Supplementary Material for
"Studying Relationships Between Human Gaze, Description, and Computer Vision"

Kiwon Yun[1], Yifan Peng[1], Dimitris Samaras[1], Gregory J. Zelinsky[2], Tamara L. Berg[1]

[1]Department of Computer Science, [2]Department of Psychology

Stony Brook University, Stony Brook, NY 11794, USA

{kyun, yipeng, samaras, tlberg}@cs.stonybrook.edu, {gregory.zelinsky}@stonybrook.edu

1 Selected scenes and objects from the SUN09 dataset

Selected scenes	bar, bathroom, bedroom, kitchen, dinning room, classroom, living room, office
Selected objects	bathtub, bed, bouquet, box, carpet/rug, cabinet/cupboard, chair/stool, curtain, desk, dishwasher, drawer, door, microwave/oven, person, picture/painting, plant , refrigerator, sofa, tv/screen, toilet, window

Table 1: Selected scene categories and common object categories from the SUN09

2   Examples of datasets

Figure 1 shows example images with corresponding fixations and descriptions from the SUN09 dataset. Figure 3 shows examples from the PASCAL dataset.
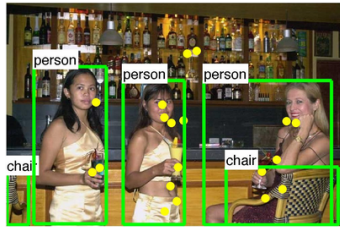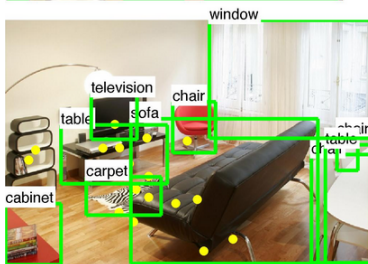

b_bar_bar_0386.jpg

Um Three females at a bars. Two of them standing one sitting. All of them have drinks in their hand and the bar is stocked fully.

Extracted words: females,bars,drinks,hand,bar
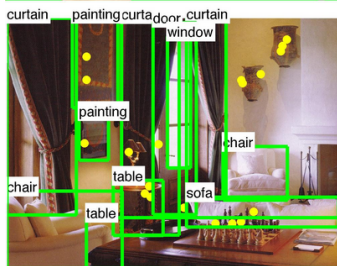**Extracted objects: person**
Extracted scenes: bar

l_living_room_indoor_0466.jpg

Living fru-/Living room with a zebra rug. Black leather couch in front of a flat screen TV.

Extracted words: room,zebra,rug,leather,couch,front,screen,tv
Extracted objects: carpet,rug,screen,sofa,television
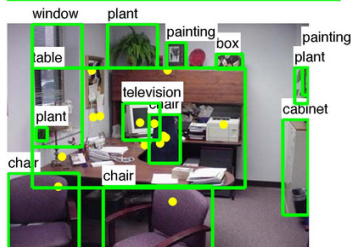**Extracted scenes:**

l_living_room_living49.jpg

A living room with two long windows both have dark blue curtains, uhh coffee table in the middle uh between chairs with chess table on it

Extracted words: living,room,windows,curtains,coffee,table,chairs,chess,table
Extracted objects: chair,curtain,desk,stool,table,window
**Extracted scenes: living**

o_office_emba_offices_before_move.jpg

It reminded me of my principals office in High School. Mmmm there was dark wood desks with dark wood cabinets in the back, white walls, blue chairs and burgundy chairs.

Extracted words: principals,office,high,school,wood,desks,wood,cabinets,back,walls,chairs,chairs
Extracted objects: cabinet,chair,cupboard,desk,stool,table
Extracted scenes: office
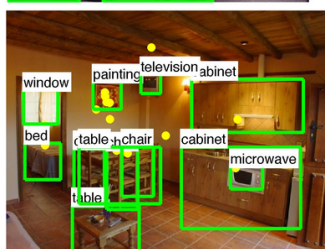
k_kitchen_dsc04183.jpg

Um a kitchen dining room that's bare with a bedroom off to the side and it has one microwave and one oven and that's all.

Extracted words: kitchen,room,bare,bedroom,side,microwave
Extracted objects: microwave,oven
**Extracted scenes: bedroom,dining,kitchen**

Figure 1: Examples of fixations and descriptions on the SUN09 dataset. Fixations are from a single subject, and marked as yellow dots. Descriptions are from the same subject. Green indicates ground truth bounding boxes.
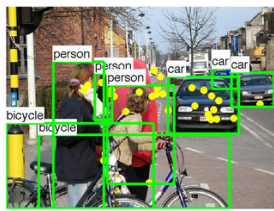
2008_003140.jpg

Two bicyclists and a pedestrian are waiting to cross the road. (bicyclists,pedestrian,road => person)
Two bicyclists wait to cross a busy road. (bicyclists,road => person)
Two blonde girls attempt to cross a busy street in town with their bicycles. (girls,street,town,bicycles => bicycle,person)
Two women watching the street with their bikes to see when they may cross. (women,street,bikes => bicycle,motorbike,person)
Two women with bicycles are trying to cross the street. (women,bicycles,street => bicycle,person)

Extracted words: bicycles,bicyclists,bikes,girls,pedestrian,road,street,town,women
Extracted objects: bicycle,person
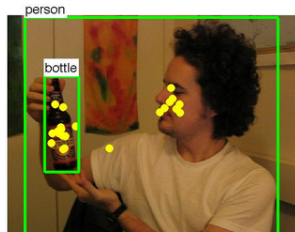
2008_005812.jpg

