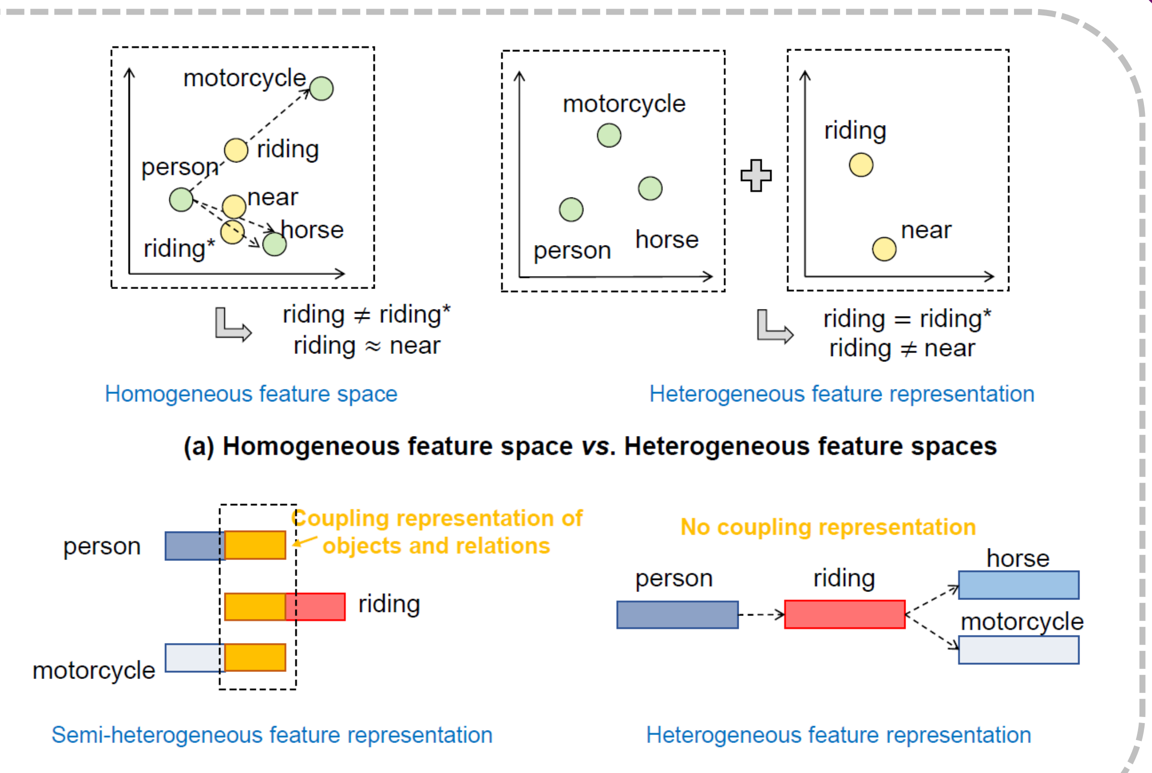
Heterogeneous Learning for Scene Graph Generation

Yunqing He, Tongwei Ren, Jinhui Tang, Gangshan Wu

Introduction

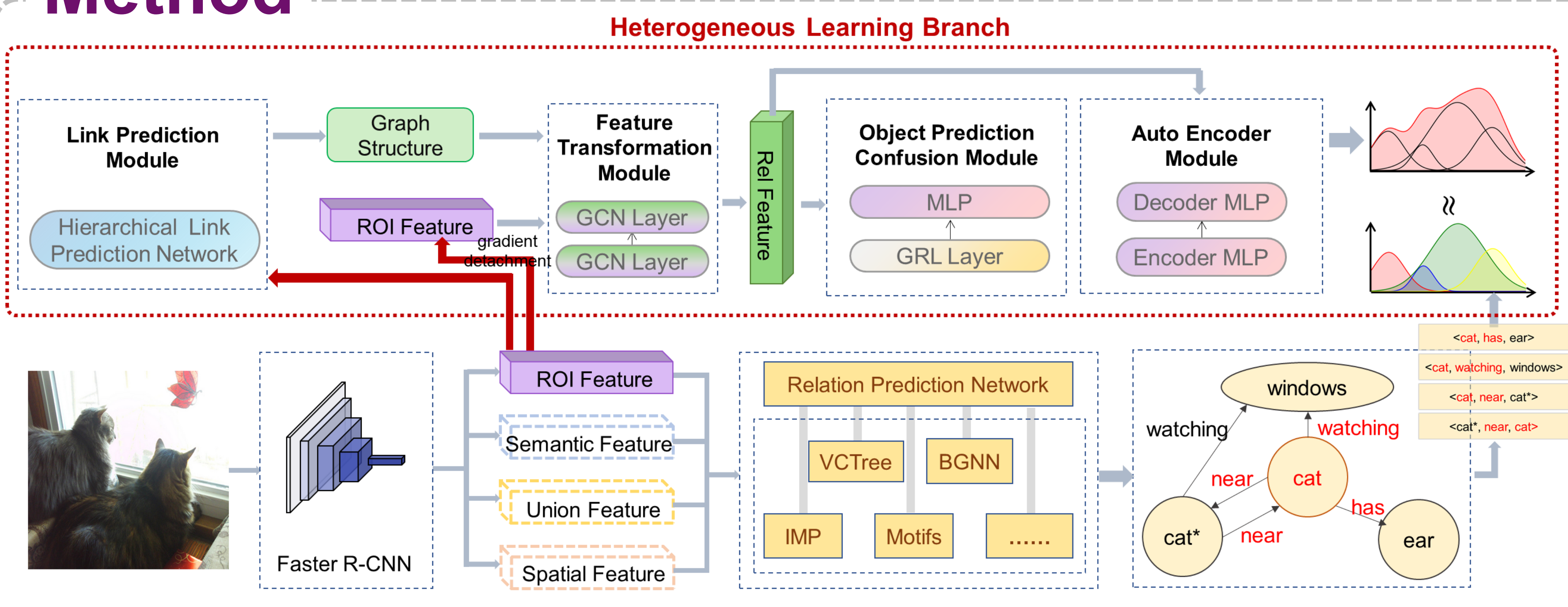
Scene Graph Generation (SGG) aims to construct a graph to express objects and their relationships. We propose a plug-and-play Heterogeneous Learning Branch (HLB) to enhance the independent representation capability of relation features and relieve the problem of large intra-class variation and inter-class ambiguity. We apply HLB to typical SGG methods and. The experimental results on VG-150 dataset demonstrate that HLB significantly improves the performance of all these methods in the common evaluation criteria.



