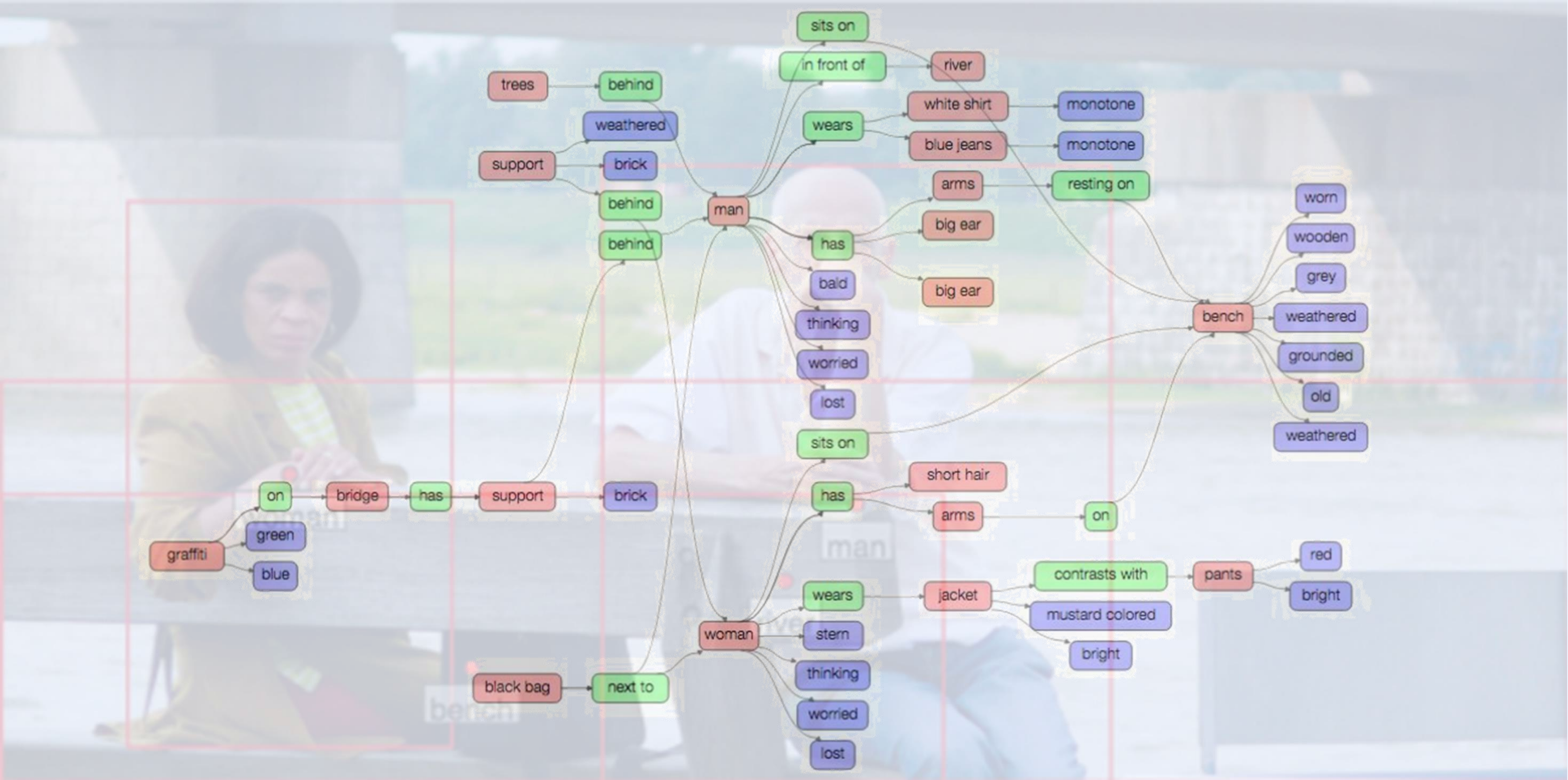
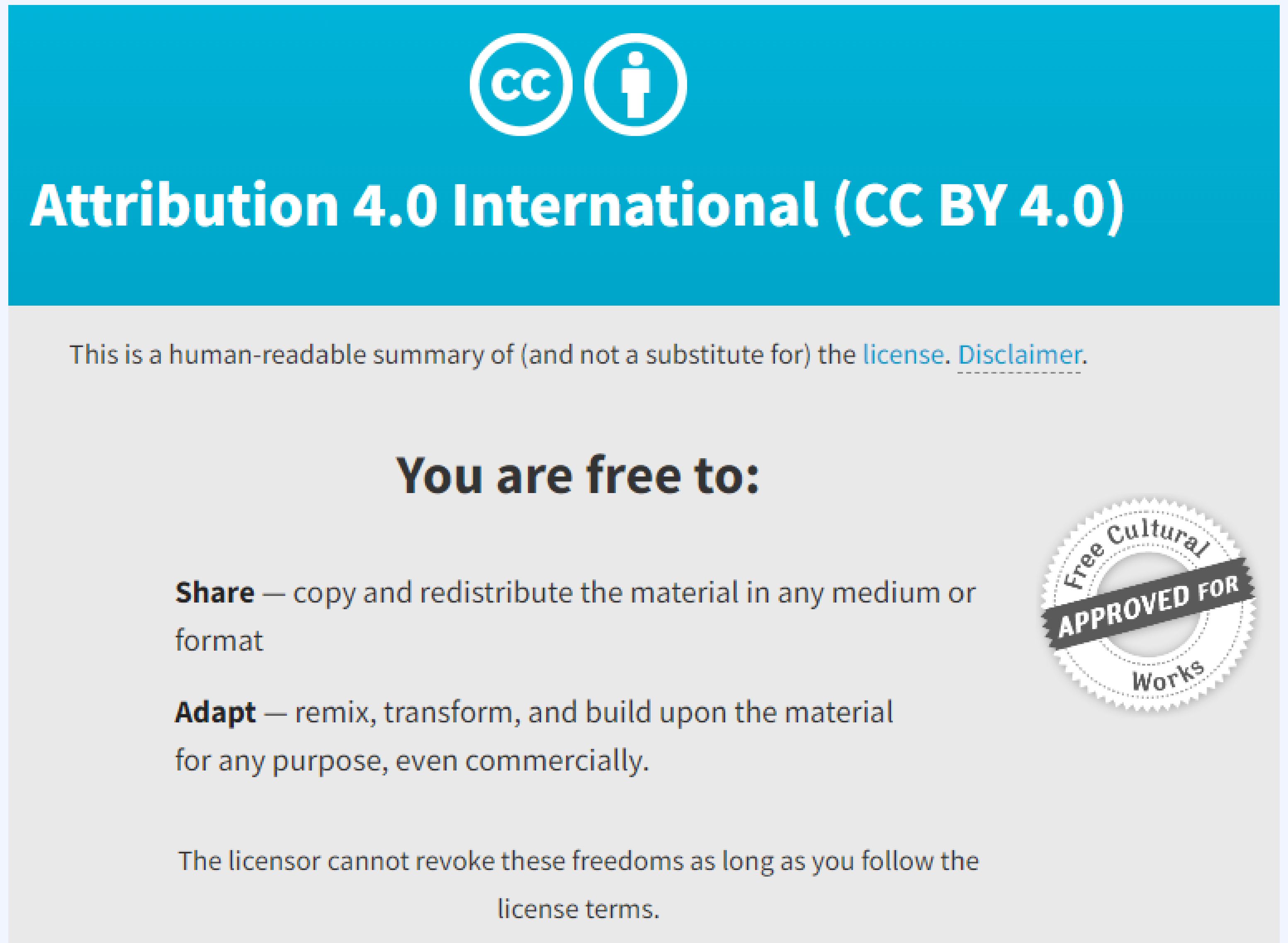
Scene Understanding

2 Scene Representation Learning (data)

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv





Attribution 4.0 International (CC BY 4.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

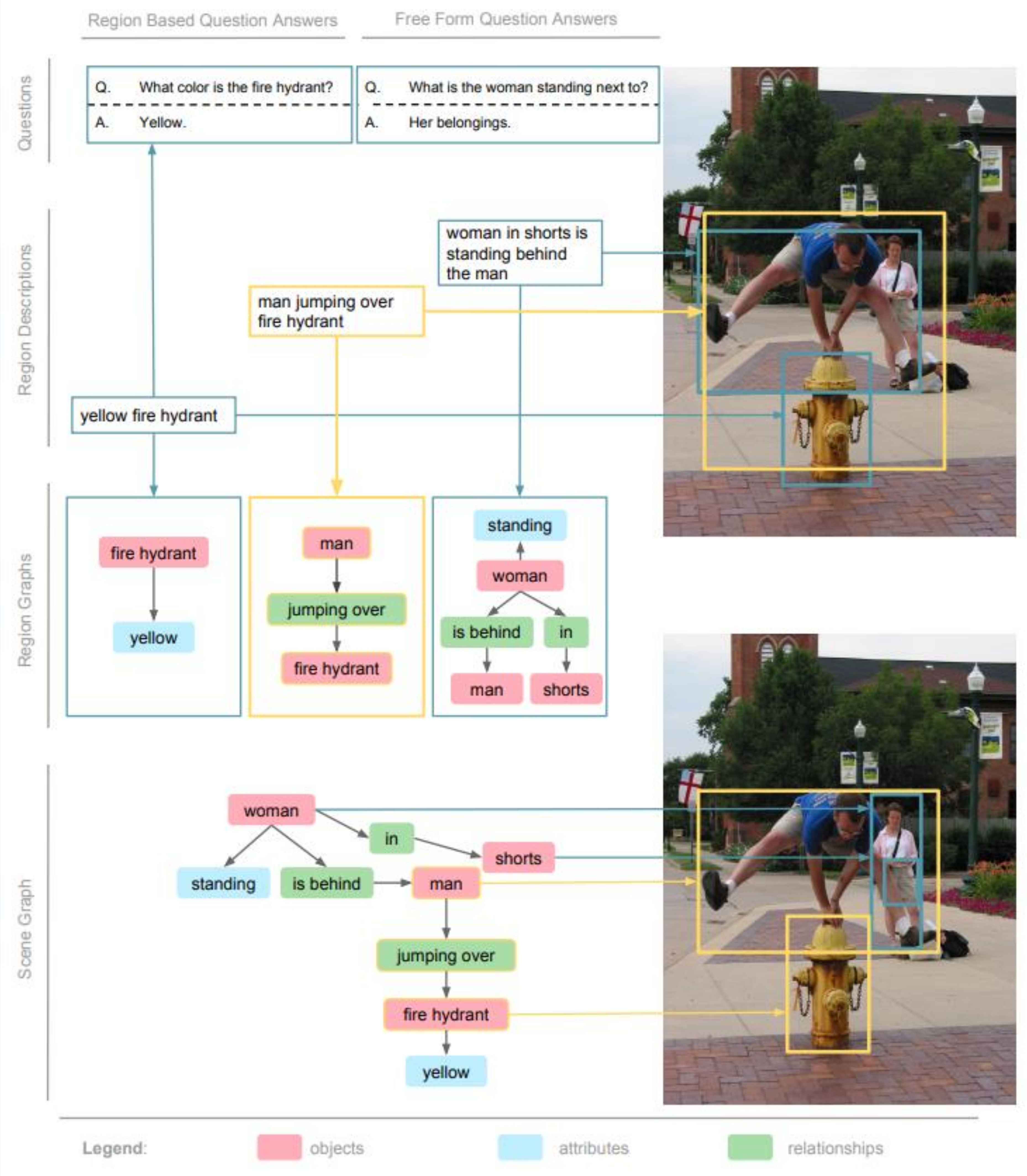
Share — copy and redistribute the material in any medium or format