A man is reading the label on a beverage bottle. (man,label,beverage,bottle => bottle,person)
A man looking at the bottle of beer that he is holding. (man,bottle,beer => bottle,person)
The man in a white tee shirt is holding a beer bottle and looking at it. (man,tee,shirt,beer,bottle => bottle,person)
The scraggly haired man is holding up and admiring his bottle of beer. (man,bottle,beer => bottle,person)
Young man with curly black hair holding a beer bottle. (man,hair,beer,bottle => bottle,person)

Extracted words: beer,beverage,bottle,hair,label,man,shirt,tee
Extracted objects: bottle,person
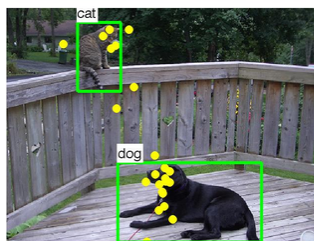
2008_005181.jpg

A black dog laying on the deck. (dog,deck => dog)
A cat sits on a railing, looking down at a black dog on a leash. (cat,railing,dog,leash => cat,dog)
A cat sits on top of a wooden railing as a large black dog looks up at it. (cat,top,railing,dog => cat,dog)
a dog on the floor of a patio looks at a cat on the fence (dog,floor,patio,cat,fence => cat,dog)
Black dog with red leash on staring up at the cat on the fence. (dog,leash,cat,fence => cat,dog)

Extracted words: cat,deck,dog,fence,floor,leash,patio,railing,top
Extracted objects: cat,dog
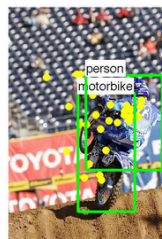
2008_000328.jpg

A man riding a dirt bike over a ridge. (man,dirt,bike,ridge => bicycle,motorbike,person)
A professional motorbike riding in a professional stadium. (motorbike,riding,stadium => motorbike)
A rider on a dirt bike. (rider,dirt,bike => bicycle,motorbike)
Blue motorcycle rider roaming a dirt ramp in front of fans. (motorcycle,rider,dirt,ramp,front,fans => motorbike)
The man wearing a blue jumpsuit riding a dirt bike tops the dirt hill. (man,jumpsuit,dirt,bike,dirt,hill => bicycle,motorbike,person)

Extracted words: bike,dirt,fans,front,hill,jumpsuit,man,motorbike,motorcycle,ramp,rider,ridge,riding,stadium
Extracted objects: motorbike,person
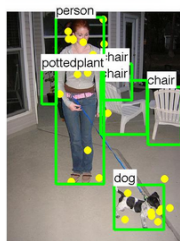
2008_005111.jpg

a girl holding a black white dog with a blue dog chain (girl,dog,dog,chain => dog,person)
A women holds a small dog on a leash with chairs in the background. (women,dog,leash,chairs,background => chair,dog,person)
Cute redhead walking a puppy (Cute,redhead =>)
The girl stands holding leash attached to a small dog. (girl,leash,dog => dog,person)
The small dog protects its owner. (dog,owner => dog)

Extracted words: Cute,background,chain,chairs,dog,girl,leash,owner,redhead,women
Extracted objects: chair,dog,person
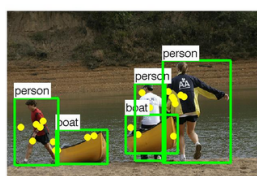
2008_000195.jpg

A family getting ready to go canoe on a small lake. (family,canoe,lake => boat,person)
Three people are by the water with yellow canoes (people,water,canoes => person)
Three people preparing to canoe. (people,canoe => boat,person)
Two people are carrying a yellow canoe down to the water, while a man in red and black walks away from a canoe already near the water.
(people,canoe,water,man,walks,canoe,water => boat,person)
Two people carrying a canoe into a river while a third person walks out of the picture. (people,canoe,river,person,picture => boat,person)

Extracted words: canoe,canoes,family,lake,man,people,person,picture,river,walks,water
Extracted objects: boat,person

Figure 2: Examples of fixations and descriptions on the PASCAL dataset. Fixations are from all three subjects, and marked as yellow dots. Descriptions are collected from different subjects by Amazon Mechanical Turk. Green indicates ground truth bounding boxes.

