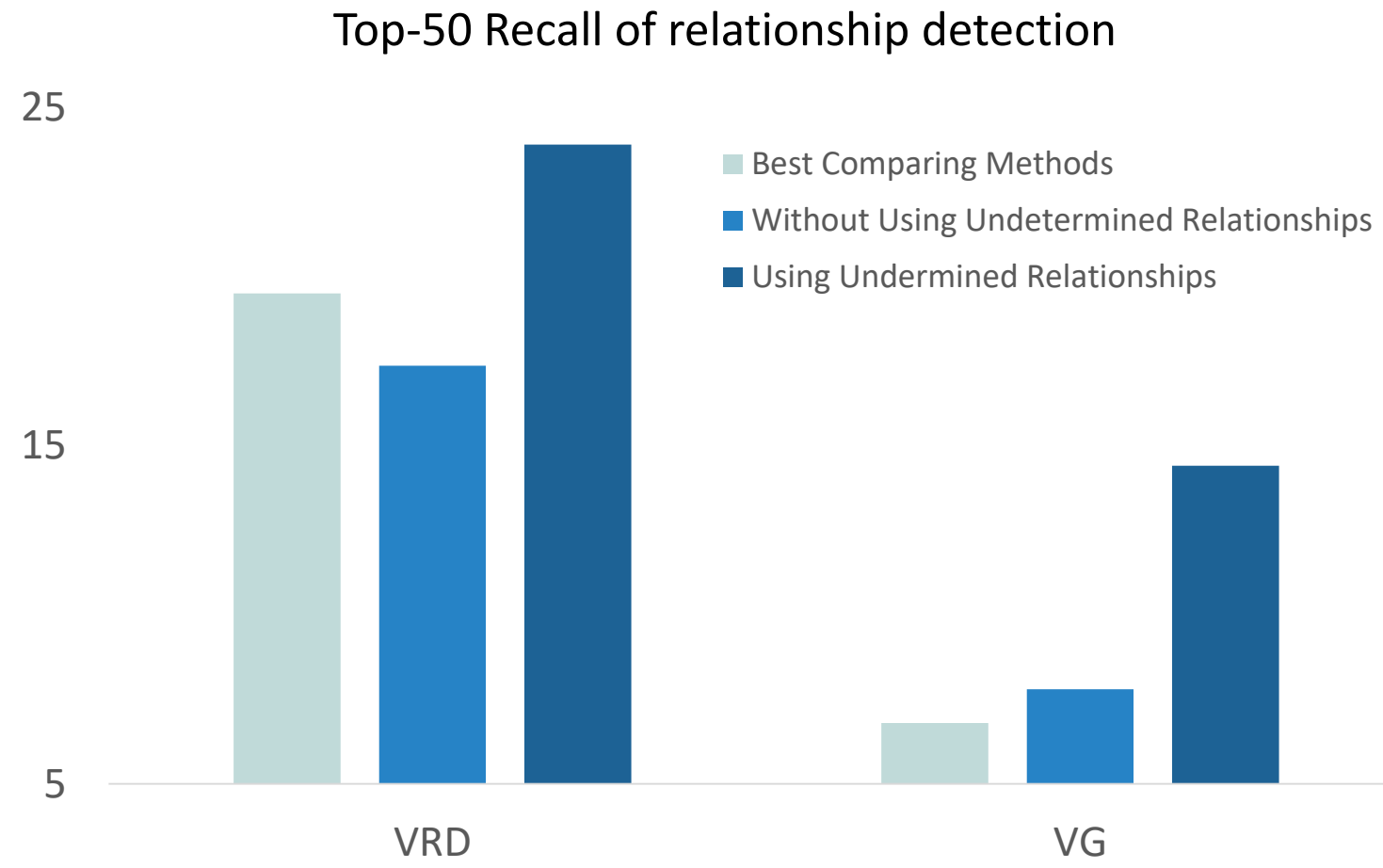
Table 3. Performance comparison on the zero-shot set of the VRD dataset. "-" denotes that the result is unavailable.