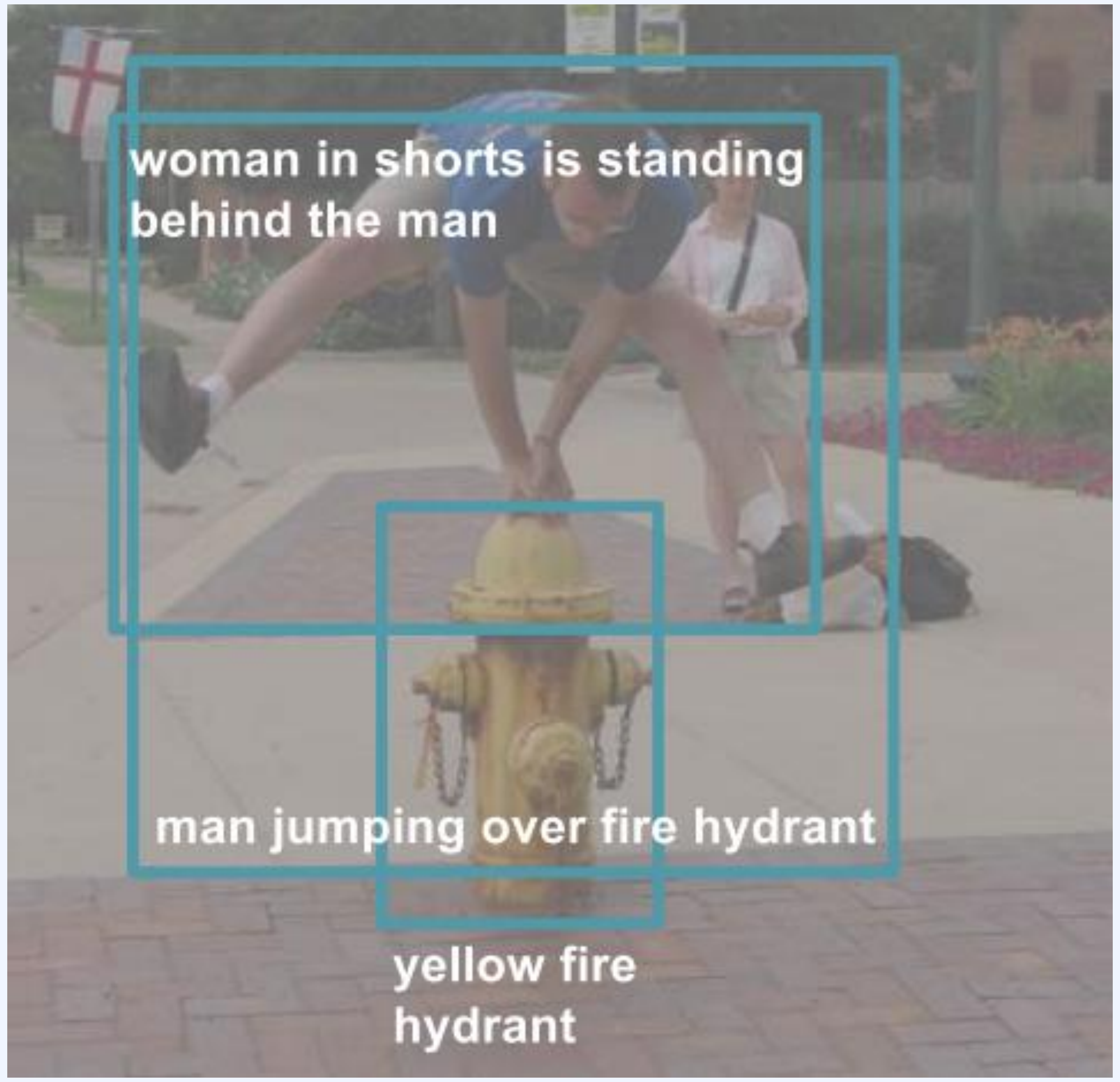
Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

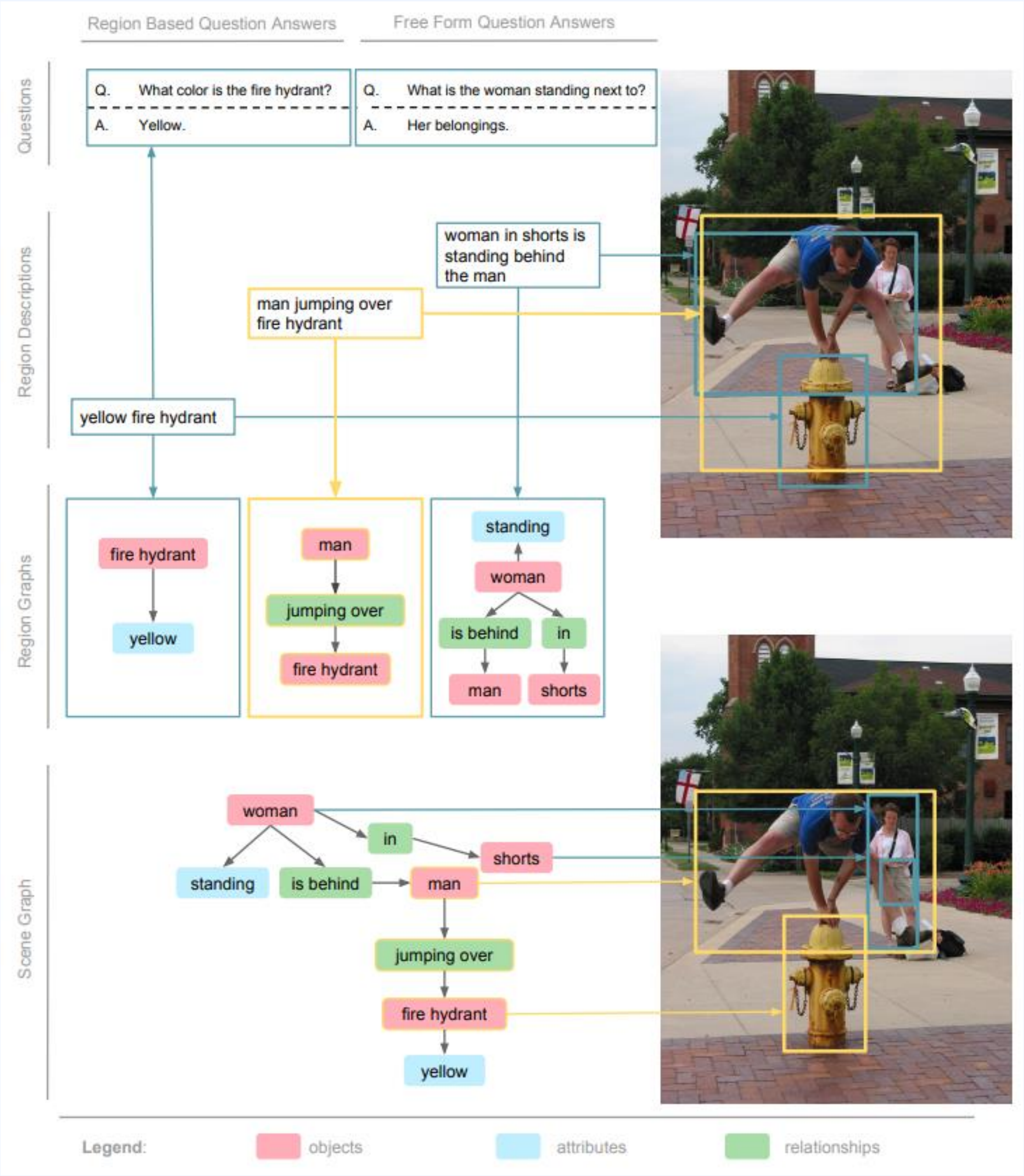



woman in shorts is standing behind the man
man jumping over fire hydrant
yellow fire hydrant

- **Region descriptions**
- Objects
- Attributes
- Relationships
- Region graphs
- Scene graphs
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv



Q. What color is the fire hydrant?
A. Yellow.

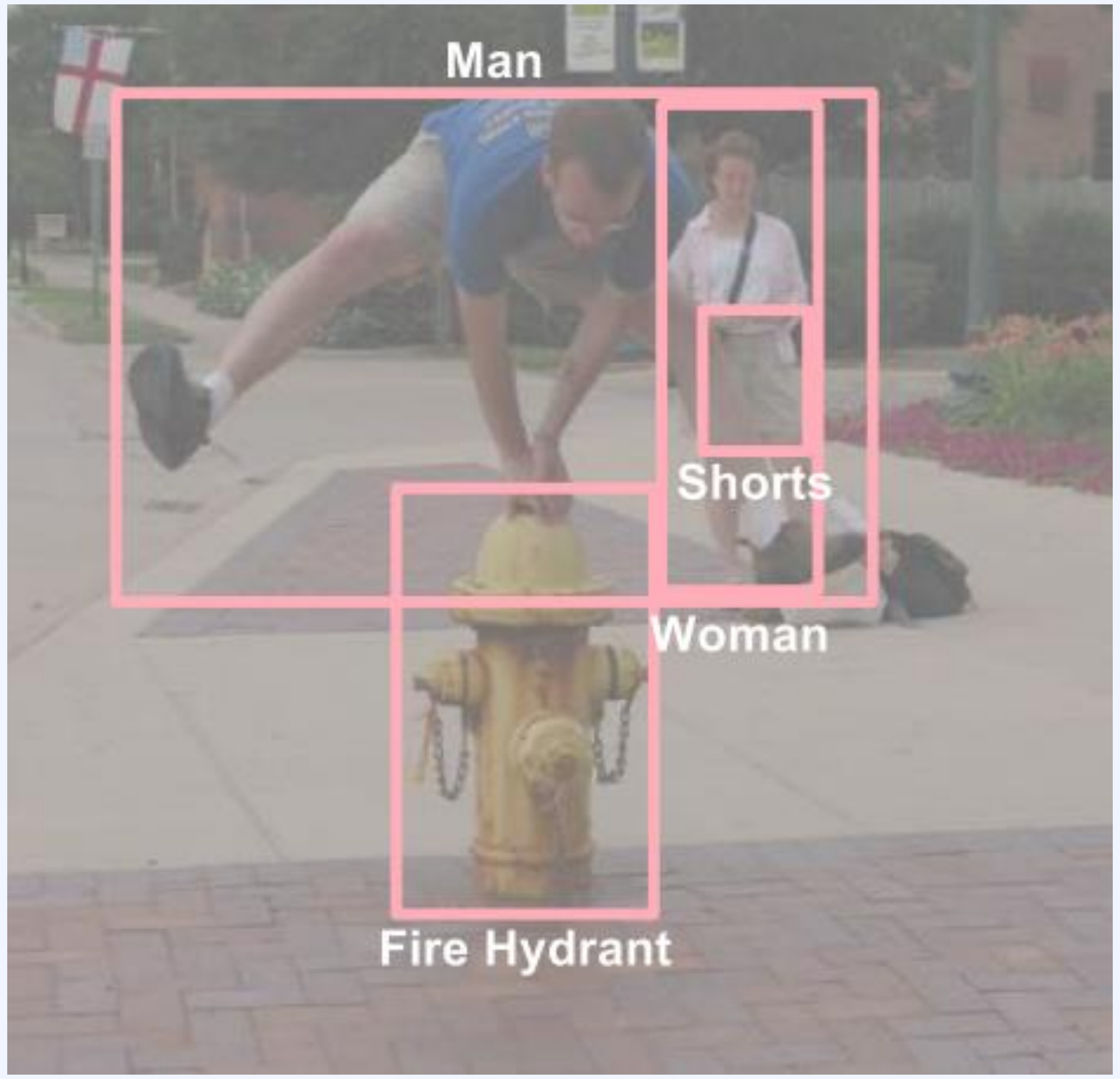
Q. What is the woman standing next to?
A. Her belongings.

woman in shorts is standing behind the man

man jumping over fire hydrant

yellow fire hydrant

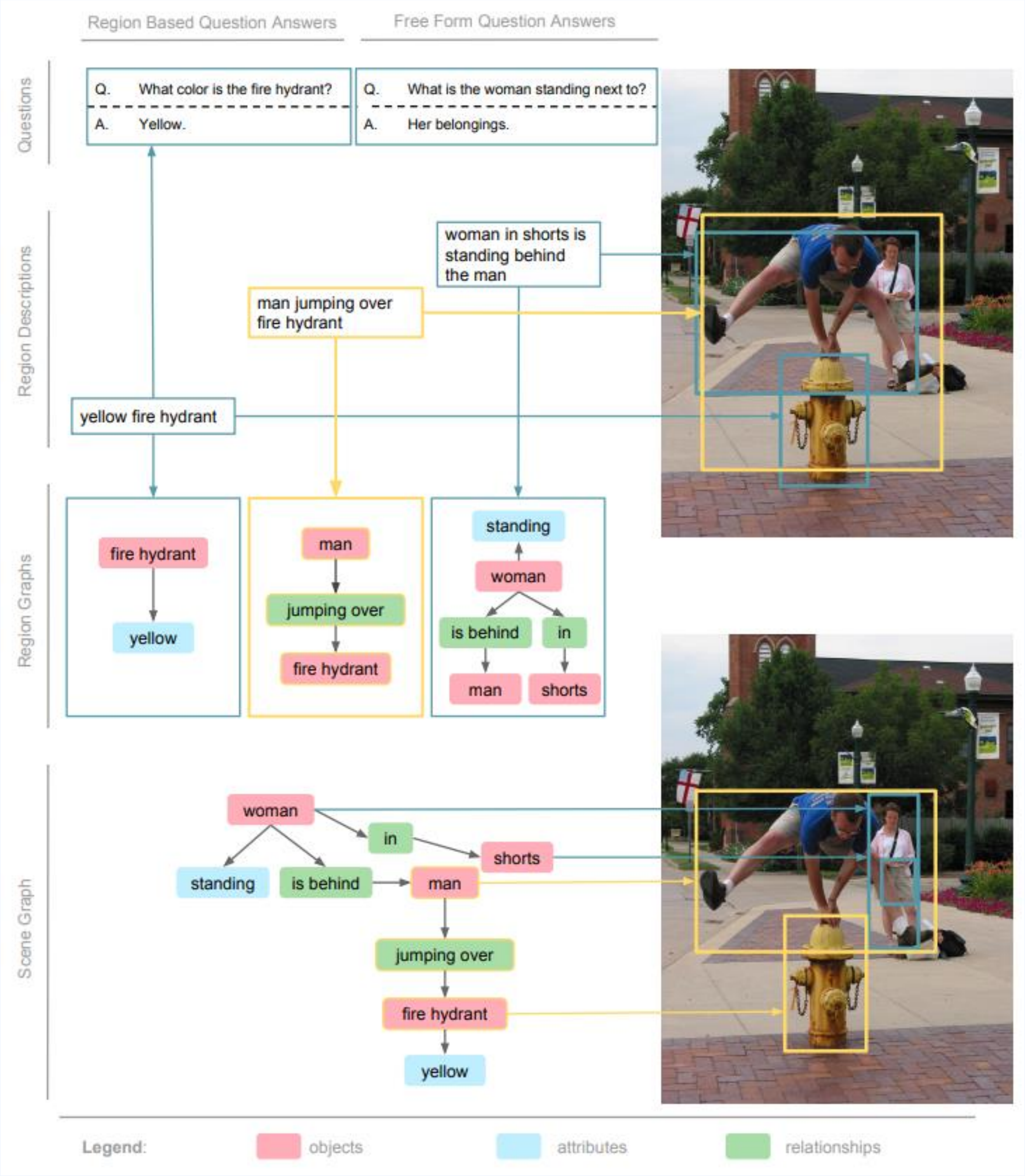
Legend: 🟥 objects 🟦 attributes 🟩 relationships

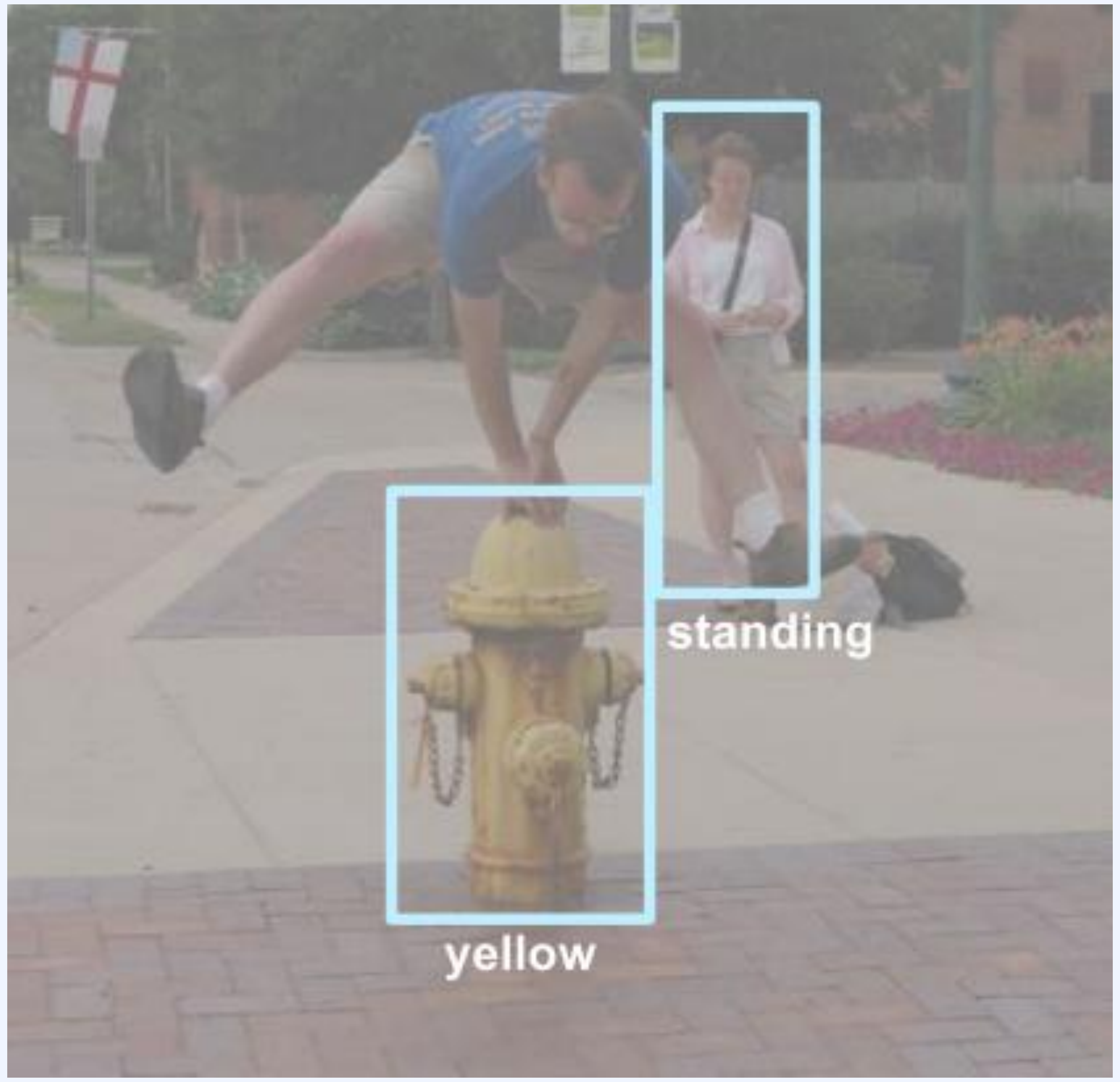


- Region descriptions
- **Objects**
- Attributes
- Relationships
- Region graphs
- Scene graphs
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

