RGB-D Scene Recognition with Object-to-Object Relation

Xinhang Song[1,2], Chengpeng Chen[1,2], Shuqiang Jiang[1,2]

[1]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, CAS, China

[2]University of Chinese Academy of Sciences, China

xinhang.song@vipl.ict.ac.cn

Introduction

➢Observations

- ☐ Objects are helpful to recognize scenes
- ☐ Object co-occurrences may confuse the scene recognition
- ☐ RGB-D data is helpful to capture the spatial information

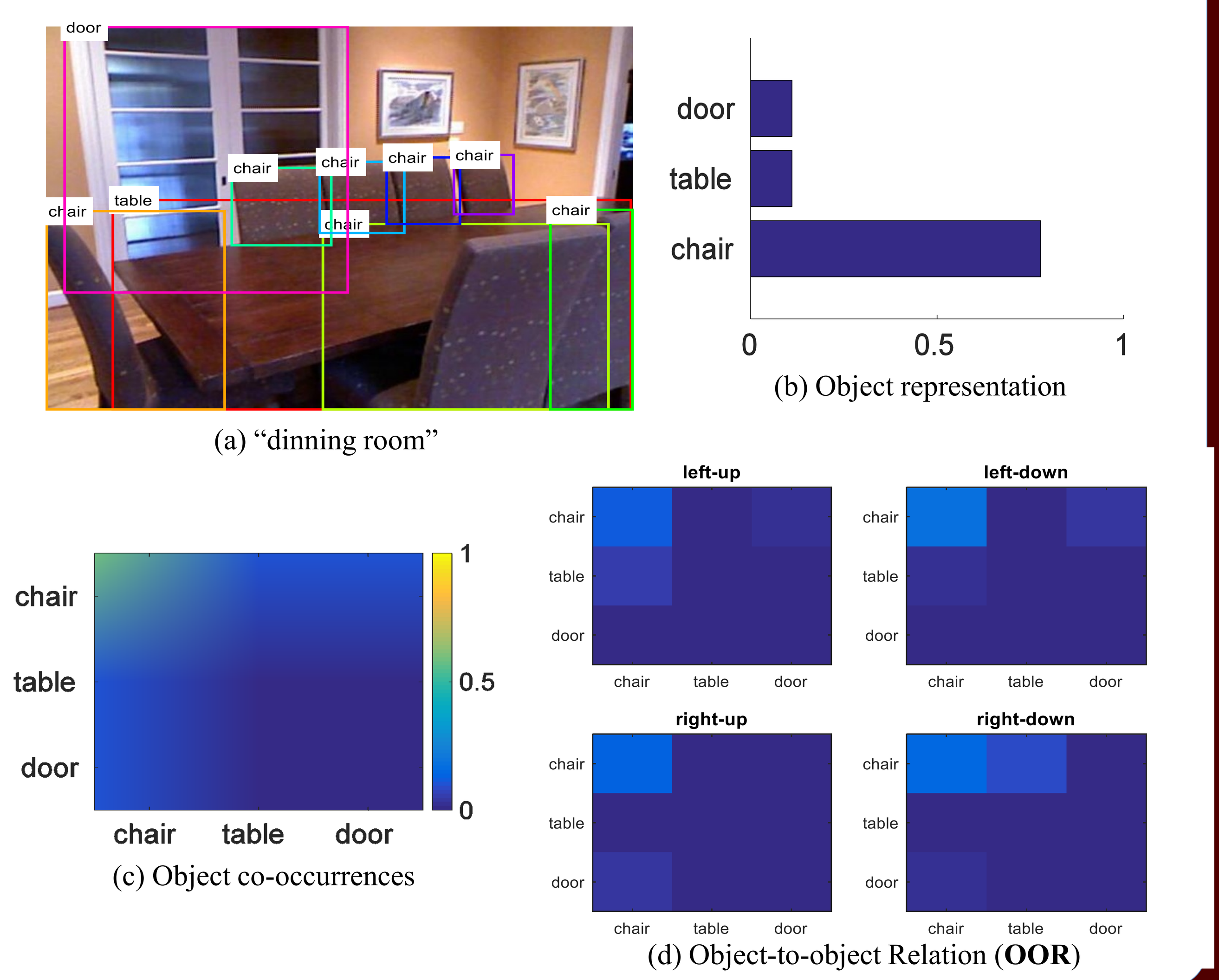
➢Motivation

- ☐ Improving scene recognition with spatial information, i.e., object-to-object relations (OOR)