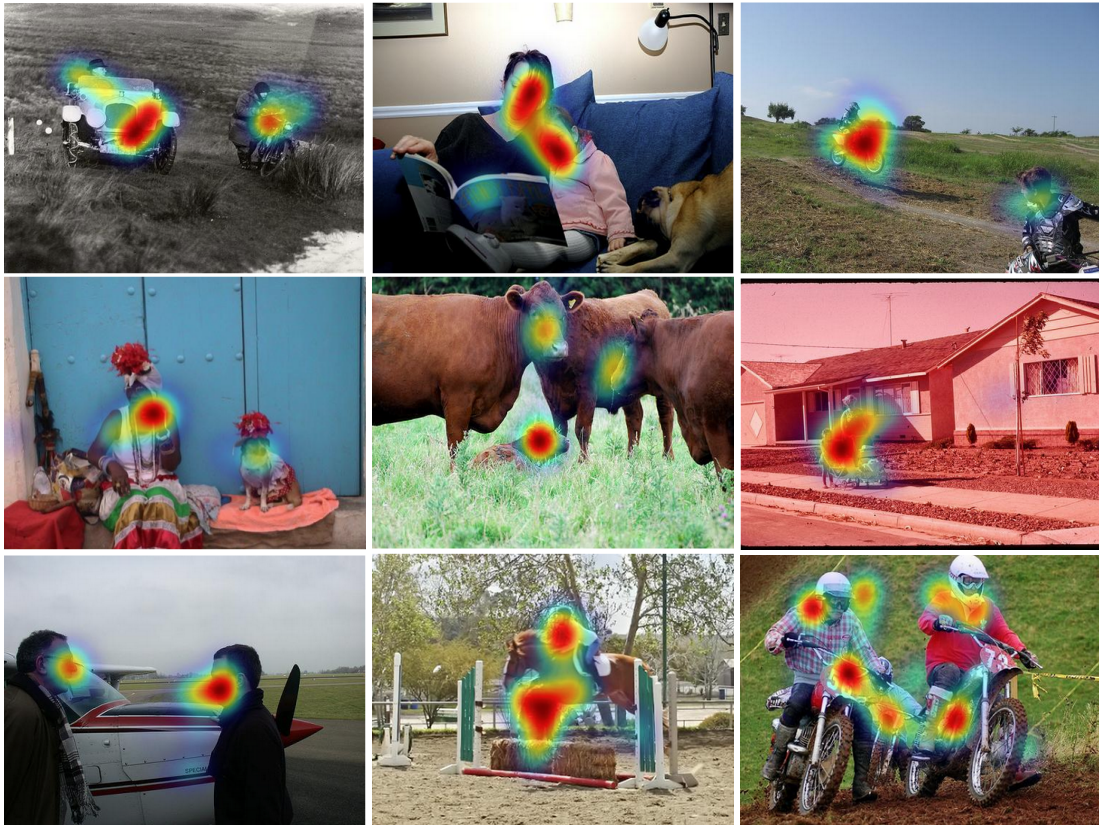# 3 Examples of fixation density maps on the PASCAL dataset

Figure 3: Examples of fixation density maps on the PASCAL dataset

4 Average fixation density maps

Fixation patterns tend to vary by object category. Figures 4, 5, 6 show how gaze patterns differ between different object categories on the PASCAL dataset. Figures 7, 8, 9 show the results on the SUN09 dataset. Note that categories present in both of our datasets e.g., person (4a, 7g), chair (5e, 7c), table (5c, 7f) and sofa (5d, 7d), display similar fixation patterns. Figure 10 shows the average fixation density map on the common object categories between two datasets.

| (a) person | (b) bird | (c) horse | (d) cat |

| (e) dog | (f) cow | (g) sheep |
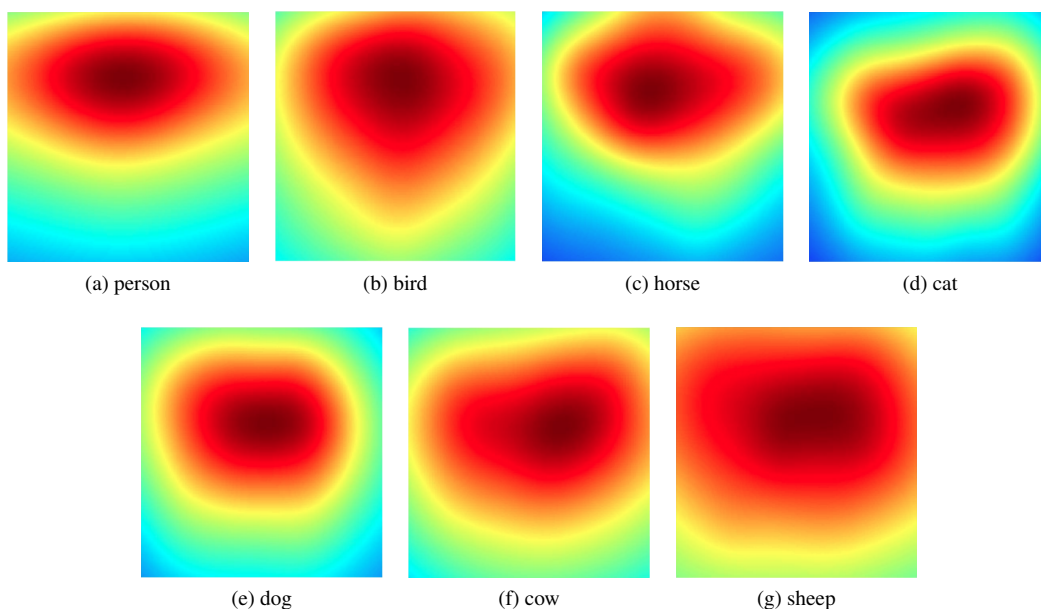
Figure 4: Average fixation density maps on the PASCAL dataset. When people look at animals, e.g. person or horse, they tend to look near the animal's head.