(a) Homogeneous feature space *vs.* Heterogeneous feature spaces

(b) Semi-heterogeneous feature representation *vs.* Heterogeneous feature representation

Method



We propose a heterogeneous learning branch which add heterogeneity constraints between object and relation feature spaces. It consists of four components:

(1) feature transformation module,
(2) link prediction module,
(3) object prediction confusion module,
(4) auto encoder module.

Experiments

SGG tasks and evaluation criteria

R@K: the Recall performance of the top-K predicted results,
mR@K: the mean results of R@K over classes.

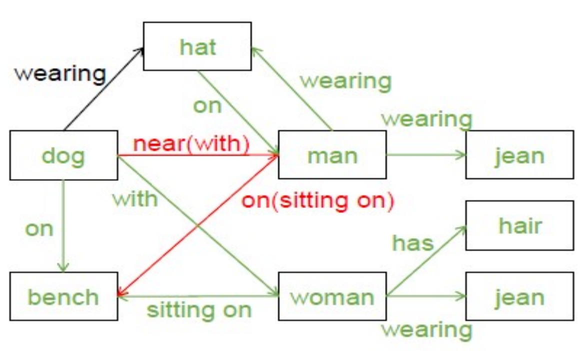
Predicate Classification (PredCls): Given image and objects' locations and labels, construct scene graph,
Scene Graph Classification (SGCls): Given image and objects' locations, construct scene graph,
Scene Graph Detection (SGDet): Only image is given.

