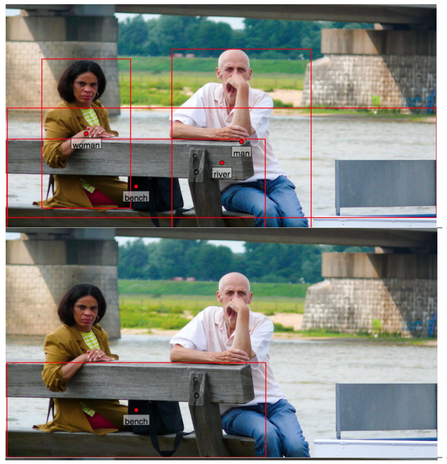
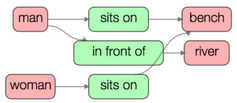
Scene Graph Generation from Images

Satoshi Tsutsui, Manish Kumar

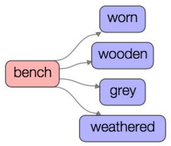
Motivation:Image Understanding

- What are the objects in a image?
- What are the attributes of these object?
- What are the relations between these objects?





A man and a woman sit on a park bench along a river.



Park bench is made of gray weathered wood

figure source Krishna et al. (2016)

Scene Graph

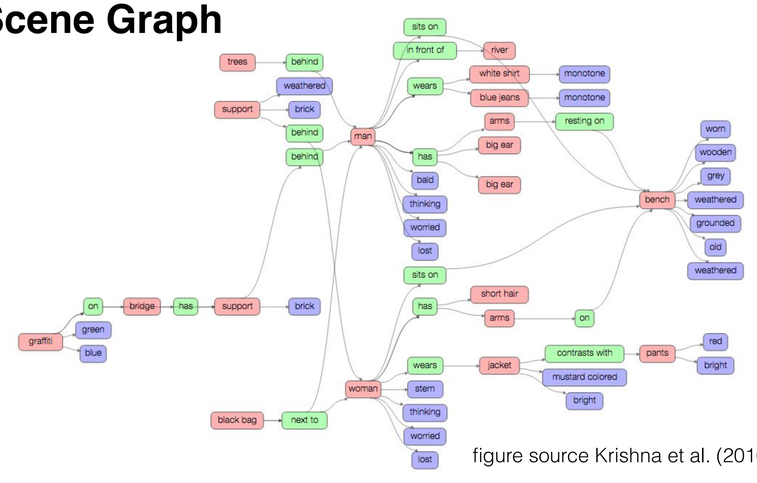
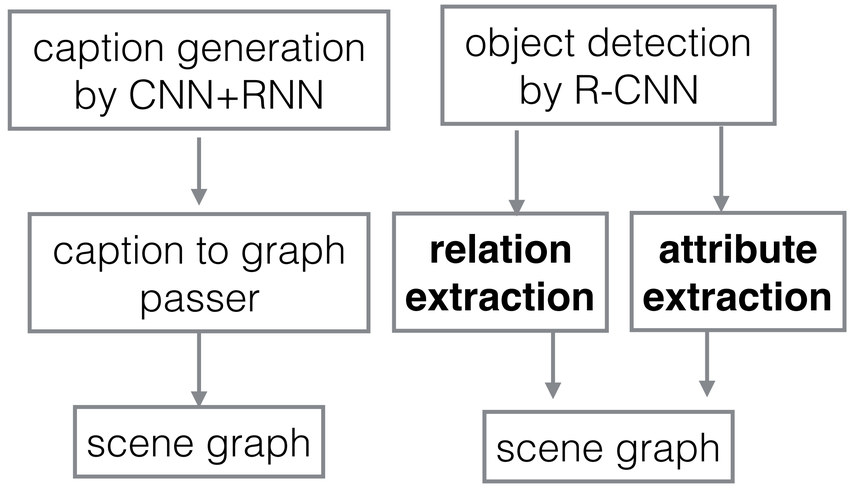


figure source Krishna et al. (2016)