| (a) bicycle | (b) motorbike | (c) diningtable |

| (d) sofa | (e) chair |
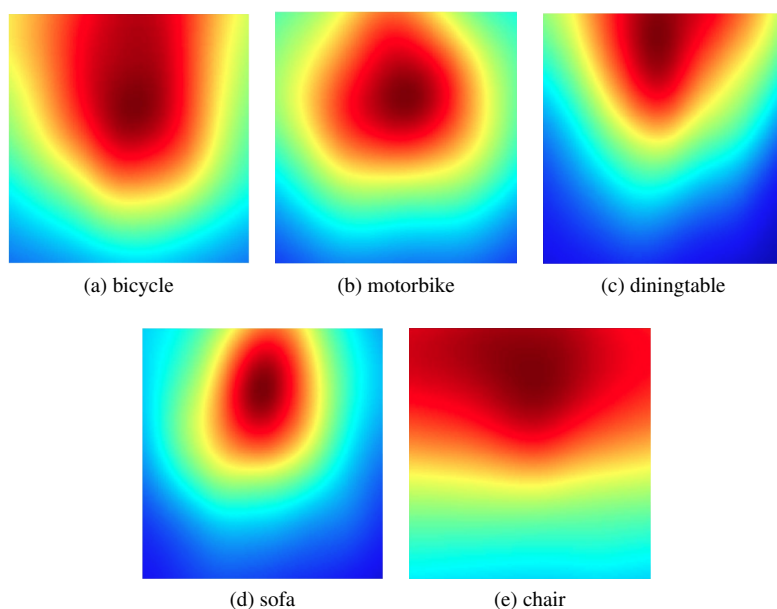
Figure 5: Average fixation density maps on the PASCAL dataset. For some categories such as bicycle and chair, humans are usually sitting on them pulling many fixations toward the top/middle of the bounding box. Similarly, there are often objects on dining tables.
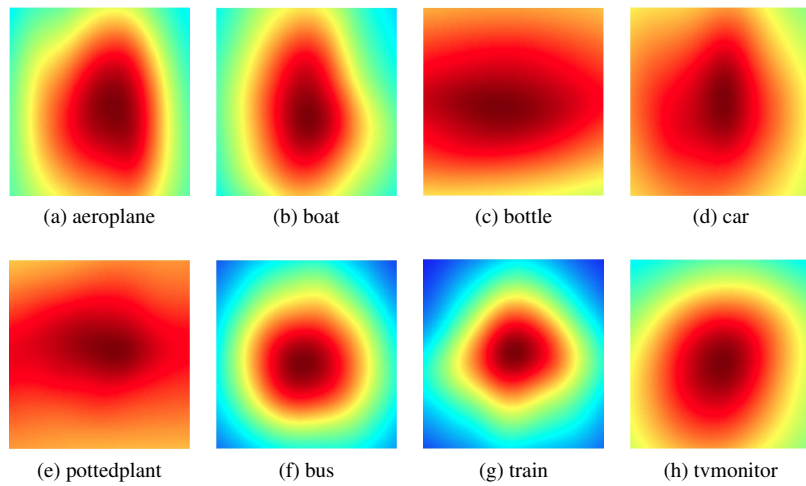
Figure 6: Average fixation density maps on the PASCAL dataset. For all the other categories like tvmonitor, people mainly look at the center of the object.
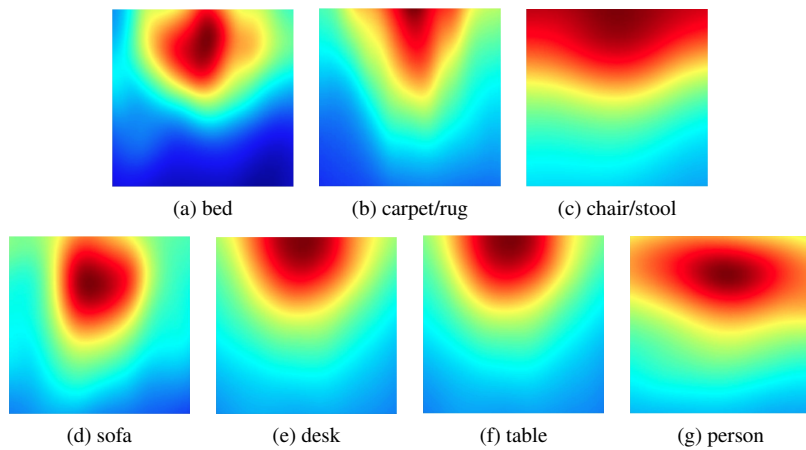

Figure 7: Average fixation density maps on the SUN09 dataset. For some categories such as bed and carpet, humans are usually sitting on them pulling many fixations toward the top/middle of the bounding box. Similarly, there are often objects on tables. When people look at person, they tend to look near his/her head.
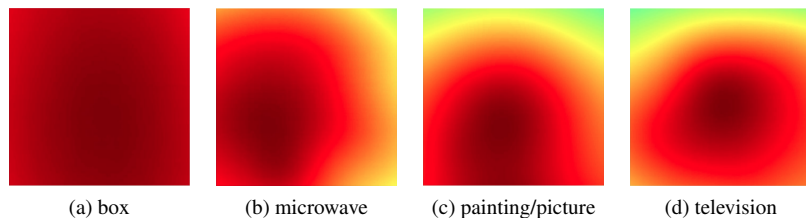