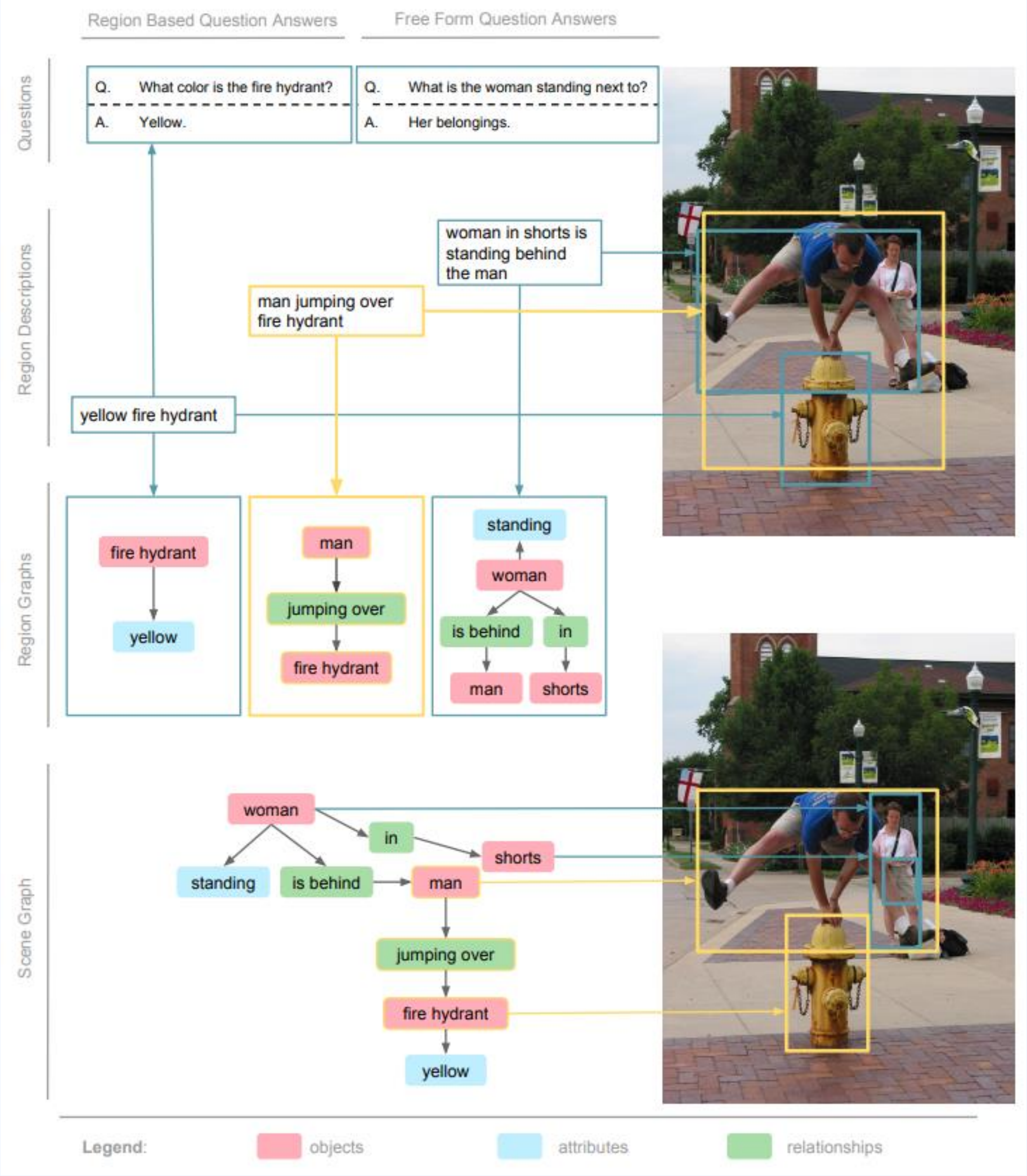


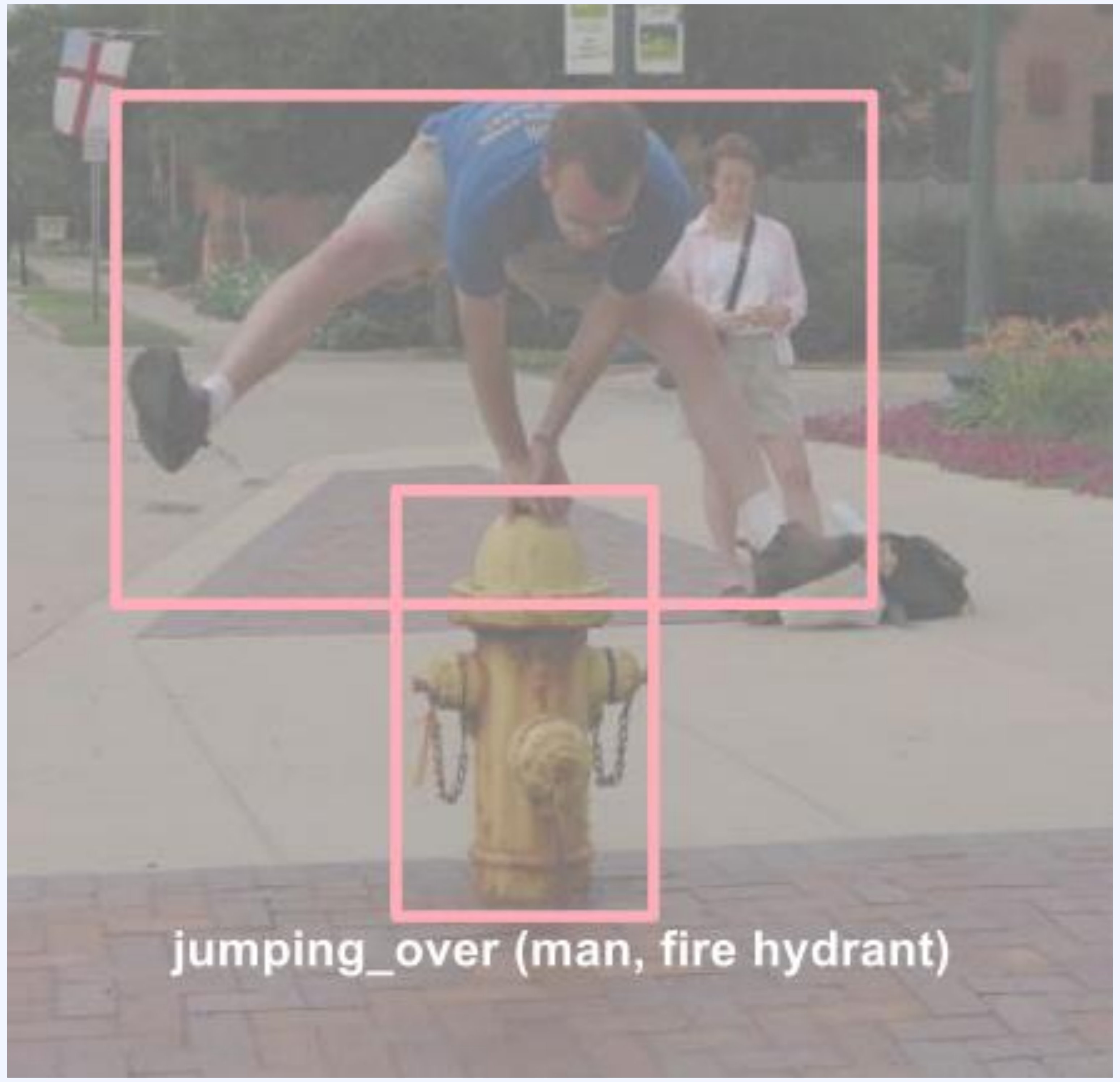


- Region descriptions
- Objects
- **Attributes**
- Relationships
- Region graphs
- Scene graphs
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

