rcp0064

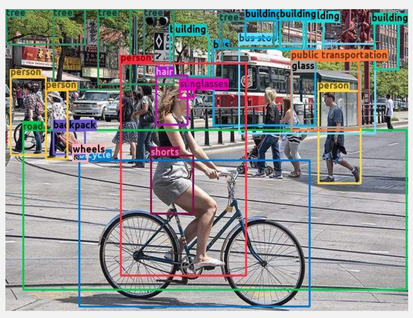
Reproducibility Companion Paper: Visual Relation of Interest Detection

Fan Yu, Haonan Wang, Tongwei Ren*, Jinghui Tang, Gangshan Wu, Jingjing Chen, Zhenzhong Kuang

Original Paper

■ **Visual Relation of Interest Detection (VROID)** a further extension of the traditional Visual Relation Detection (VRD) task to pursue the most semantically important visual relations among all detected ones for describing the main content of an image. We call such a relation **"visual relation of interest" (VROI)**.

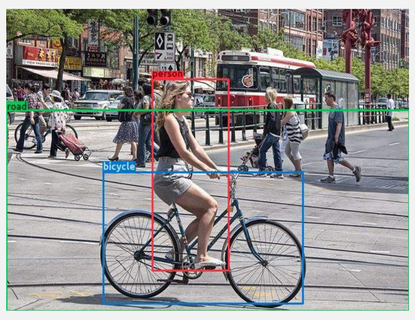
- visual relations



<person, on, road><tree, along, road>
<person, wear, sunglasses><tree, along, road>
<person, cross, road><tree, along, road>
<person, on, bicycle><tree, along, road>
<person, wear, shorts><person, pull, backpack>
<person, with, hair><person, on, road>
<building, on, road><person, cross, road>
<building, on, road><person, on, road>
<building, on, road><person, cross, road>
<building, on, road><person, on, road>
<building, on, road><person, cross, road>
<bus stop, on, road><bus, on, road>
<tree, along, road><bicycle, on, road>
<public transportation, encased in, glass>

<person, ride, bicycle>
<person, ride on, road>
<backpack, with, wheels>
<person, dressed in, shorts>

- **visual relations of interest** and captions

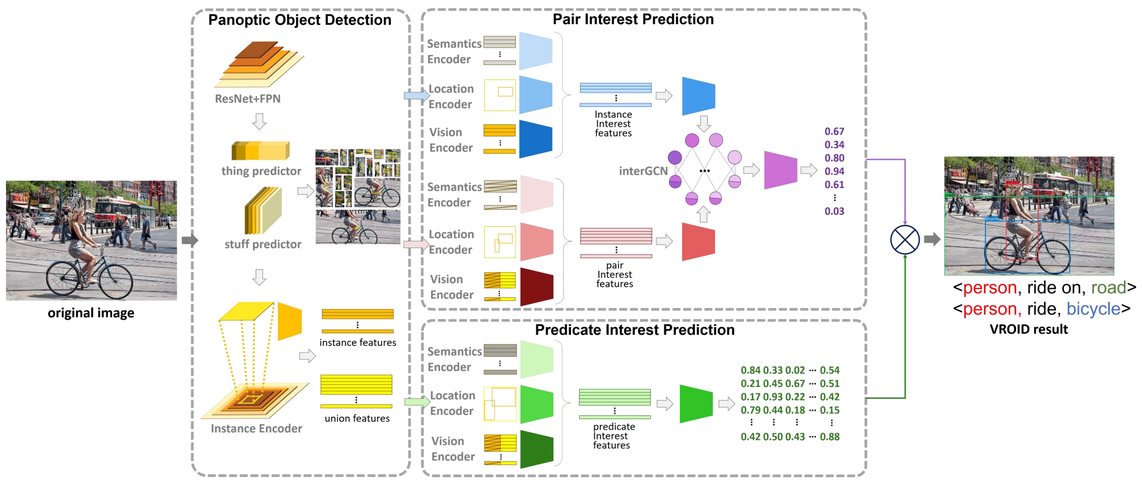


<person, ride, bicycle>
<person, ride on, road>

A woman riding a bike on a somewhat busy road.
A woman rides on a bike down a road.
A girl riding her bike along a busy road.

■ **We propose an Interest Propagation Network (IPNet) for VROID.**

- **Panoptic Object Detection (POD)**: extracts instances
- **Pair Interest Prediction (PaIP)**: predicts the interest score of each instance pair
- **Predicate Interest Prediction (PrIP)**: predicts the interest score of each predicate for each instance pair



■ Experiments show the effectiveness of the method and the components.