Figure 8: Average fixation density maps on the SUN09 dataset. For some categories like television and box, people mainly look at the center of the object.
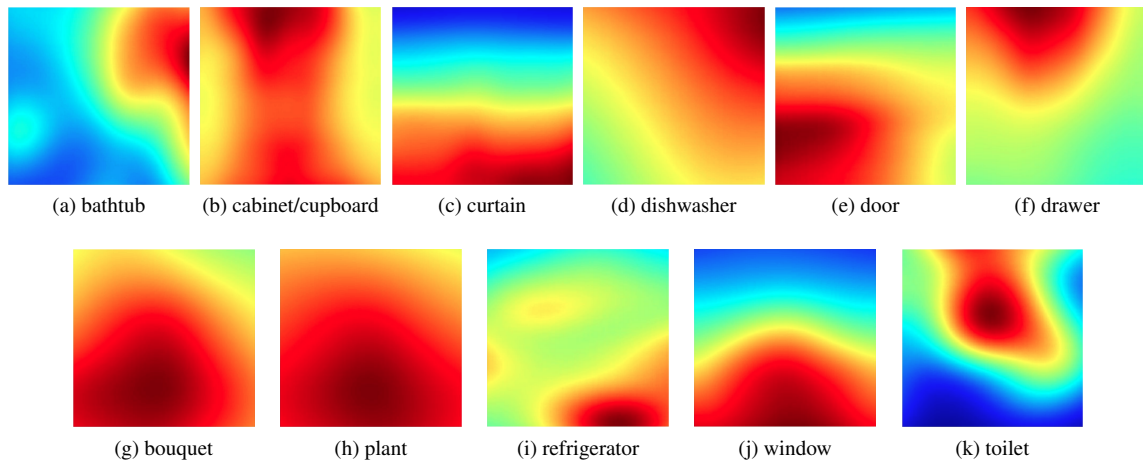
Figure 9: Average fixation density maps on the SUN09 dataset. For all the other categories like bathtub, curtain or window, it is difficult to say what its pattern is. The reason is that people do not often fixate on these objects since they are not important objects in the image. Fixations on these objects sometimes depend on the other objects they are close to.
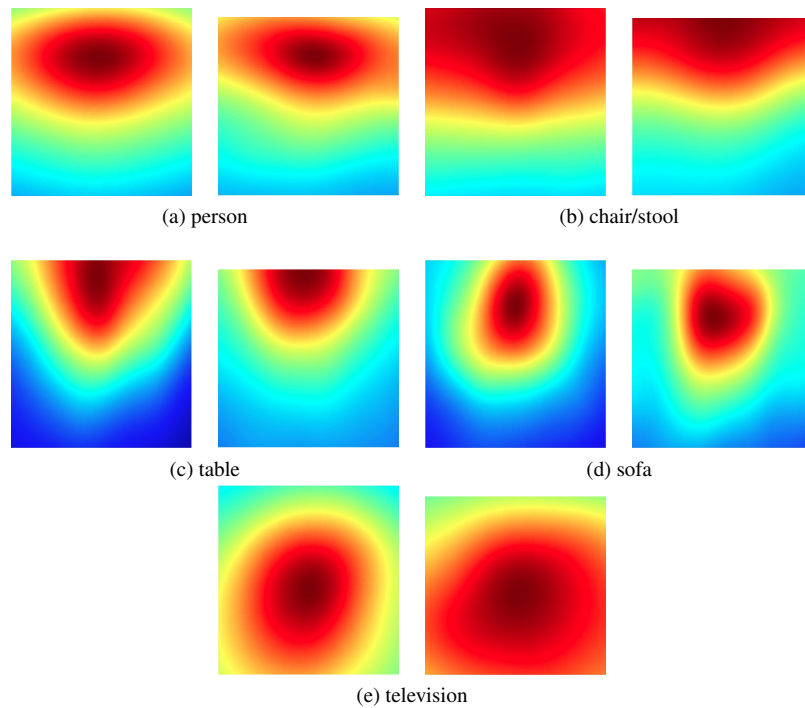

Figure 10: Categories present in both of our test datasets such as person, chair, table and sofa have similar fixation patterns. For each sub-figures, the left is the average fixation density map on the PASCAL dataset, while the right is the average fixation density map on the SUN09 dataset.