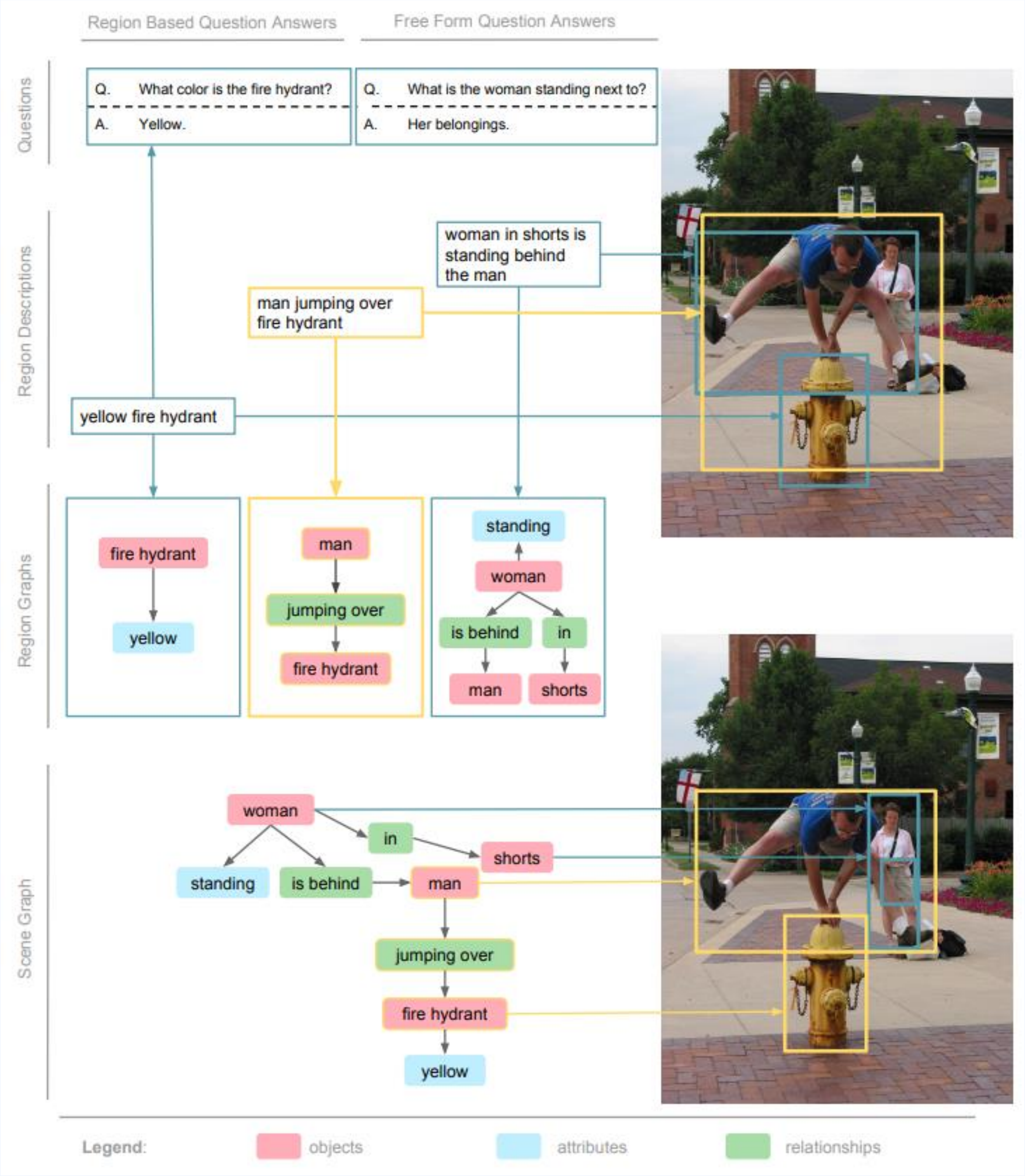


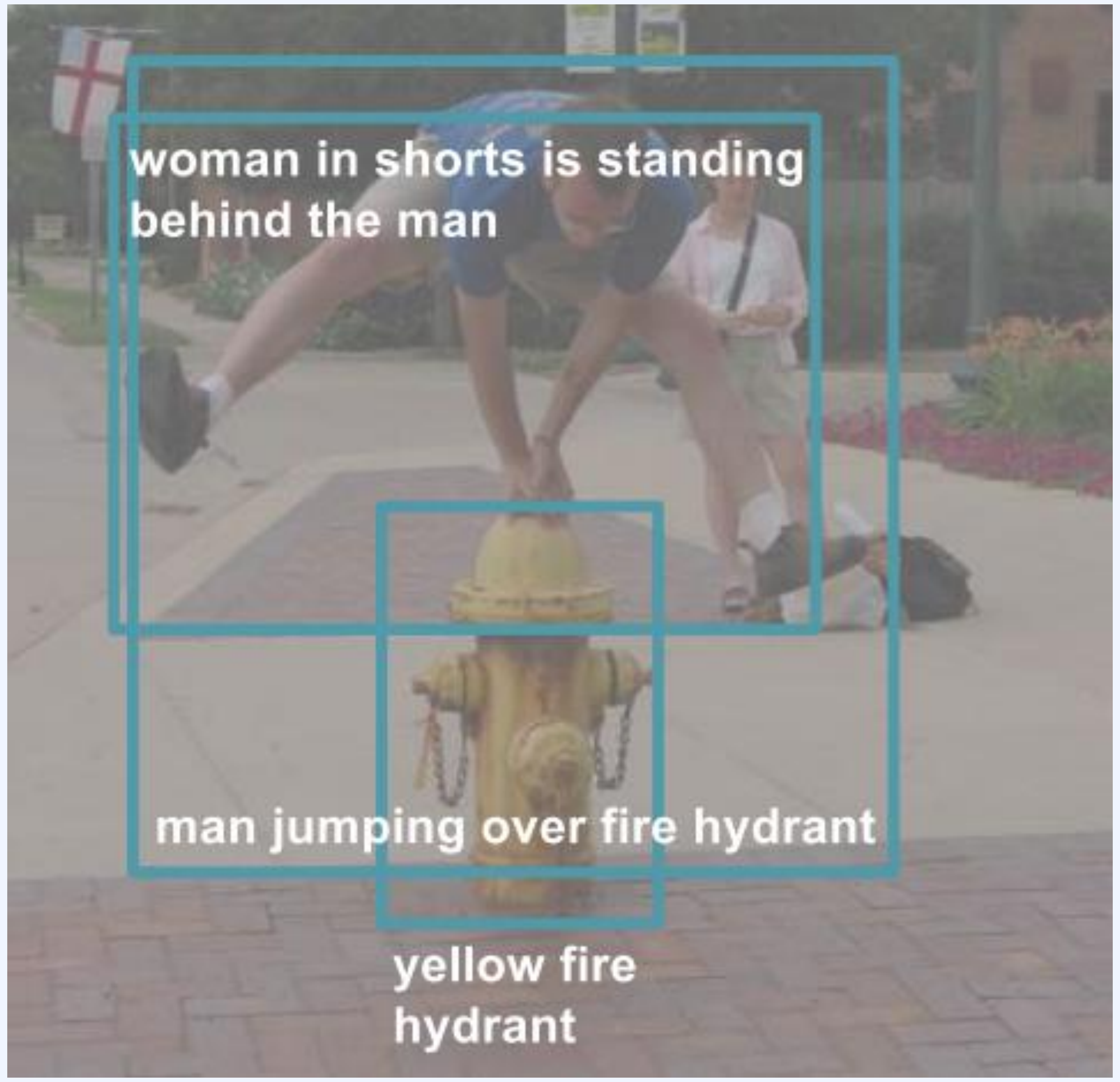


- Region descriptions
- Objects
- Attributes
- **Relationships**
- Region graphs
- Scene graphs
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