	Pre.	Phr.		Rel.	
	$R_{50/100}$	R_{50}	R_{100}	R_{50}	R_{100}
VRD-Full [23]	12.3	5.1	5.7	4.8	5.4
VTransE [37]	-	2.7	3.5	1.7	2.1
Weak-S [26]	21.6	6.8	7.8	6.4	7.4
LKD:S [34]	17.0	**10.4**	**10.9**	**8.9**	**9.1**
LKD:T [34]	8.8	6.5	6.7	6.1	6.4
DVSRL [22]	-	9.2	10.3	7.9	8.5
TFR [15]	17.3	5.8	7.1	5.3	6.5
STA [32]	20.6	-	-	-	-
MF-URLN	26.9	5.9	7.9	4.3	5.5
MF-URLN-IM	**27.2**	6.2	9.2	4.5	6.4

Undetermined Relationship Learning Network

➢ Experimental Results:

✓Ablation Analysis

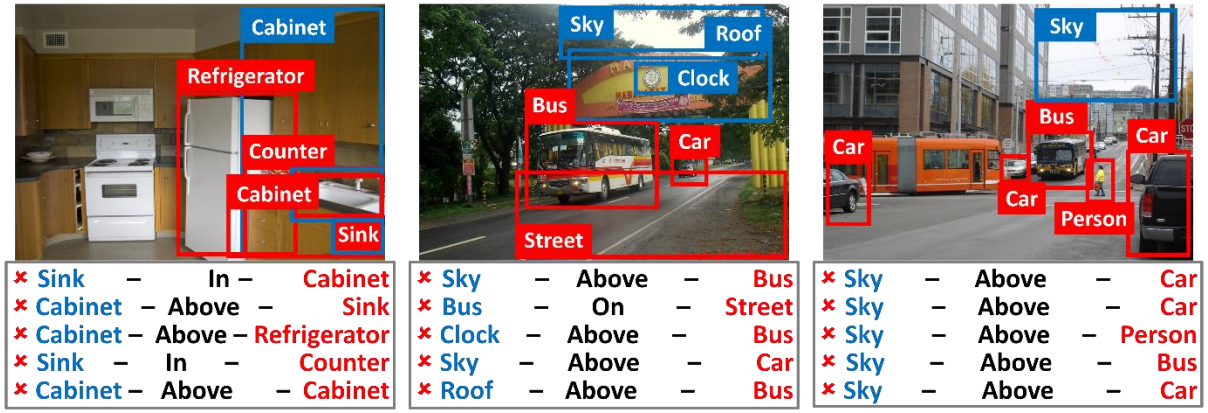


Undetermined Relationship Learning Network

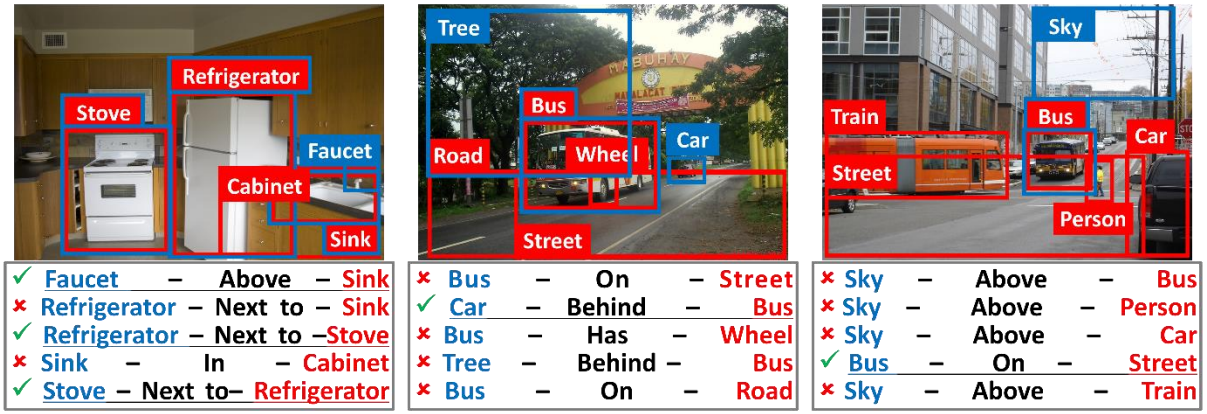
➢ Experimental Results:

✓Qualitative Performance of relationship detection

Top-5 Relationships of MF-URLN without using undetermined relationships



Top-5 Relationships of MF-URLN using undetermined relationships

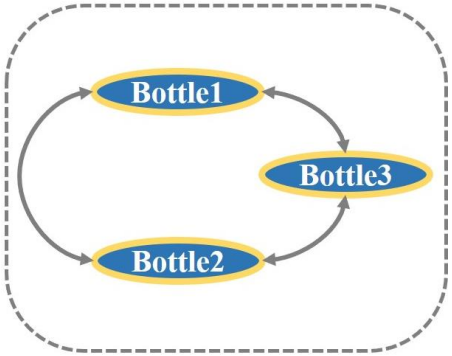


Undetermined Relationship Learning Network

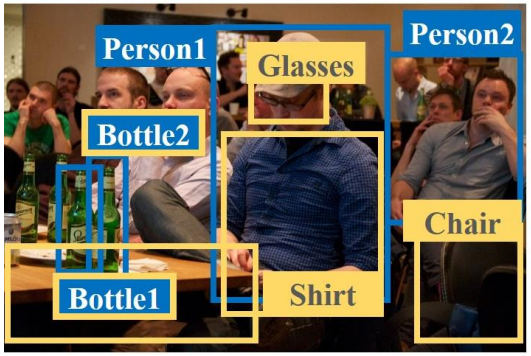
➢ Experimental Results:

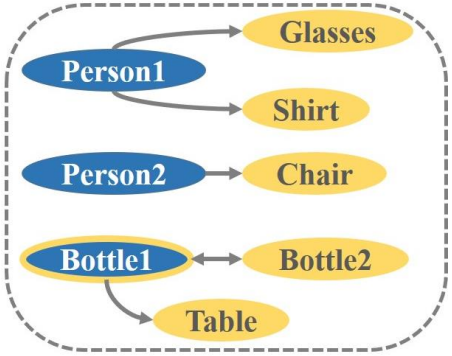
✓Qualitative Performance of determinate detection

The top-6 object pairs of MF-URLN without using undetermined relationships.



The top-6 object pairs of MF-URLN using undetermined relationships.



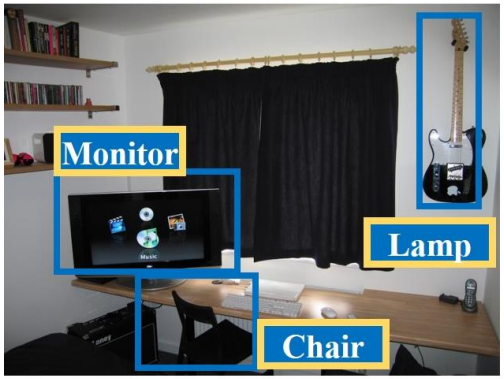


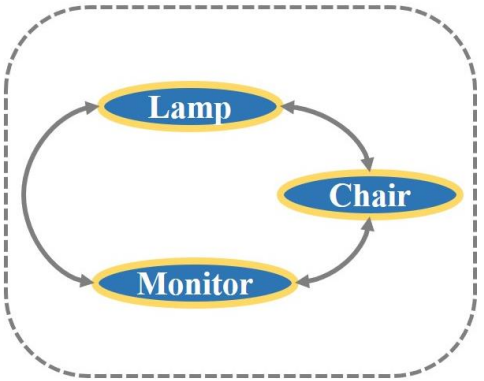
Undetermined Relationship Learning Network

➢ Experimental Results:

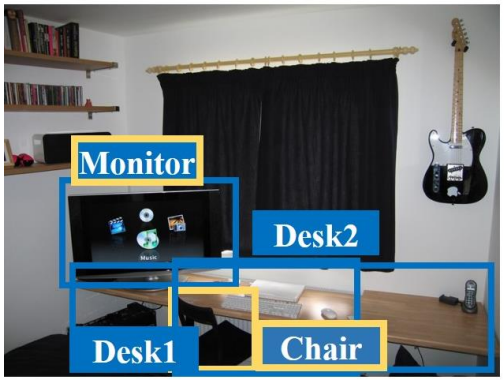
✓Qualitative Performance of determinate detection

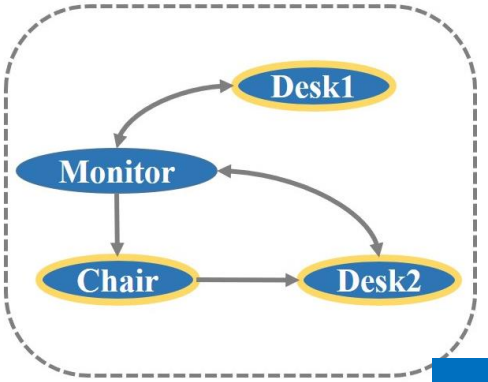
The top-6 object pairs of MF-URLN without using undetermined relationships.





The top-6 object pairs of MF-URLN using undetermined relationships.





Conclusions

➢ Conclusions:

- ✓We use undetermined relationships to improve the visual relationship detection.
- ✓We propose a novel visual relationship detection method, the MF-URLN, by using multi-modal features based on determinate and undetermined relationships.

➢ Future work:

- ➢Better generation and utilization of undetermined relationships.

Thanks for your attention!

➢ Our Paper:

✓On Exploring Undetermined Relationships for Visual Relationship Detection, CVPR 2019.

➢ Our Code:

➢Will be released soon. https://github.com/Atmegal/On-Exploring-Indeterminate-Relationships-for-Visual-Relationship-Detection.