5 The average normalized percentages of fixations falling into detection boxes

Figure 11: Average normalized percentage of fixations falling into detection boxes, grouped into 10 bins corresponding to detection boxes scored top 10%, top 10%-20%, etc. In the PASCAL dataset, people tend to look more at detection boxes from bicycle, bird, boat, car, person, tvmonitor detectors
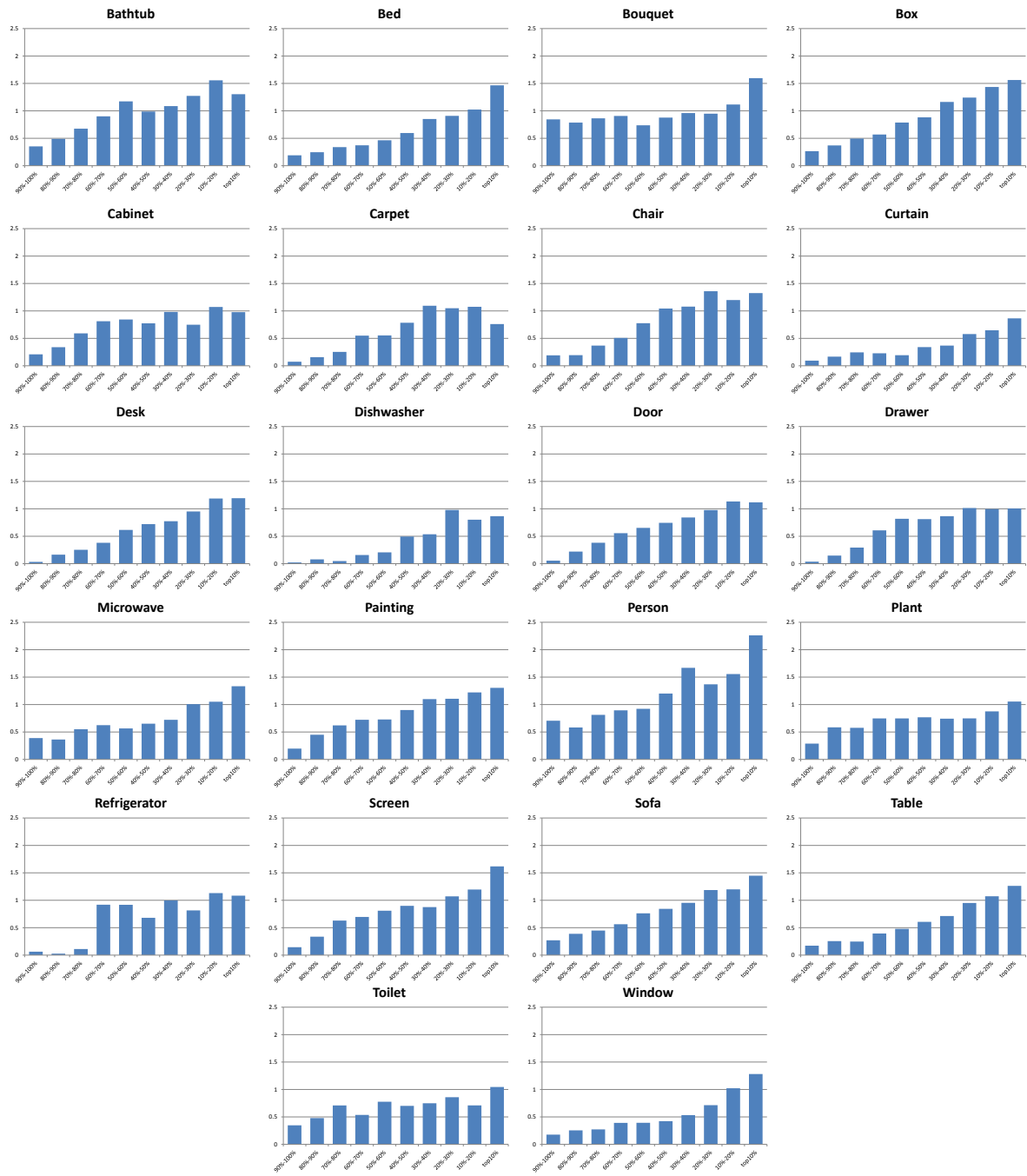
Figure 12: Average normalized percentage of fixations falling into detection boxes in the SUN09 dataset. In the SUN09 dataset, people tend to look more at detection boxes from bed, box, curtain, desk, microwave, person, screen, sofa, table, window detectors.

9

6 The percentage of bounding box type on detection boxes with default thresholds

Figure 13a is Figure 7 in the paper, and Figure13b depicts the same analysis for the SUN09 dataset.
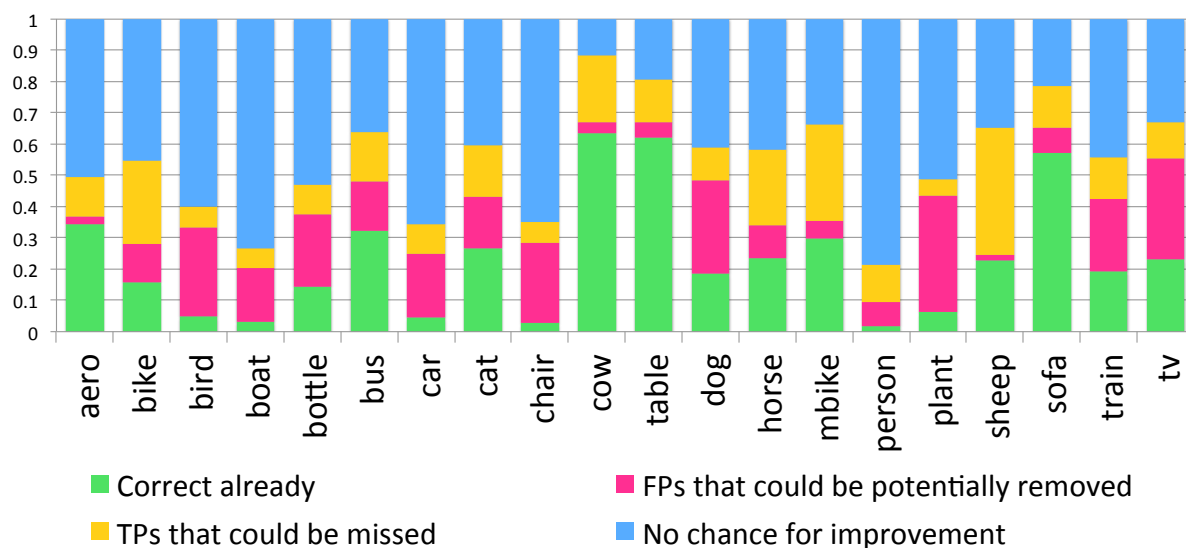
(a) PASCAL
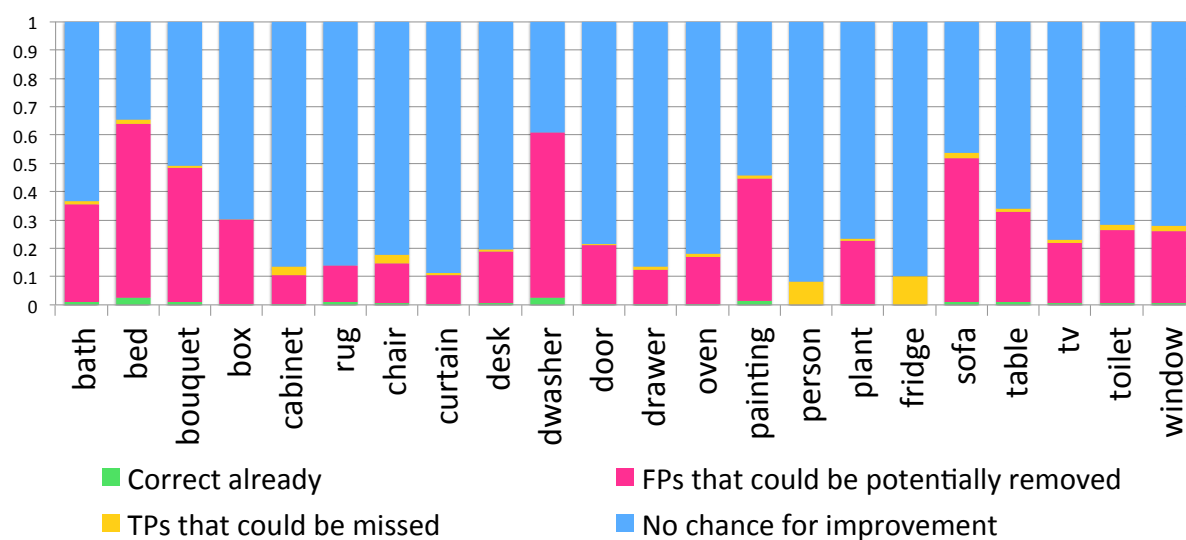
(b) SUN09