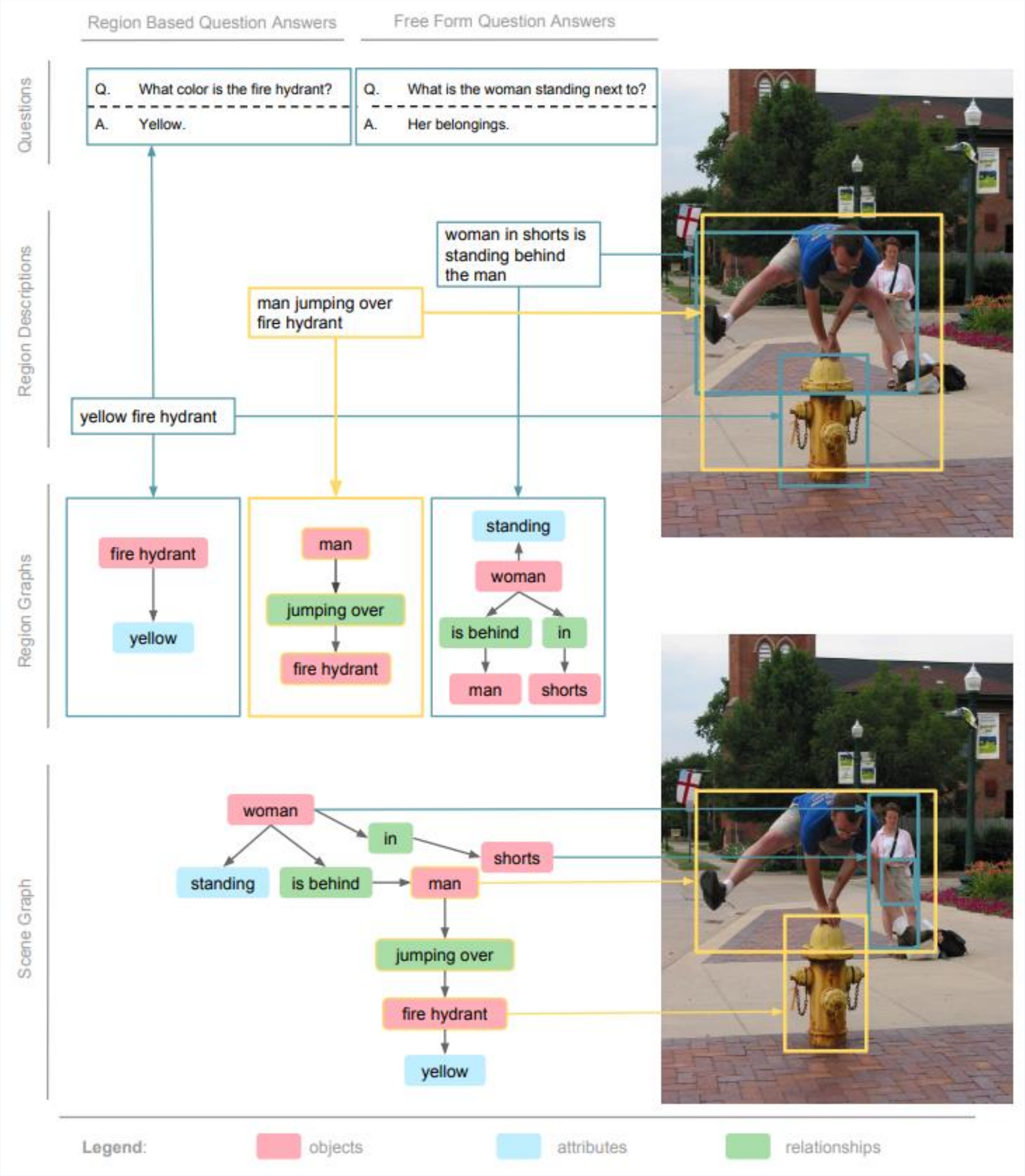


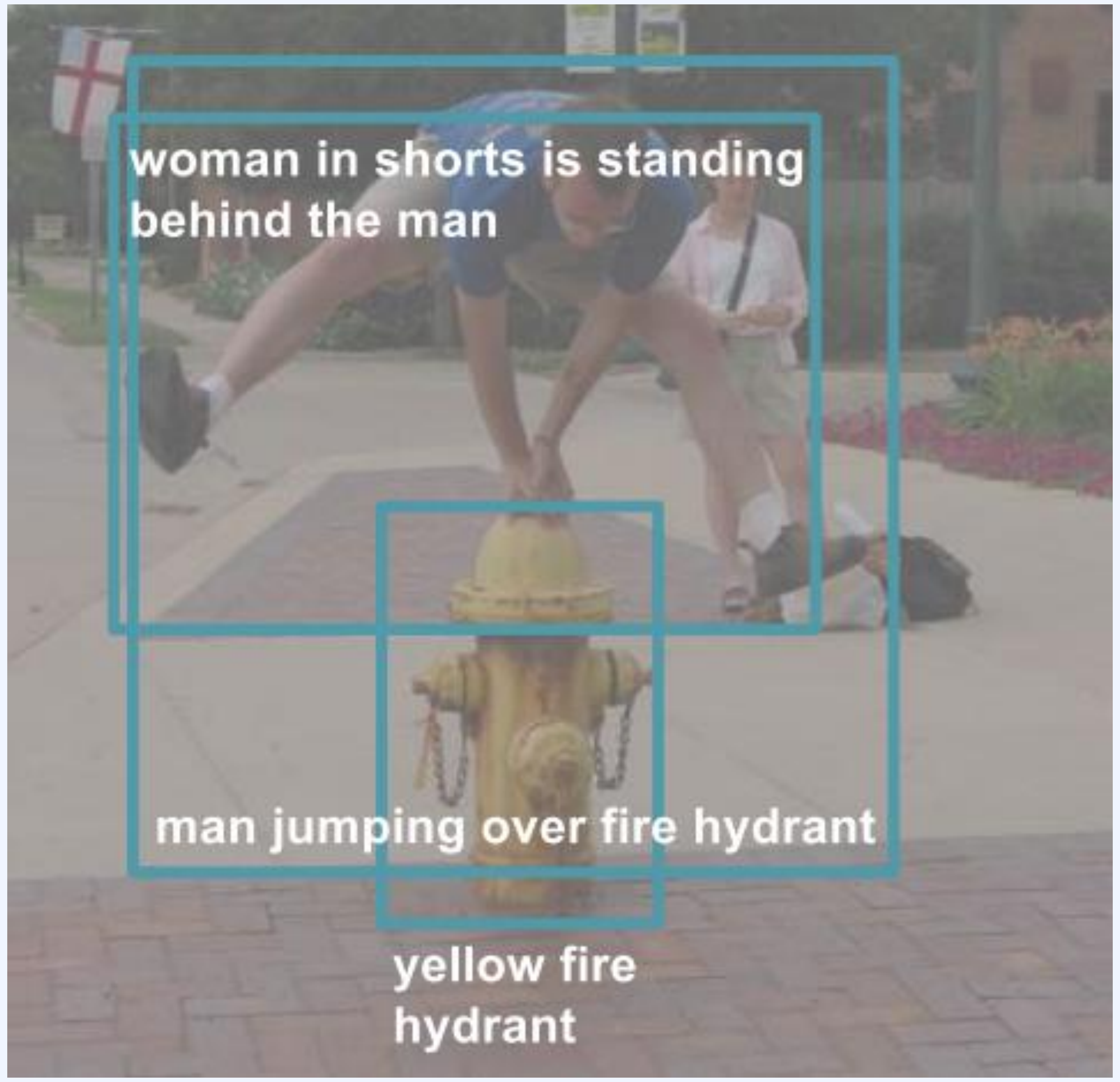

woman in shorts is standing behind the man
man jumping over fire hydrant
yellow fire hydrant

- Region descriptions
- Objects
- Attributes
- Relationships
- **Region graphs**
- Scene graphs
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

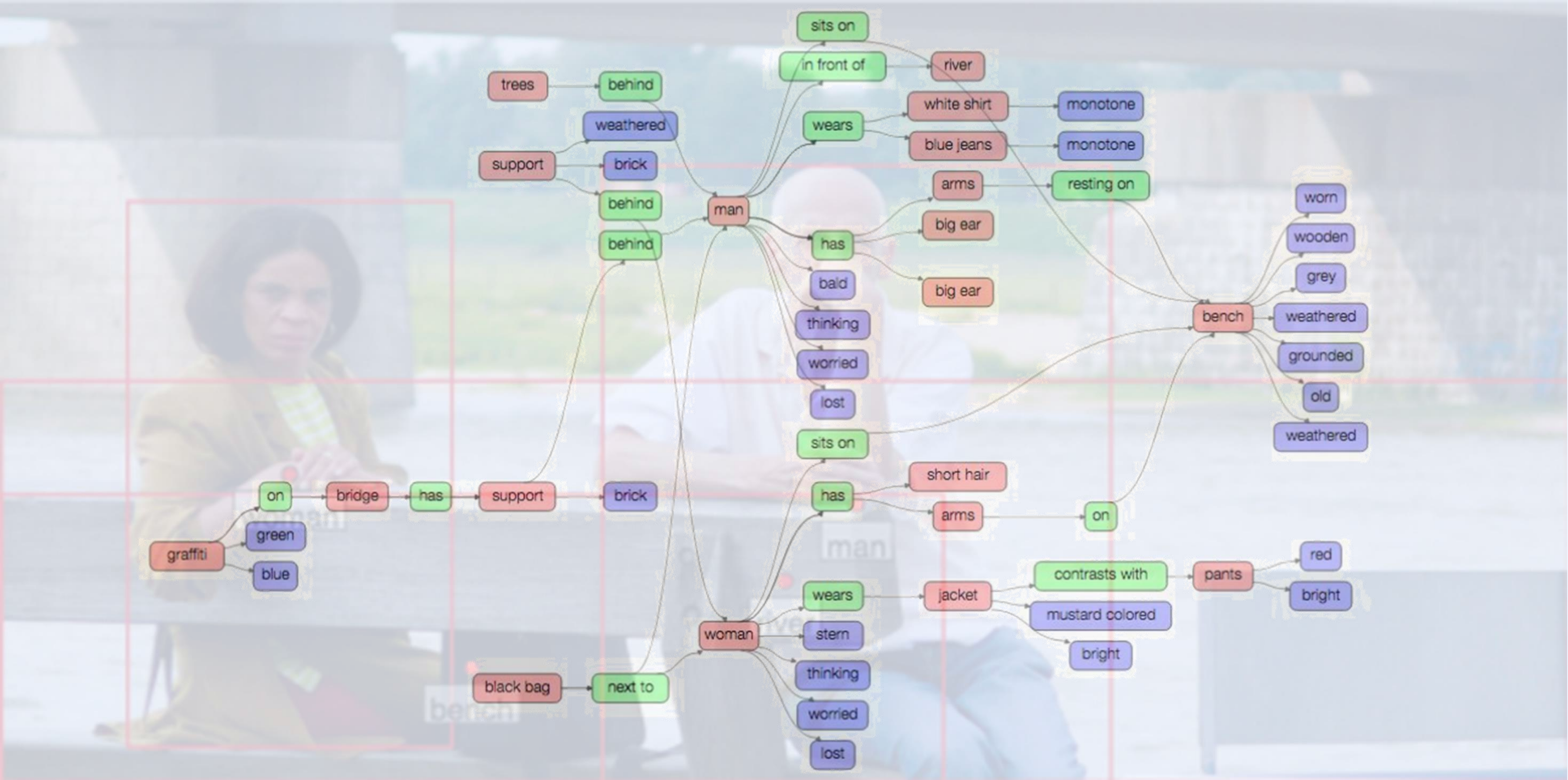



woman in shorts is standing behind the man
man jumping over fire hydrant
yellow fire hydrant

- Region descriptions
- Objects
- Attributes
- Relationships
- Region graphs
- **Scene graphs**
- Question answer pairs

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv



Basic Statistics

- Total images: 108,077
- Average image size (px): 500.14
- Maximum image size (px): 1,280
- Minimum image size (px): 72

- Total region descriptions: 4,297,502
- Total image object instances: 1,366,673
- Unique image objects: 75,729

- Total object-object relationship instances: 1,531,448
- Unique relationships: 40,480
- Total attribute-object instances: 1,670,182
- Unique attributes: 40,513