- metrics

$$Recall@k = \frac{TP_k}{TP_k + FN_k} \qquad Precision@k = \frac{TP_k}{TP_k + FP_k} \qquad \Psi@k = \frac{TP_k}{TP_k^{max}}$$

- component analysis

Method	R@θ	R@10	P@10	Ψ@10	R@20	P@20	Ψ@20	R@50	P@50	Ψ@50	R@100	P@100	Ψ@100
triplet as output	15.20	23.53	8.55	26.88	31.20	5.67	32.51	42.42	3.08	42.59	51.05	1.85	51.05
output with triplet	20.01	30.18	10.96	34.49	38.44	6.98	40.05	48.93	3.55	49.13	57.05	2.07	57.05
output without pair	0.18	1.62	0.59	1.85	3.47	0.63	3.61	7.89	0.58	8.01	13.38	0.49	13.38
only raw predicate	13.03	22.21	8.07	25.38	30.61	5.56	31.90	41.86	3.04	42.03	50.21	1.82	50.21
no instance	20.14	29.76	10.81	34.01	37.71	6.85	39.30	48.35	3.51	48.55	56.23	2.04	56.23
output with instance	18.37	27.53	10.00	31.46	35.48	6.44	36.97	46.11	3.35	46.30	54.29	1.97	54.29
no semantics features	19.48	29.12	10.58	33.27	37.23	6.76	38.79	47.63	3.46	47.83	55.55	2.02	55.55
no locations features	20.20	29.95	10.88	34.23	38.20	6.94	39.80	48.75	3.54	48.95	57.04	2.07	57.04
bce loss	13.58	20.95	7.61	23.94	27.21	4.94	28.35	36.39	2.64	36.54	43.42	1.58	43.42
Ours	**20.93**	**30.75**	**11.17**	**35.13**	**38.79**	**7.05**	**40.43**	**49.60**	**3.60**	**49.80**	**57.50**	**2.09**	**57.50**

- comparison with baselines

Method	R@θ	R@10	P@10	Ψ@10	R@20	P@20	Ψ@20	R@50	P@50	Ψ@50	R@100	P@100	Ψ@100
STA [38]	4.52	7.71	2.81	8.81	12.08	2.20	12.59	20.02	1.46	20.10	27.03	0.98	27.03
MFURLN [42]	5.73	9.32	3.39	10.65	13.24	2.41	13.79	19.84	1.44	19.93	25.28	0.92	25.28
IMP [35]	3.99	6.32	2.38	7.22	8.87	1.67	9.25	12.46	0.94	12.51	15.56	0.59	15.56
Graph R-CNN [37]	11.34	16.92	6.15	19.33	22.19	4.03	23.12	28.86	2.10	28.98	33.03	1.20	33.03
neural motifs [40]	15.09	21.93	7.97	25.06	27.34	4.97	28.49	33.67	2.45	33.80	37.60	1.37	37.60
VCTree [32]	17.78	25.96	9.43	29.67	32.26	5.86	33.62	40.38	2.93	40.55	46.05	1.67	46.05
VCTree [32]+DSS [8]	17.74	25.93	9.42	29.63	32.23	5.85	33.59	40.38	2.93	40.55	46.05	1.67	46.05
VCTree [32]+NLDF [20]	17.68	25.89	9.41	29.58	32.23	5.85	33.58	40.38	2.93	40.54	46.05	1.67	46.05
ARNet [3]	3.98	-	-	-	-	-	-	-	-	-	-	-	-
MMT [4]	4.94	-	-	-	-	-	-	-	-	-	-	-	-
Frequency	11.25	16.30	5.92	18.62	23.88	4.34	24.89	34.56	2.51	34.71	42.57	1.55	42.57
Ours	**20.93**	**30.75**	**11.17**	**35.13**	**38.79**	**7.05**	**40.43**	**49.60**	**3.60**	**49.80**	**57.50**	**2.09**	**57.50**

Dataset

■ We construct a **ViROI** dataset based on **MSCOCO** and **IOID**.