Figure 13: Percentage of bounding box type on all detection boxes from each detector with default thresholds in the PASCAL and the SUN09 dataset

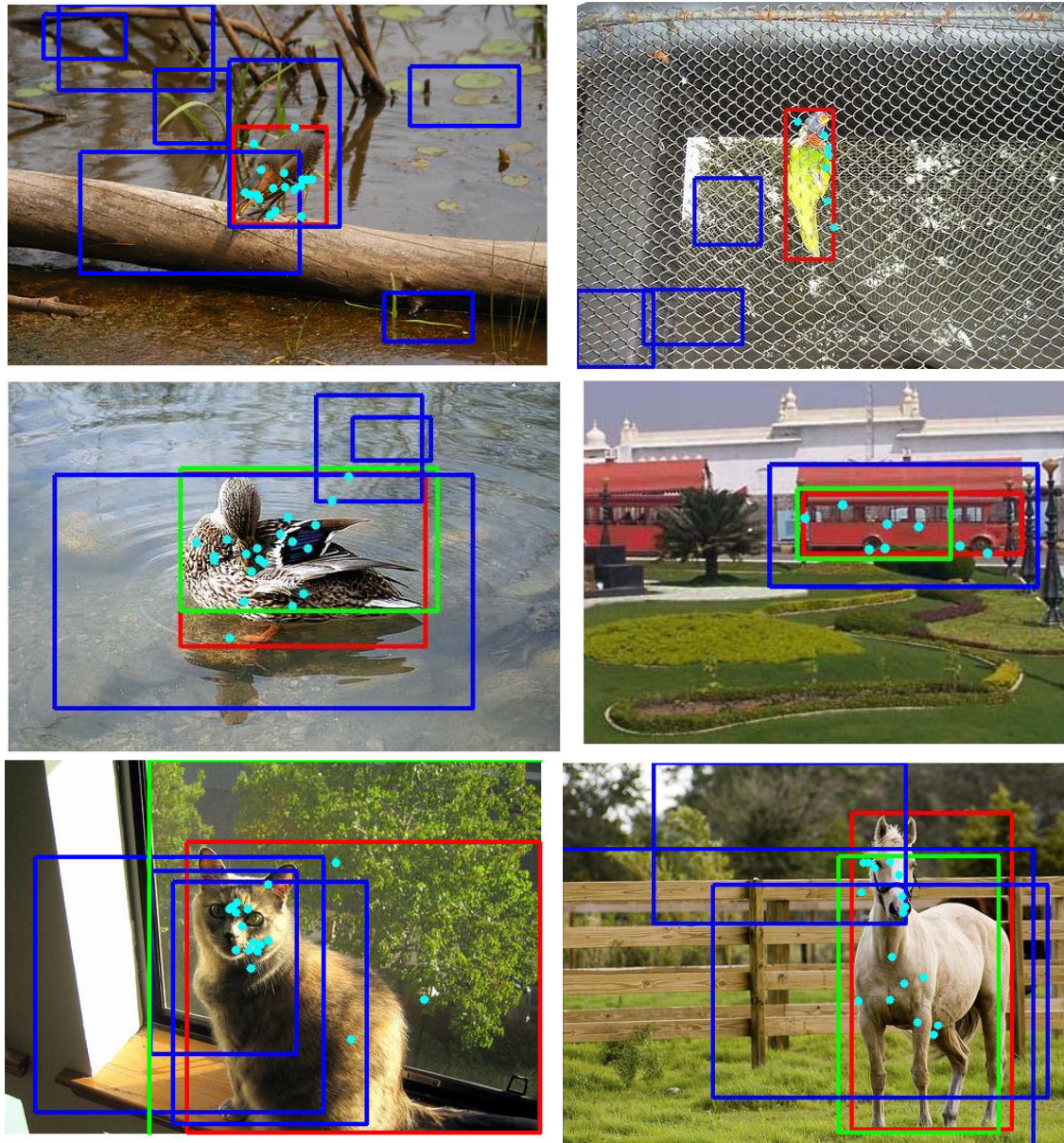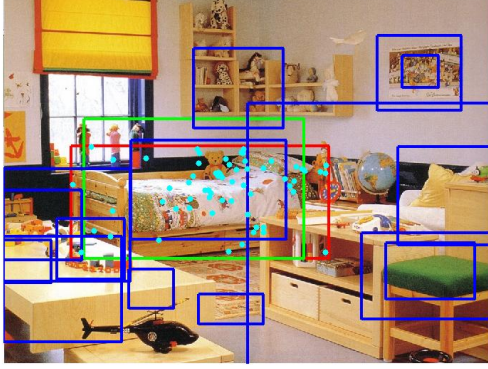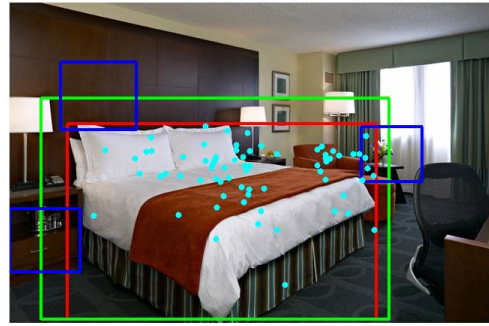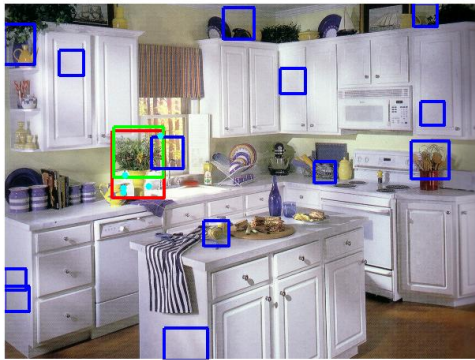# 7 Examples where fixations can improve detections