➢Contributions

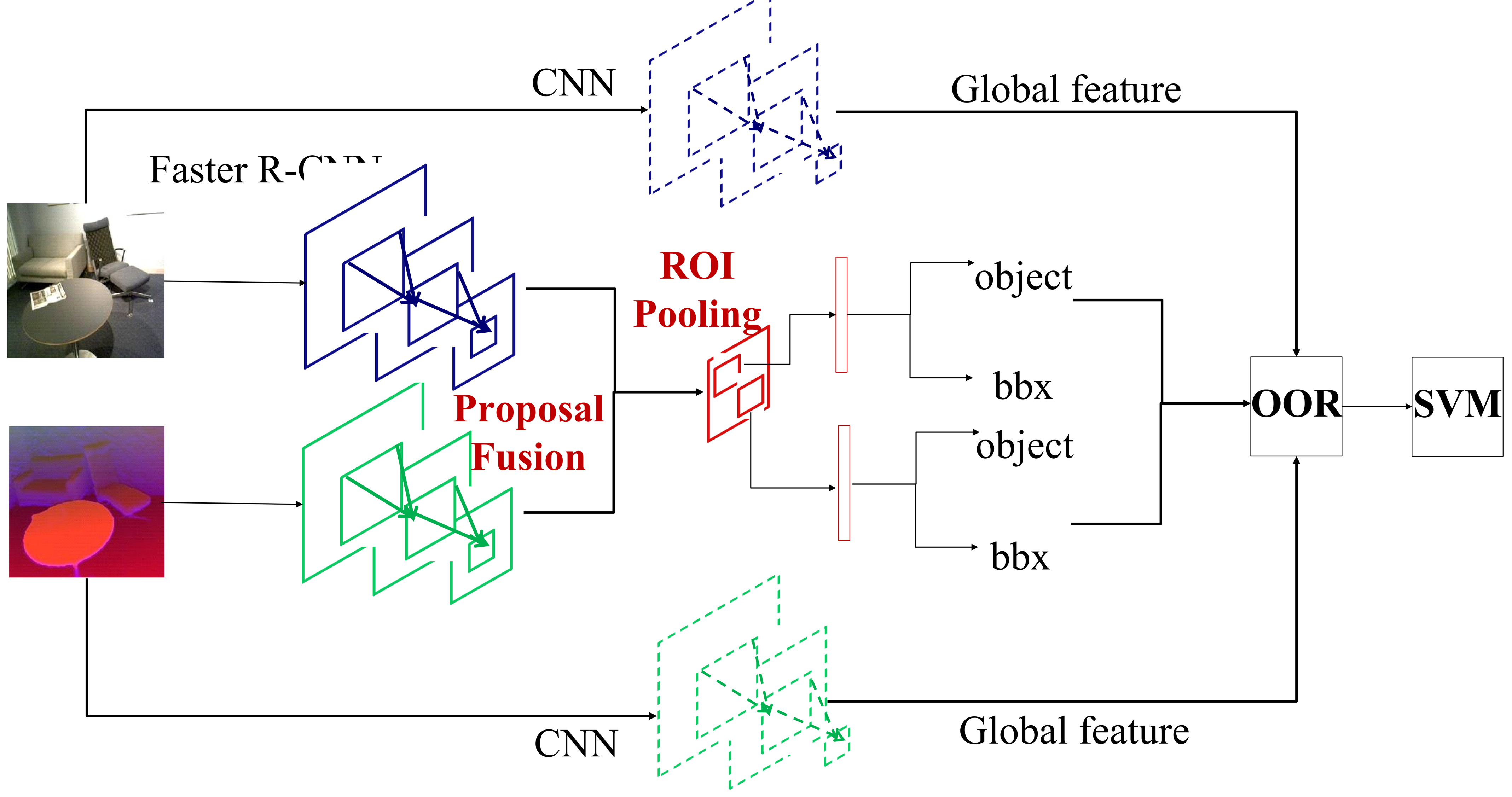
- ☐ Propose to detect OOR for image representation
- ☐ Propose to combine RGB-D representations with multi-modal fusion of object proposals