- 133 object categories (80 things, 43 stuff)
- 249 predicate categories (77 verbs, 17 prepositions, 5 spatial phrases, 150 preposition phrases)
- 12,713 unique VROIs

	training	test	overall
images	25,091	5,029	30,120
VROIs	91,496	18,268	109,764
things per image	6.684	6.634	6.676
stuff per image	4.018	4.029	4.020
VROIs per image	3.647	3.632	3.644

- data format

```
{
  <image_id>:{
    "image_id": int,
    "height": int,
    "width": int,
    "image_name": string,
    "instances": {
      <instance_id>: {
        "instance_id": int,
        "class_id": int,
        "box": [y1, x1, y2, x2],
        "segmentation": encoded dict,
        "labeled": boolean,
        "iscrowd": int
      }
    }
  }
}
```

***_images_dict.json**

```
{
  <image_id>:{
    "image_id": int,
    "height": int,
    "width": int,
    "image_name": string,
    "triplets": {
      <triplet_id>: {
        "id": int,
        "subject_instance_id": int,
        "object_instance_id": int,
        "relation_id": int
      }
    }
  }
}
```

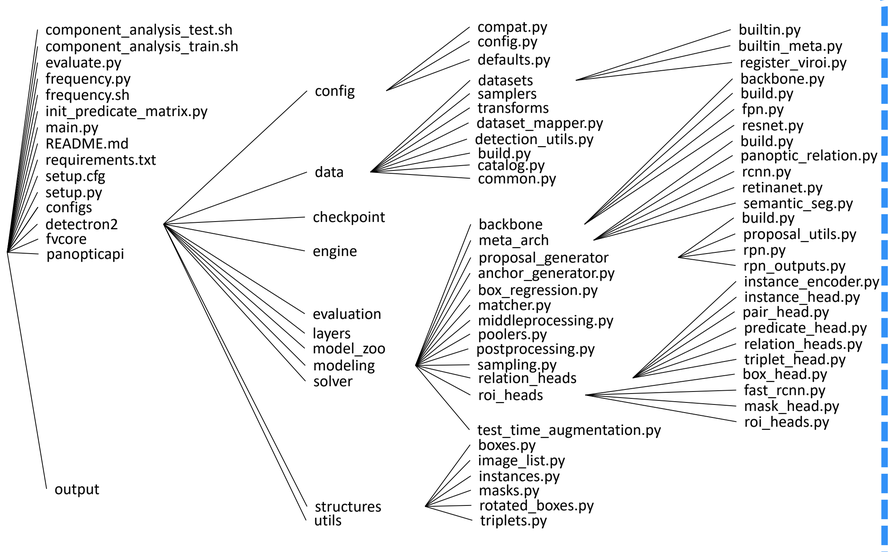
***_triplets_dict.json**

Implementation

■ Environment

- Ubuntu 18.04
- CPU E5-2680 v4
- 64GB memory
- 1TB free space
- GPU 3090
- CUDA 11.1
- cuDNN 8.0
- pytorch 1.7
- Python 3.8.3

VROID

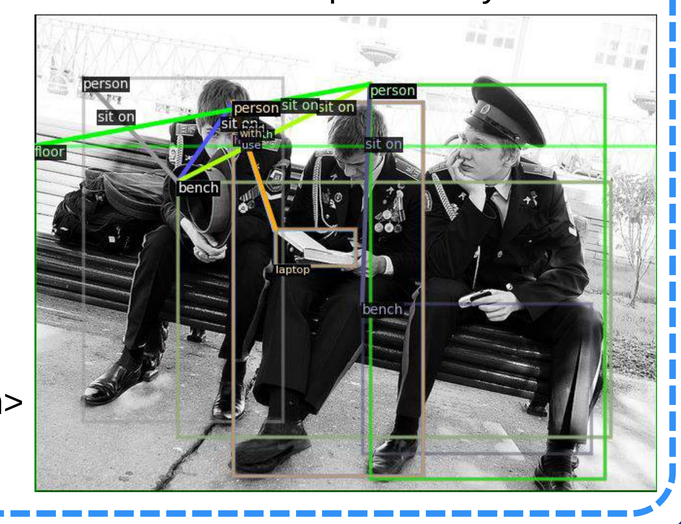


■ Scripts

- train
  python main.py −− config <configuration file path > −−mode train_relation
- predict
  python main.py −−config <configuration file path > −−mode test_relation
- evaluate
  python evaluate.py −−pred_json <output file path > −−top_n <K>

- comparison with other methods
  python main.py −−config configs/VROID/Base−Panoptic−FPN.yaml −−mode test_panoptic
  ./frequency.sh
- component analysis
  ./component_analysis_train.sh
  ./component_analysis_test.sh

- visualization
  python main.py −−config <configuration file path> −−mode demo −−image_path <image path> −−visible −−visible_num <N>



Fan Yu: yf@smail.nju.edu.cn Tongwei Ren: rentw@nju.edu.cn Code&Dataset: https://github.com/njumagus/VROID