Figure 14: Examples of object detection results that could potentially be improved by incorporating fixation information in the PASCAL dataset. Red bounding boxes show ground truth(GT) bounding boxes. Green bounding boxes show true positive detection. Blue bounding boxes show false positive detection. Cyan dots denote fixations inside a GT box. Row 1: No chance for improvement since there is no detection close enough to ground truth. Row 2: False positive bounding boxes shared more fixations in GT box than TP boxes. Using fixation would select the false positive instead. Row 3: cases that use fixation would be helpful. TP boxes take more fixations in GT boxes than all other FP boxes. FP boxes can be discarded.
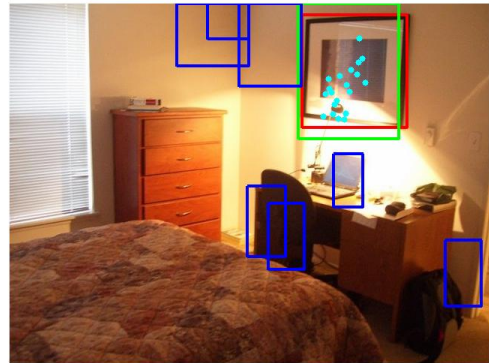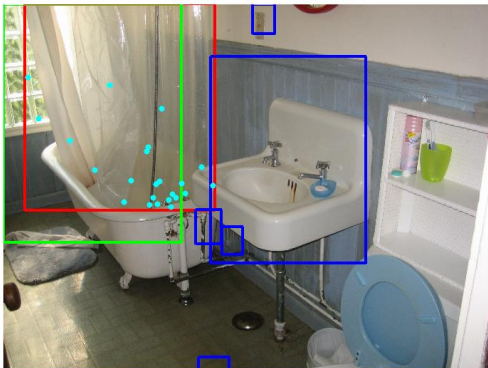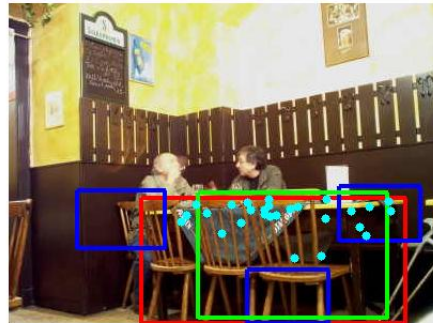
(a) bed

(b) bed

(c) bouquet

(d) painting

(e) curtain

(f) table

Figure 15: Examples of object detection results that could potentially be improved by incorporating fixation information in the SUN09 dataset. Red bounding boxes show ground truth(GT) bounding boxes. Green bounding boxes show true positive detection. Blue bounding boxes show false positive detection. Cyan dots denote fixations inside a GT box. For example, in (a) there is a ground truth box for bed. A TP box for bed has more fixation scores than all FP boxes. All FP boxes can be potentially discarded.