Future of COCO and Discussion

Tsung-Yi Lin

Cornell NYC Tech
COCO Challenge – Accurate Object Detection
Beyond Detection

a cat laying on top of a wooden bench.
cat sleeping on a bench in a city area
a cat is curled up and sleeping on an old bench.
a cat is sitting outside on a bench.
an orange and white cat resting on a wooden bench.
Review: 2015

• Weekly meeting (10+ people)
• Captioning challenge (CVPR15)
• Detection challenge (ICCV15)
• Evaluation metrics
• Evaluation server
• Test set protocol
• Labeling 40k additional test images
Figure 1: Visual Semantic Role Labeling: We want to go beyond classifying the action occurring in the image to being able to localize the agent, and the objects in various semantic roles associated with the action.
SALICON

SALICON: Saliency in Context, Jiang et al., CVPR15

Visual Saliency
Visual Actions
Visual Role
Targeted visual Description
Visual Q&A
Grounded Visual Q&A
Structured Visual concepts to language
COCO-a

Describing Common Human Visual Actions in Images, Ronchi and Perona., BMVC 2015

http://www.vision.caltech.edu/~mronchi/projects/Cocoa/

Visual Saliency
Visual Actions
Visual Role
Targeted visual Description
Visual Q&A
Grounded Visual Q&A
Structured Visual concepts to language
v-COCO

Visual Semantic Role Labeling, Saurabh et al., Arxiv2015

Figure 1: Visual Semantic Role Labeling: We want to go beyond classifying the action occurring in the image to being able to localize the agent, and the objects in various semantic roles associated with the action.
Visual Madlibs

Visual Madlibs: Fill in the blank Description Generation and Question Answering, Yu et al., ICCV2015

1. This place is a park.
2. When I look at this picture, I feel competitive.
3. The most interesting aspect of this picture is the guys playing shirtless.
4. One or two seconds before this picture was taken, the person caught the frisbee.
5. One or two seconds after this picture was taken, the guy will throw the frisbee.
6. Person A is wearing blue shorts.
7. Person A is in front of person B.
8. Person A is blocking person B.
9. Person B is a young man wearing an orange hat.
10. Person B is on a grassy field.
11. Person B is holding a frisbee.
12. The frisbee is white and round.
13. The frisbee is in the hand of the man with the orange cap.
14. People could throw the frisbee.
15. The people are playing with the frisbee.
VQA

VQA: Visual Question Answering, Antol et al, ICCV15

Visual Saliency
Visual Actions
Visual Role
Targeted visual Description
Visual Q&A
Grounded Visual Q&A
Structured Visual concepts to language

What color are her eyes?
What is the mustache made of?
Visual7W: Grounded Question Answering in Images, Yuke et al., arxiv 2015

Where does this scene take place?
A) In the sea. ✓
B) In the desert.
C) In the forest.
D) On a lawn.

What is the dog doing?
A) Surfing. ✓
B) Sleeping.
C) Running.
D) Eating.

Why is there foam?
A) Because of a wave. ✓
B) Because of a boat.
C) Because of a fire.
D) Because of a leak.

What is the dog standing on?
A) On a surfboard. ✓
B) On a table.
C) On a garage.
D) On a ball.

Which paw is lifted?
VISUAL GENOME

Visual Genome (http://visualgenome.org)

Visual Saliency
Visual Actions
Visual Role
Targeted visual Description
Visual Q&A
Grounded
Visual Q&A
Structured
Visual concepts
to language
How can We Help You?

Figure 1: Visual Semantic Role Labeling: We want to go beyond classifying the action occurring in the image to being able to localize the agent, and the objects in various semantic roles associated with the action.

where does this scene take place?
A) In the sea ✓
B) In the desert
c) In the forest
d) On a beach

What is the dog doing?
A) Surfing ✓
B) Sleeping
C) Running
D) Eating

Which paw is lifted?

Why is there a wave?
A) Because of a wave ✓
B) Because of a boat
C) Because of a fire
d) Because of a leak

What is the dog standing on?
A) On a skateboard ✓
B) On a table
c) On a garage
d) On a ball

What color are her eyes?

What is the mustache made of?

How can We Help You?

- Annotation UIs
  
  [http://github.com/tylin/coco-ui](http://github.com/tylin/coco-ui)

- Evaluation
  
  [http://Codalab.org](http://Codalab.org)

- Funding
  
  MTurk
What’s Next?
Object Pose
Scene Structure

- Amodal segmentation
- Depth order
- “Stuff” segmentation

Semantic Amodal Segmentation, Zhu et al., 2015
Increase the Size of Dataset

- More categories?
- More same/different images?
- How many are enough?
Playground to Connect Domains

- Vision
- Language
- Commonsense
- Knowledge
The Detection Challenge Next Year

• Better annotation and detection file formats
• Better evaluation server
• What do you want to do next?

We’d love to hear from you!

cocodataset@outlook.com