Representation of Object-to-Object Relation

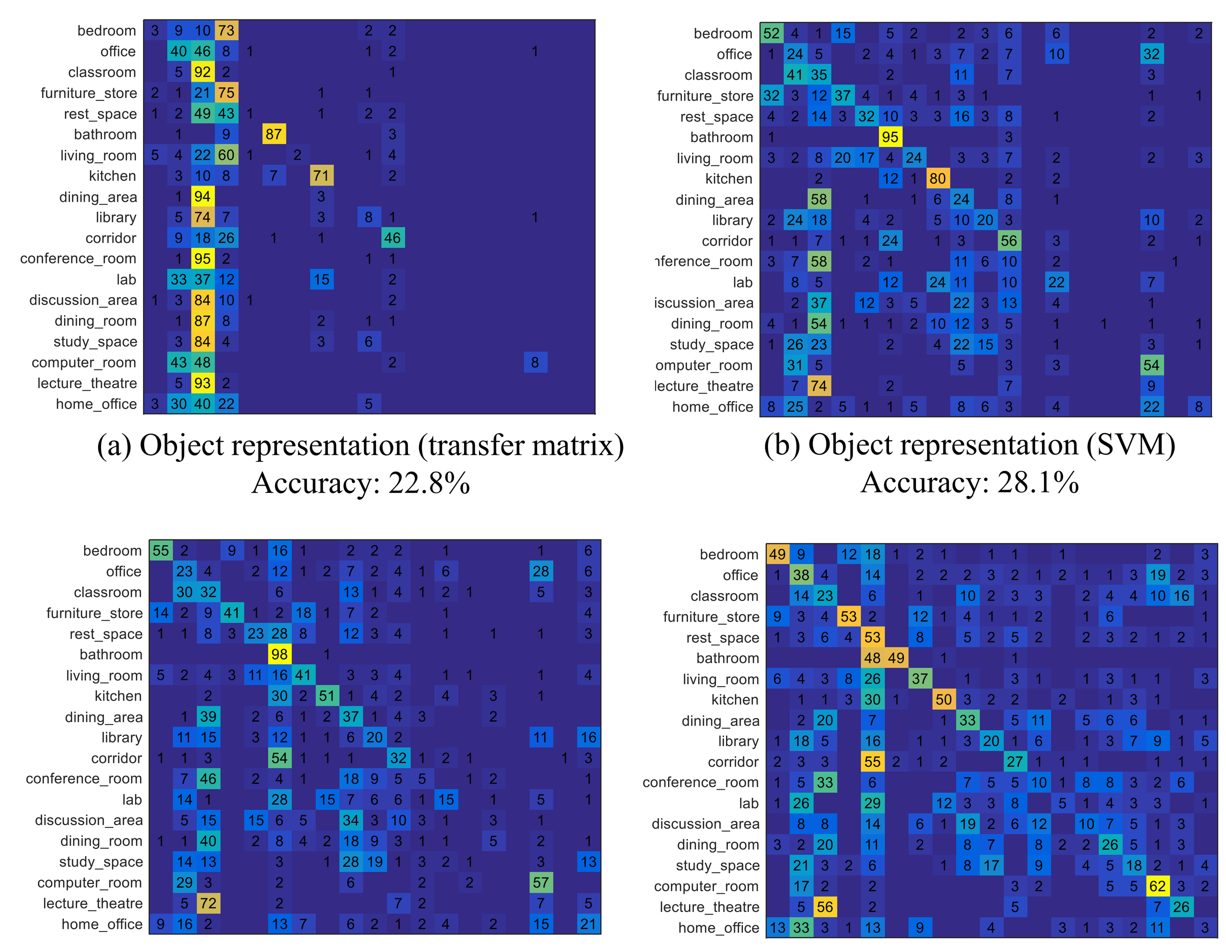


(a) "dinning room" (b) Object representation (c) Object co-occurrences (d) Object-to-object Relation (**OOR**)

