Human-Centric Parsing Network for Human-Object Interaction Detection

Guanyu Chen[1], Chong Chen[1] , Zhicheng Zhao[1,2] , and Fei Su[1,2]

[1] Beijing University of Posts and Telecommunications

[2] Beijing Key Laboratory of Network System and Network Culture, Beijing, China

Speaker：Guanyu Chen

BUPT MCPRL

Introduction

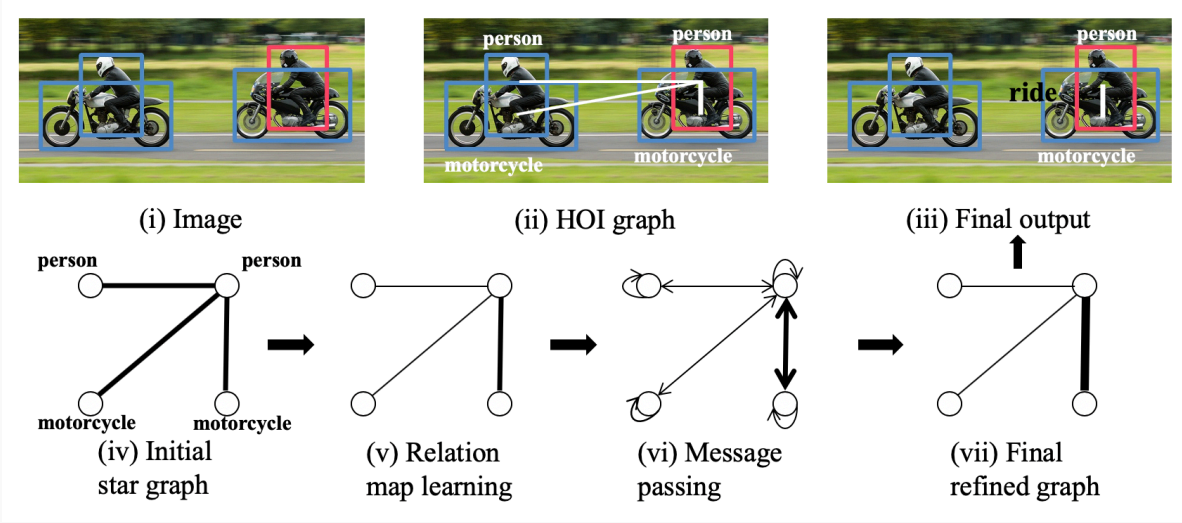
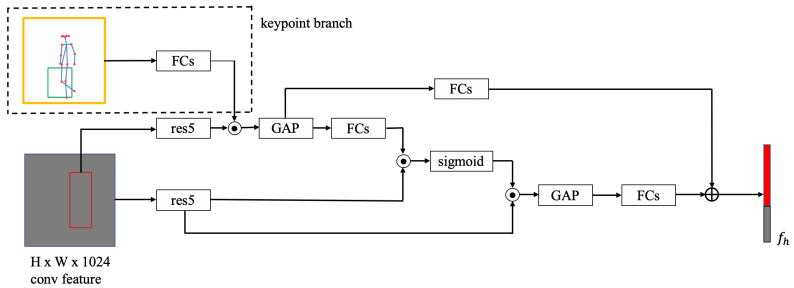


Fig. 1. Illustration of the proposed HCPN.

Human-object interactions detection is an essential task of image inference, but current methods can't efficiently make use of global knowledge in the image. To tackle this challenge, in this paper, we propose a Human-Centric Parsing Network (HCPN), which integrates global structural knowledge to infer human-object interactions. We evaluate our model on V-COCO dataset, and a great improvement is achieved compared with state-of-the-art methods. An overview of HCPN is illustrated in Fig. 1. It mainly includes four parts.