Dataset: visualgenome.org

	# of images			vocabulary size
	train	val	test	
relations	42728	4000	4000	4265
attributes	38918	4000	4000	2834

Approaches



Example

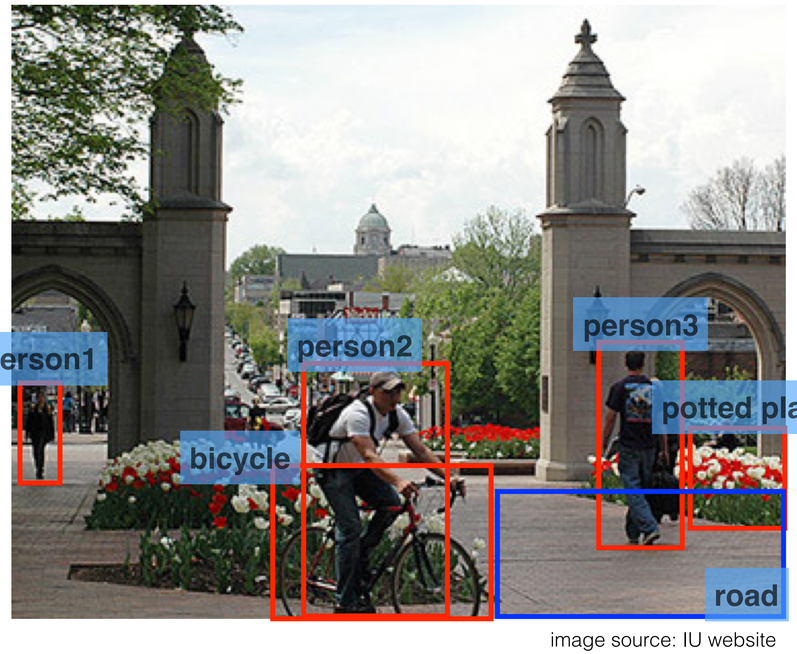
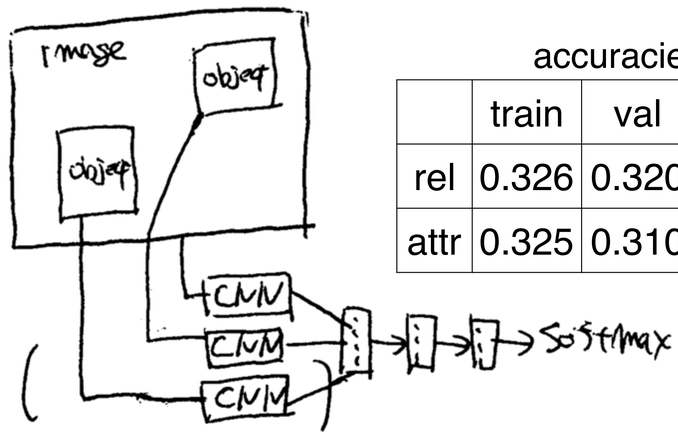


image source: IU website

"a group of people riding bikes down a street"

scene graph parser: [group, of, people]
[people, ride, bike]

Neural net for relations and attributes



accuracies

	train	val	test
rel	0.326	0.320	0.317
attr	0.325	0.310	0.314

What is the relation from person3 to road?

1	0.630	on
2	0.071	on the
3	0.054	in
4	0.036	is on
5	0.031	walking on

What is the relation from person2 to bicycle?

1	0.224	on
2	0.117	riding
3	0.056	is on a
4	0.032	sitting on
5	0.020	on a

What is the attribute of potted plant?

1	0.006	white
2	0.005	black
3	0.004	blue
4	0.003	visible
5	0.003	green

What is the attribute of bicycle?

1	0.006	white
2	0.005	black
3	0.004	blue
4	0.003	visible
5	0.003	green

Limitations and Future Plan

- Complete the system to output the graph
- Consider positional information for relations
- Train end to end deep net with object proposal