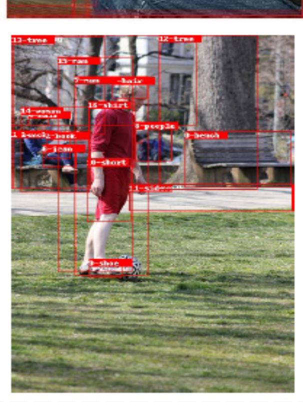
Comparison with the state-of-the-arts

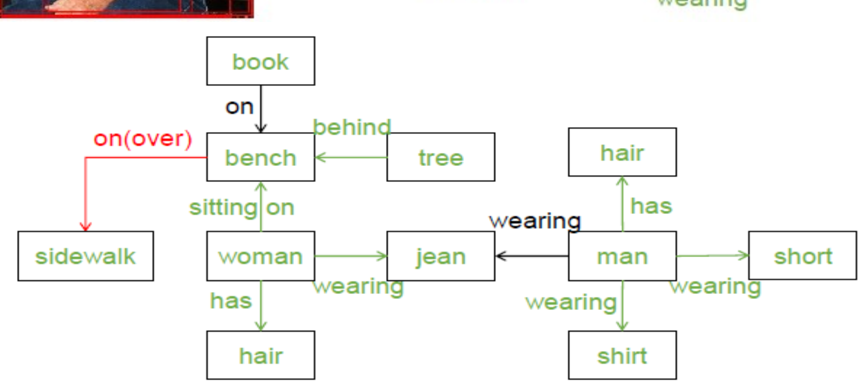
	PredCls				SGCls				SGDet			
Model	mR@50	R@50	mR@100	R@100	mR@50	R@50	mR@100	R@100	mR@50	R@50	mR@100	R@100
GBNet-β [34]	22.1	66.6	24.0	68.2	12.7	37.3	13.4	38.0	7.1	26.3	8.5	29.9
Graph R-CNN [32]	16.4	54.2	17.2	59.1	9.0	29.6	9.5	31.6	5.8	11.4	6.6	13.7
RelDN [38]	15.8	68.7	17.2	68.8	9.3	38.9	9.6	38.9	6.0	31.0	7.3	36.7
FCSGG [17]	6.3	41.0	7.1	45.0	3.7	23.5	4.1	25.7	3.6	21.3	4.2	25.1
GPS-Net [31]	19.2	69.7	21.4	69.7	11.7	42.3	12.5	42.3	7.4	28.9	9.5	33.2
IMP [29]	9.8	59.3	10.5	61.3	5.8	34.6	6.0	35.4	3.8	20.7	4.8	24.5
IMP+HLB	**10.63**	**60.91**	**11.37**	**62.95**	**6.62**	**38.10**	**6.98**	**39.01**	**4.19**	**26.67**	**5.23**	**31.85**
IMP-H	10.17	58.89	10.97	61.31	6.05	34.89	6.47	36.59	5.37	31.21	6.30	35.36
IMP-H+HLB	**10.44**	**59.43**	**11.17**	**61.52**	**7.07**	**38.21**	**7.47**	**39.09**	**5.87**	**31.79**	**6.84**	**35.91**
VTransE [37]	14.7	65.7	15.8	67.6	8.2	38.6	8.7	39.4	5.0	29.7	6.0	34.3
VTransE+HLB	**15.26**	65.68	**16.40**	67.60	8.24	**39.72**	8.74	**40.61**	**5.14**	29.74	**6.22**	**34.47**
KERN [3]	**17.7**	**65.8**	**19.2**	**67.6**	**9.4**	36.7	**10.0**	37.4	6.4	27.1	7.3	29.8
KERN+HLB	15.89	61.17	17.15	64.17	9.01	**38.16**	9.69	**39.37**	**7.11**	**28.70**	**8.58**	**33.41**
MOTIFS [36]	14.0	**65.2**	15.3	**67.1**	7.7	35.8	8.2	36.5	5.7	27.2	6.6	30.3
MOTIFS+HLB	**15.39**	64.91	**16.74**	66.80	**8.90**	**39.48**	**9.44**	**40.32**	**7.19**	**32.57**	**8.43**	**37.01**
VCTree-SL [24]	17.0	**66.2**	18.5	**67.9**	9.8	**37.9**	10.5	**38.6**	6.7	27.7	7.7	31.1
VCTree-SL+HLB	**17.47**	65.73	**18.79**	67.35	**11.98**	36.95	**12.73**	38.50	**7.46**	**32.04**	**8.75**	**36.34**
BGNN [13]	**30.4**	59.2	**32.9**	61.3	14.3	**37.4**	16.5	**38.5**	10.7	**31.0**	12.6	**35.8**
BGNN+HLB	28.20	**61.06**	30.43	**63.22**	**16.72**	35.27	**18.09**	36.64	**12.57**	27.80	**15.03**	32.28
On Average	+1.45%	-0.21%	+0.96%	-0.02%	+11.73%	+4.08%	+10.74%	+4.39%	+12.63%	+8.83%	+14.20%	+10.48%
	+0.54%				+7.73%				+11.53%			

We extend seven typical SGG methods with distinct characteristics to validate the effectiveness of HLB approach, including 4 semi-heterogeneous methods, 2 homogeneous methods and 1 de-homogenized method. The experimental results evidently show that HLB can significantly improve the performance of the existing SGG methods, especially on the SGDet task.

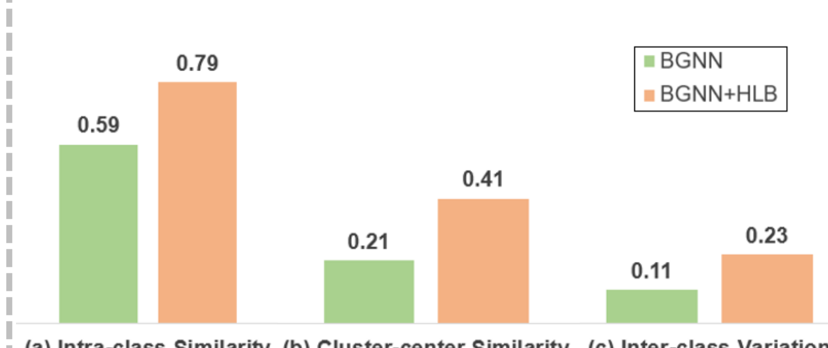
Qualitative results







The words marked with green denote the correctly detected objects and relations, the red words and lines represent the wrongly predicted ones with notated labels in brackets, and the words marked with black color refer to the predicted relations which are considered positive but unlabeled.



南京大学 NANJING UNIVERSITY
南京理工大学 NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY
#2754
heyq@smail.nju.edu.cn
MAGUS MediA recoGnition and UnderStanding