BUPT MCPRL

Graph Initialization



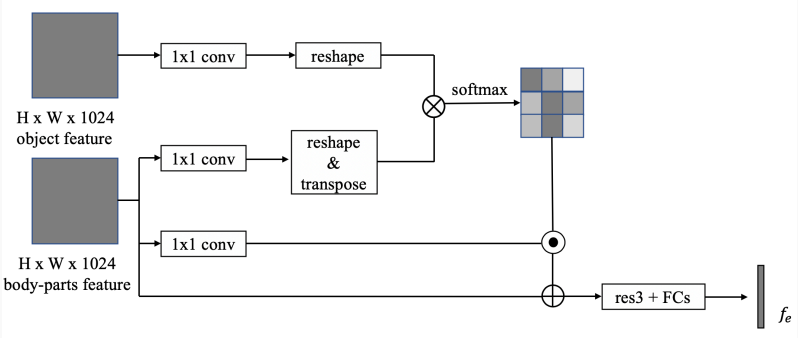


Fig. 2. Feature Generation for nodes and edges.

For a person in the image, we first construct a star graph, by centering the human node and then linking the other object nodes to it. Such kind of structure can avoid the information flow between object nodes, and enhance human feature to the maximum extent. To initialize this star graph, we adopt an attention-based method, such as that developed in iCAN[1], to generate node features. For edge features, we split human keypoints into six body-parts, and use the similarity of objects and body-parts appearance features to selectively integrate body-parts features.

[1] C. Gao, Y. Zou, and J.-B. Huang, "ican: Instance-centric attention network for human-object interaction detection," *arXiv preprint arXiv:1808.10437*, 2018.

BUPT MCPRL

Message Passing Network

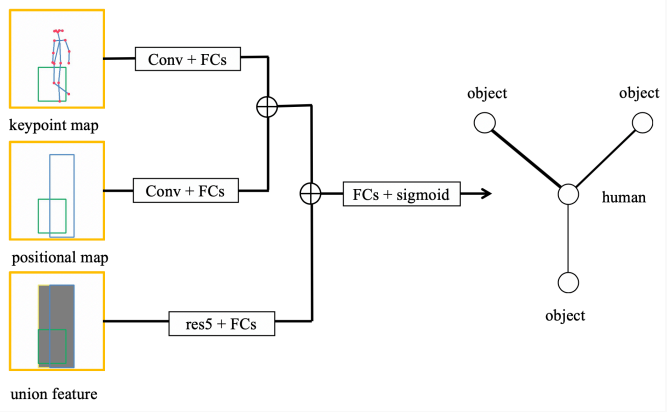


Fig. 3. Link Function.

Then the star graph will go through four parts, namely, the link, message, update and readout functions. The link function will generate a human-object relation map for the star graph, measuring the degree of interaction and calculating the weights of information flow. HCPN can limit message propagation from irrelevant nodes and focus on related nodes with such a relation graph structure. Similar to the message passing neural network (MPNN) [2], the message and update functions iteratively update node features with messages coming from other nodes. Lastly, the readout function computes the multi-class label for each node.

[2] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proceedings of the 34th International Conference on Machine Learning-Volume70*. JMLR. org, 2017, pp. 1263–1272.