Framework of RGB-D Scene Recognition



Comparisons of representations



(a) Object representation (transfer matrix) Accuracy: 22.8%

(b) Object representation (SVM) Accuracy: 28.1%

(c) Object co-occurrences (SVM) Accuracy: 28.8%

(d) Object-to-Object Relation (SVM) Accuracy: 31.1%

With OOR, the classifier obtains less confusion and better accuracy

Experimental results

Table 1: Object detection AP (%) of SUN RGB-D

Model	bathtub	bed	bookshelf	box	chair	counter	desk	door	dresser	garbage_bin
FRCN-RGB	34.4	63.2	39.8	12.5	43.9	42.2	20.3	30.7	30.0	40.0
FRCN-Depth	54.5	71.6	25.5	5.0	45.4	39.5	22.2	10.5	18.0	34.2
FRCN-RGBD	**57.5**	**75.6**	**44.2**	**17.7**	**49.6**	**48.9**	**25.4**	**33.6**	**40.2**	**49.2**
Model	lamp	monitor	night_stand	pillow	sink	sofa	table	tv	toilet	mAP
FRCN-RGB	38.5	34.3	39.2	33.0	46.9	39.5	34.6	23.2	74.5	37.9
FRCN-Depth	40.0	18.8	34.8	40.2	49.2	44.9	41.2	14.3	70.0	35.8
FRCN-RGBD	**53.0**	**44.0**	**47.6**	**48.6**	**61.1**	**50.3**	**43.2**	**35.2**	**81.7**	**47.7**

Table 2: Scene recognition accuracy (%) with intermediate representation

Intermediate representations	RGB	Depth	RGB-D
P^I_S	16.8	13.9	17.8
P^I_O	31.4	26.5	31.9
P^I_{OO}	32.7	28.7	33.4
P^I_{OOR}	**33.5**	**30.0**	**36.3**

P^I_S : inference with object representation
P^I_O: SVM classification with object representation
P^I_{OO}:SVM classification with object co-occurrence
P^I_{OOR}:SVM classification with OOR

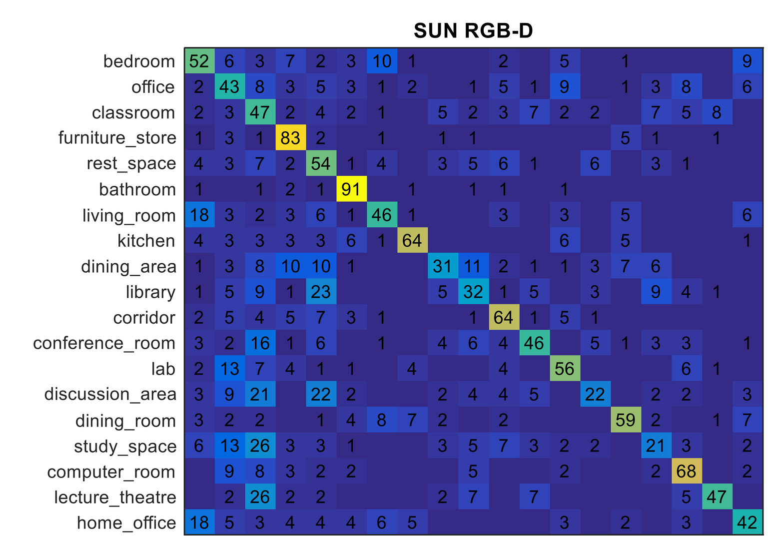
Table 4: Comparisons on SUN RGB-D in accuracy(%)

	Method	RGB-D
Proposed	Local-OOR	**50.3**
	Global+Local	**52.6**
	Global+Local-OOR	**54.0**
State-of-the-art	Song *et al.* [33]	39.0
	Zhu *et al.*[43]	41.5
	Wang *et al.* [39]	48.1
	Song *et al.* [34]	52.4

Global: CNN features of images
Local: CNN features of bounding boxes

Table 5: Comparisons on NYUD2 in accuracy(%)

Method	RGB	Depth	RGB-D
Proposed methods			
Local	51.2	46.4	56.4
OOR	45.1	40.9	48.6
Global	57.3	54.1	64.0
Local-OOR	-	-	60.1
Global+Local-OOR	-	-	**66.9**
State-of-the-art			
Gupta *et al.* [17]			45.4
Wang *et al.* [39]			63.9
Song *et al.* [34]			65.8



Confusion matrix of Global+Local-OOR.

Conclusion

- ➢Introduce some analysis and insights between objects and scenes
- ➢Propose a framework to extract object-to-object relation (OOR) for scene recognition
- ➢The propose method achieves the state-of-the-art on public RGB-D databases