- Total Scene Graphs: 108,249
- Total Region Graphs: 3,788,715

- Total Question Answers: 1,773,258

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

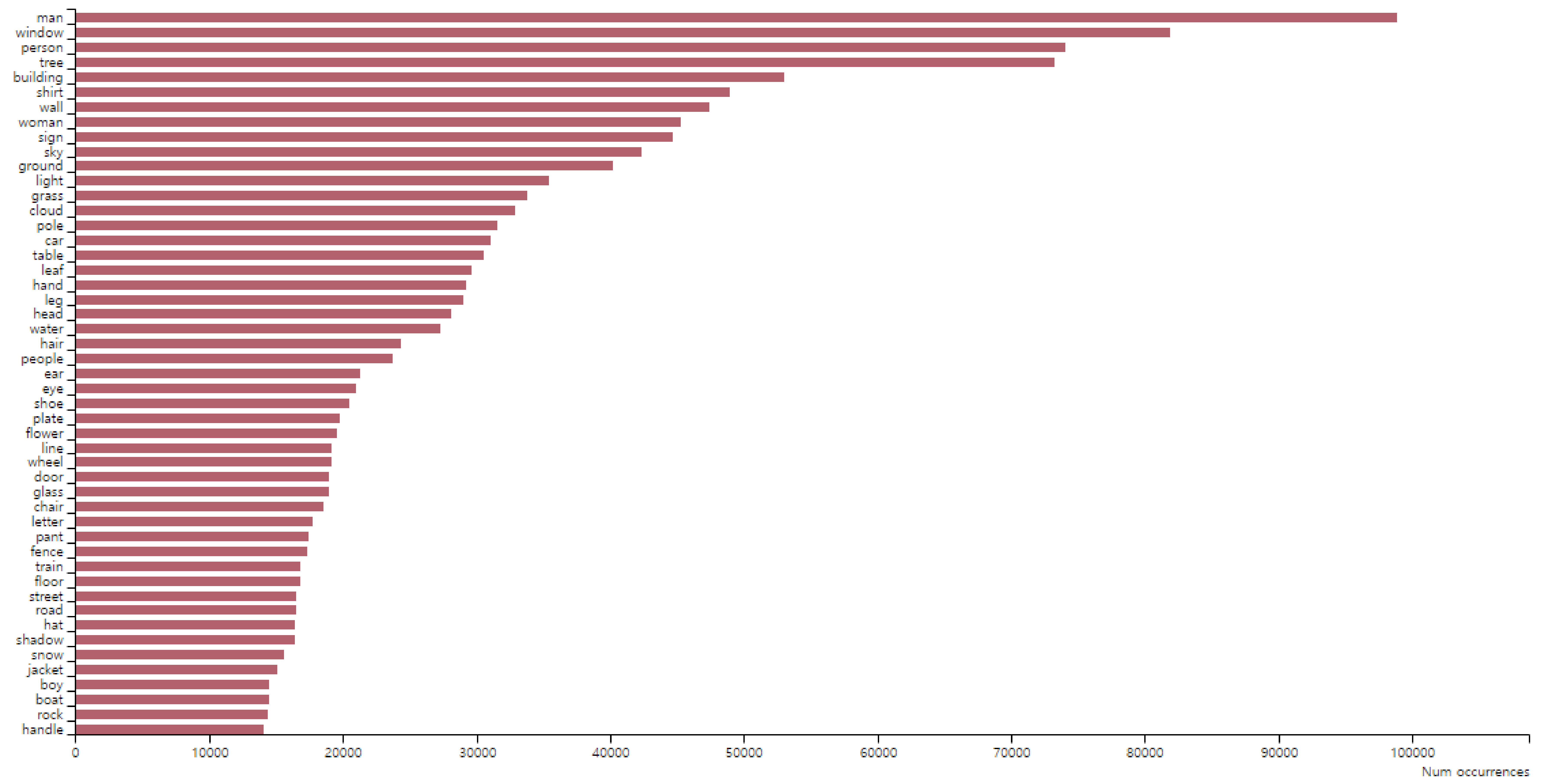
	Average number of objects	Average number of relationships	Average number of attributes
Per region annotation	1.01	0.63	0.77
Per image	21.24	17.68	16.08

Most common objects	Most common predicates	Most common attributes
•man	•ON	•white
•person	•has	•blue
•woman	•IN	•red
•building	•WEARING	•black
•sign	•wears	•green
•table	•behind	•yellow
•bus	•next to	•brown
•window	•with	•large
•sky	•near	•on
•tree	•in front of	•parked

Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

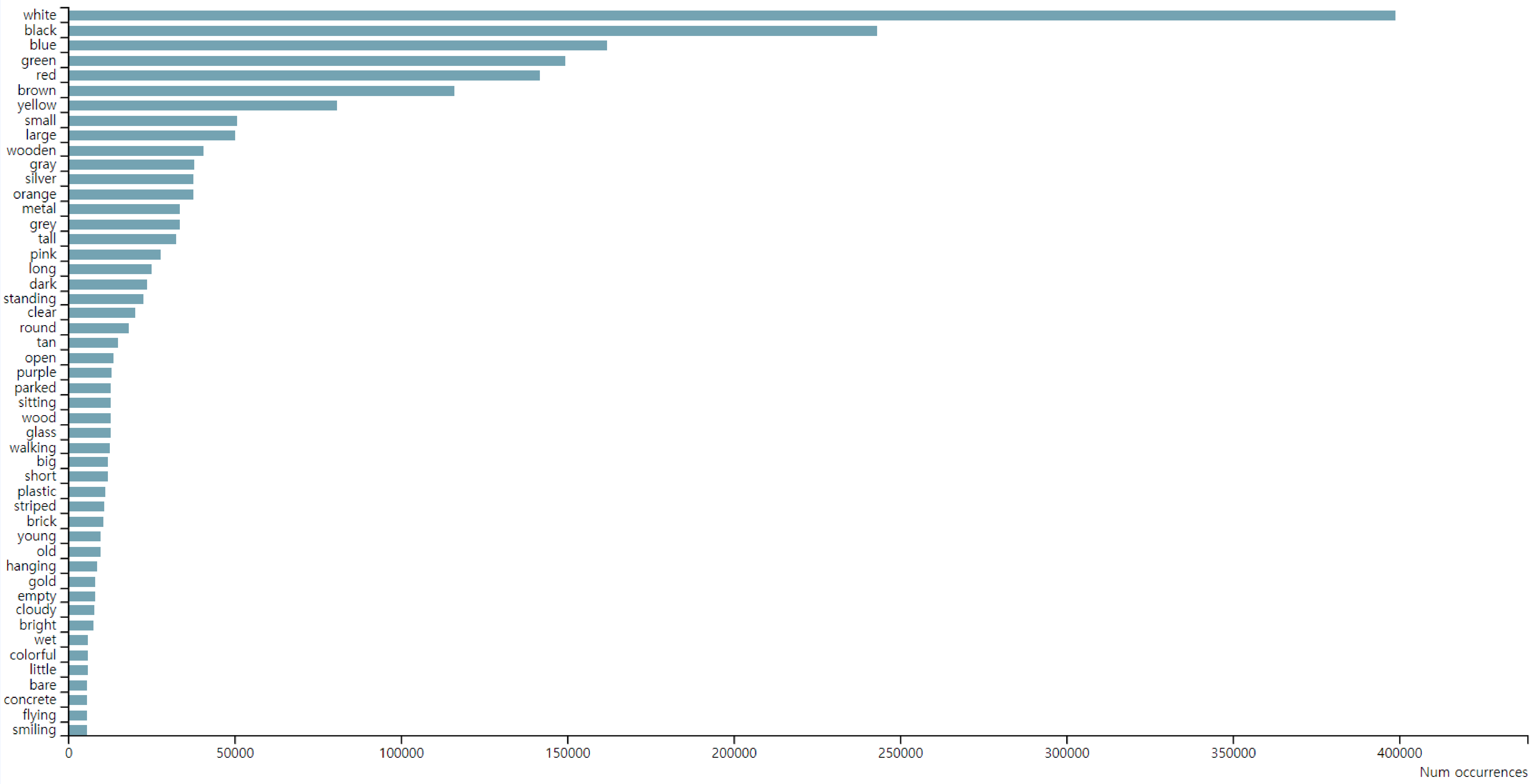
Top Object Names



Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

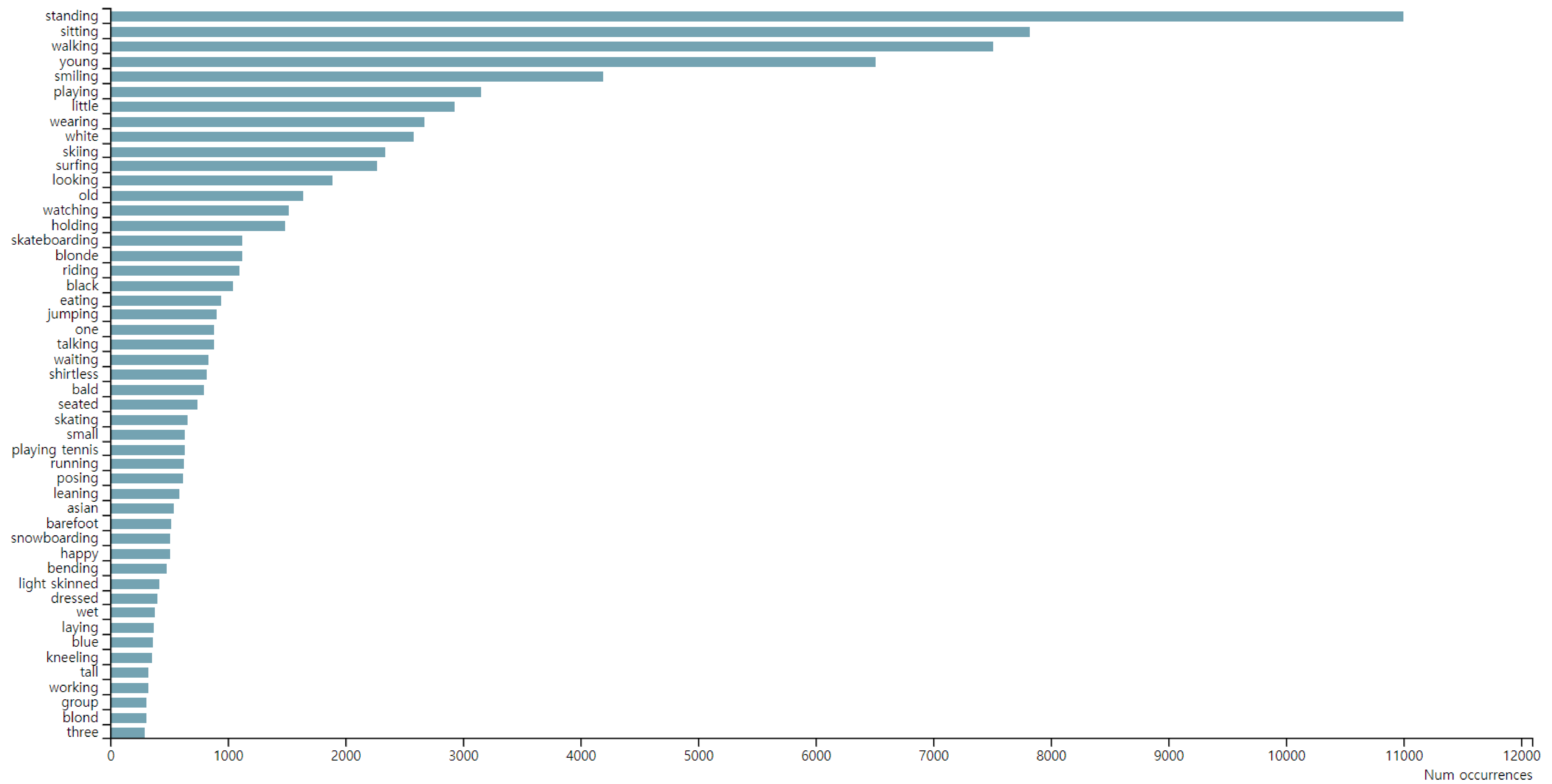
Top Attributes



Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

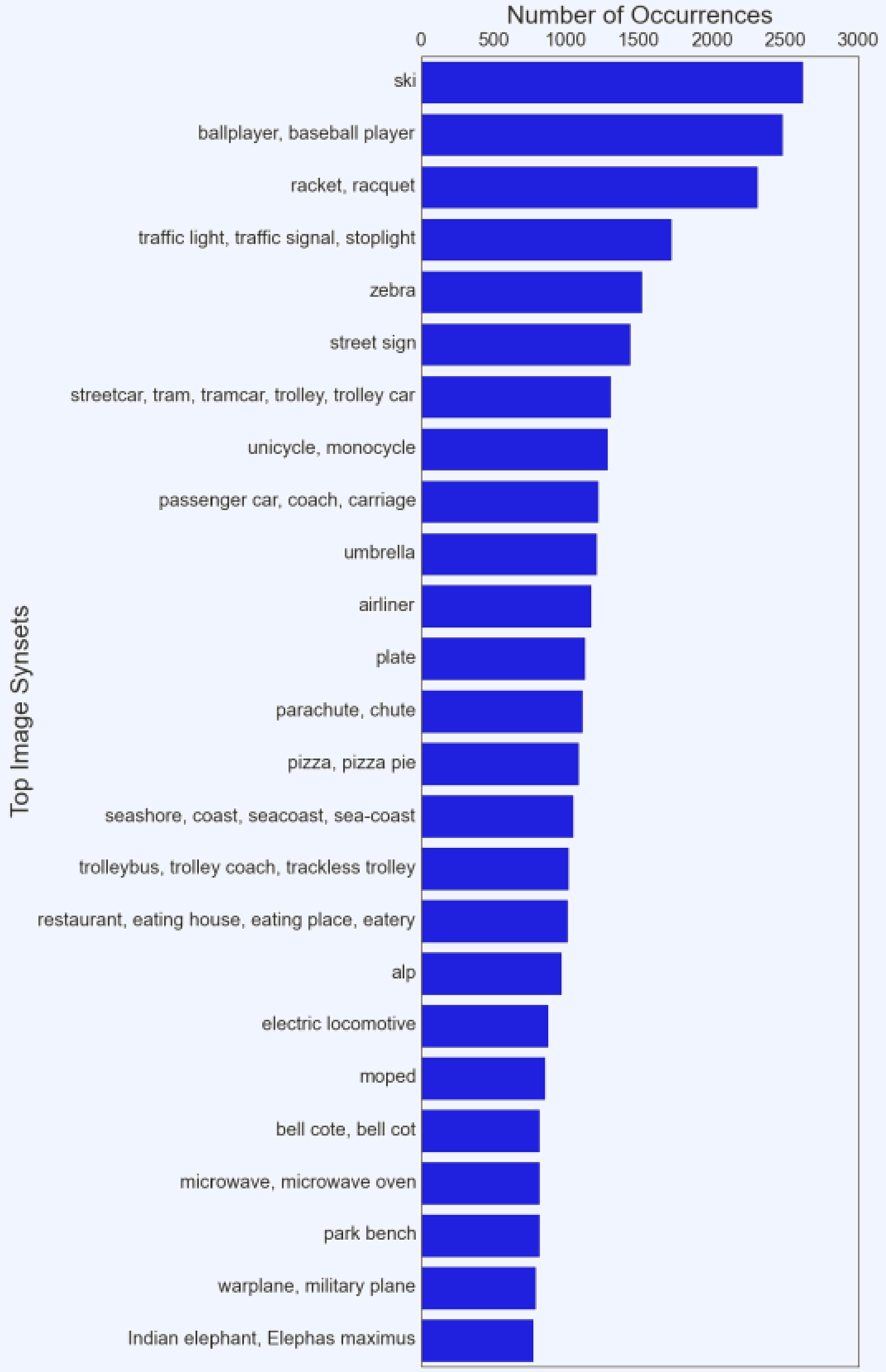
Top Attributes on People (man, women, person)



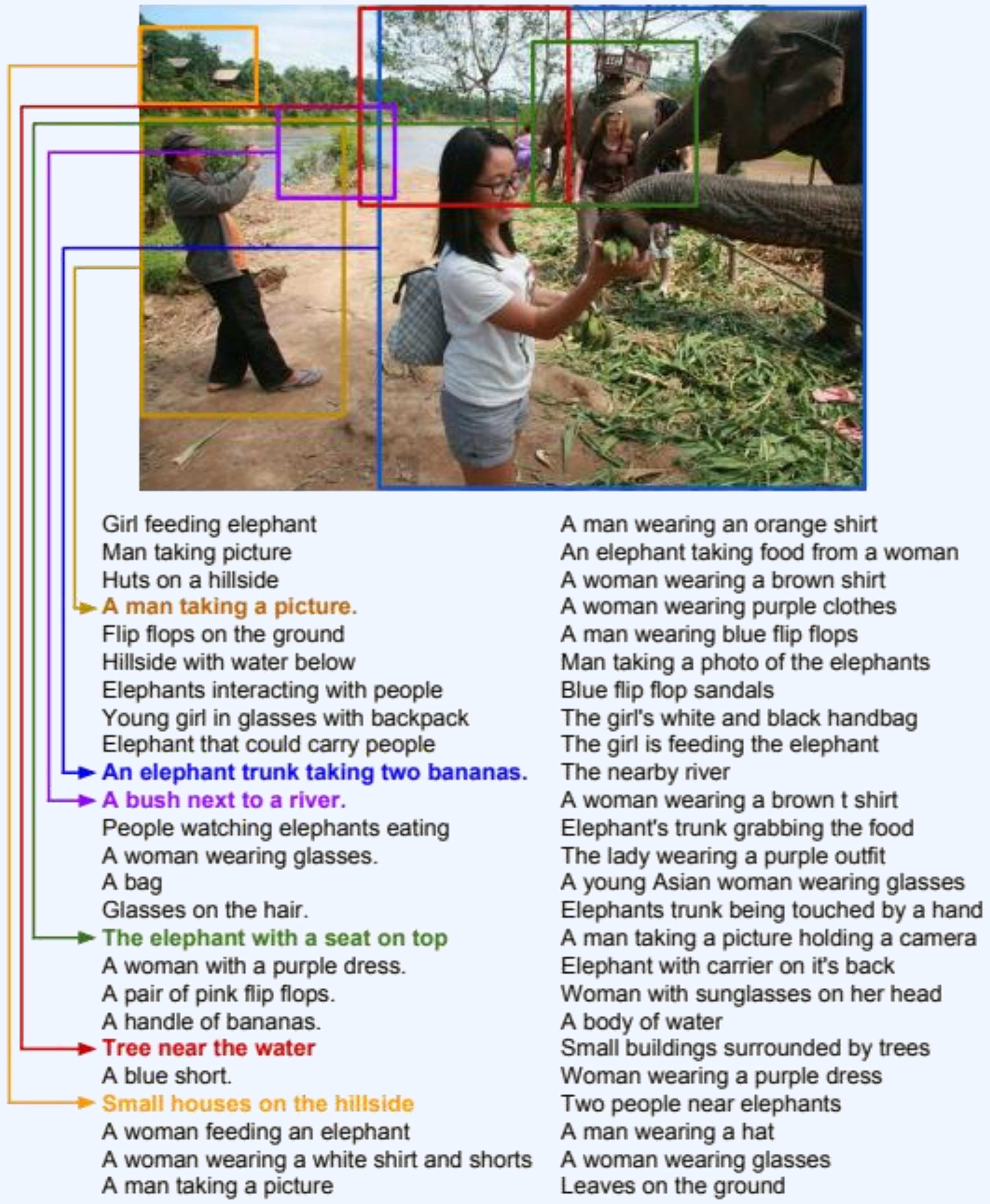
Scene Understanding
Scene representation learning (data)

R. Krishna et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations . arxiv

Top Image Synsets



Region Descriptions



Girl feeding elephant
Man taking picture
Huts on a hillside
A man taking a picture.
Flip flops on the ground
Hillside with water below
Elephants interacting with people
Young girl in glasses with backpack
Elephant that could carry people
An elephant trunk taking two bananas.
A bush next to a river.
People watching elephants eating
A woman wearing glasses.
A bag
Glasses on the hair.
The elephant with a seat on top
A woman with a purple dress.
A pair of pink flip flops.
A handle of bananas.
Tree near the water
A blue short.
Small houses on the hillside
A woman feeding an elephant
A woman wearing a white shirt and shorts
A man taking a picture

A man wearing an orange shirt
An elephant taking food from a woman
A woman wearing a brown shirt
A woman wearing purple clothes
A man wearing blue flip flops
Man taking a photo of the elephants
Blue flip flop sandals
The girl's white and black handbag
The girl is feeding the elephant
The nearby river
A woman wearing a brown t shirt
Elephant's trunk grabbing the food
The lady wearing a purple outfit
A young Asian woman wearing glasses
Elephants trunk being touched by a hand
A man taking a picture holding a camera
Elephant with carrier on it's back
Woman with sunglasses on her head
A body of water
Small buildings surrounded by trees
Woman wearing a purple dress
Two people near elephants
A man wearing a hat
A woman wearing glasses
Leaves on the ground