BUPT MCPRL

Experiments

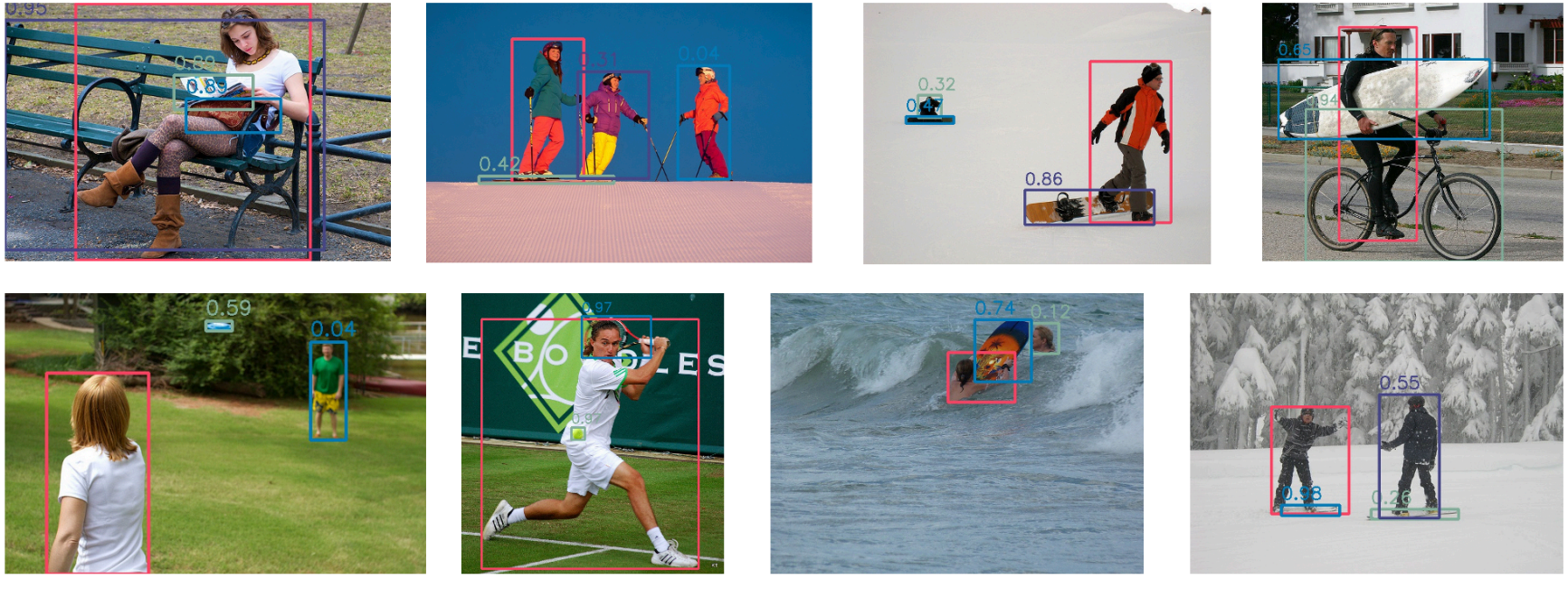


Fig. 4. HCPN is able to pay more attention to the related objects. The red boxes in the images are central human nodes, and other boxes with score denote object nodes and their connectivity with the human nodes.

BUPT MCPRL

Experiments

Our contributions can be summarized in three-fold: First, a new human-centric parsing network (HCPN) is proposed to represent Human-object interactions, in which global structural knowledge is integrated and the information redundancy is avoided effectively. Second, we propose to feed human key- point information into the graph structure to enhance network discriminability. Third, the experimental results show that our HCPN outperforms state-of-the-art methods by the star graph structure inference and message passing.

TABLE I
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS.

Methods	Feature Backbone	mAP_{role}
Gupta et al.	ResNet-50-FPN	31.8
InteractNet	ResNet-50-FPN	40.0
GPNN	Deformable ConvNets	44.0
iCAN	ResNet-50	45.3
RPNN	ResNet-50	47.53
HCPN(ours)	ResNet-50	**47.72**

TABLE II
ABLATION ON V-COCO DATASET.

Methods	mAP_{role} Scenario1	mAP_{role} Scenario2
HCPN w/o message passing	31.52	39.52
HCPN w/o relation map	35.05	43.94
HCPN w/o body-parts attention	33.45	41.16
Node feature using [3]	35.28	43.71
Body-parts spatial relation map	38.10	47.1
Message passing using [3]	38.32	47.23
HCPN	**38.62**	**47.72**

[3] S. Qi, W. Wang, B. Jia, J. Shen, and S.-C. Zhu, "Learning human-object interactions by graph parsing neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 401–417.

BUPT MCPRL

Thanks!

BUPT MCPRL