The Effect of Improving Annotation Quality on Object Detection Datasets: A Preliminary Study — *Supplementary Material*

Jiaxin Ma
OMRON SINIC X Corp.
jiaxin.ma@sinicx.com

Yoshitaka Ushiku
OMRON SINIC X Corp.
yoshitaka.ushiku@sinicx.com

Miori Sagara
Baobab Inc.
sagara@baobab-trees.com

1. Complete annotation guidelines

In the main paper, we only showed our annotation guideline of *chair* in MS COCO. In this supplementary material, we provide our complete annotation guidelines (two datasets, ten categories in total) for a better reference. Note that the descriptions provided in this section is abbreviated. The original guideline documents include more explanations and illustration samples.

1.1. COCO guideline

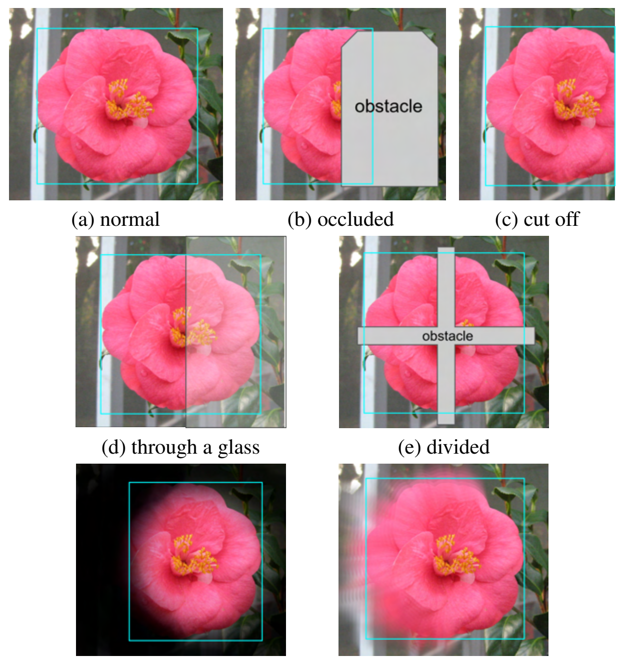
For MS COCO dataset, we have the following general instructions.

- When there are a large number of target objects in an image, annotate them as possible as you can. However, do not annotate objects that are too small or too blurred.
- Refer to Fig. 1 for examples of the proper enclosure of bounding boxes in various cases.
- If the target to be annotated is not shown in the image at all, annotate a small bounding box at the bottom right of the image with a skip tag.

The instructions related to the five categories involved in this study are shown below.

1.1.1 Traffic light

1. Definition: A device that conveys the priority of traffic for automobiles, bicycles, pedestrians, and trains, etc., and is installed at intersections, railways, railroad crossings, etc., for the purpose of smooth driving while ensuring traffic safety. Its three colors (green, yellow, and red) are internationally standardised. Red means stop and green means safe to move forward.
2. Positive examples, see Fig. 2
   (a) road traffic light
   (b) a red light installed at the railroad crossing
   (c) railroad traffic light



(a) normal (b) occluded (c) cut off (d) through a glass (e) divided (f) dark (g) blurred

Figure 1. Proper enclosure of bounding boxes in various cases

   (d) pedestrian / bicycle traffic light
3. Negative examples, see Fig. 3
   (a) traffic sign / road sign
   (b) general sign / signboard
4. Additional instructions, see Fig. 4
   (a) Annotate a traffic light that is in side view, in back view, blurred, or partially cut off from the image, if it is identifiable.
   (b) Do not enclose pillars, electric wires, or accessories that support the traffic light.
   (c) If a traffic light only shows its shining part (such as at night) but is still identifiable, annotate the shining part (unnecessary to estimate the size of

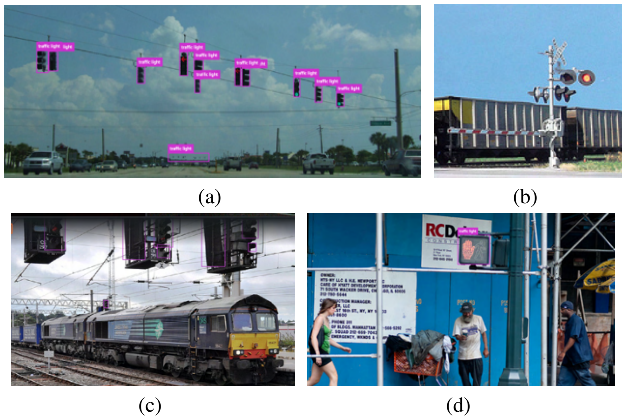


Figure 2. Positive examples of *traffic light*

Figure 3. Negative examples of *traffic light*

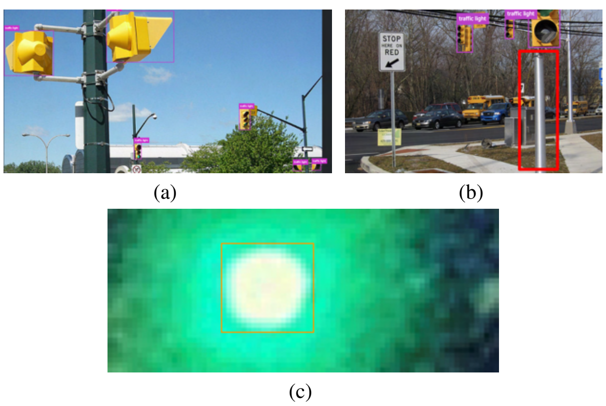


Figure 4. Additional instructions for *traffic light*

the entire traffic light).

(d) Do not annotate if you cannot clearly identify.

1.1.2 Cup

1. Definition: A small container used for drinking from. The material (glass, pottery, paper, plastic, etc.) does not matter as long as the shape and the function satisfies.

2. Positive examples, see Fig. 5

   (a) cup with a handle
   (b) rocks glass
   (c) tumbler

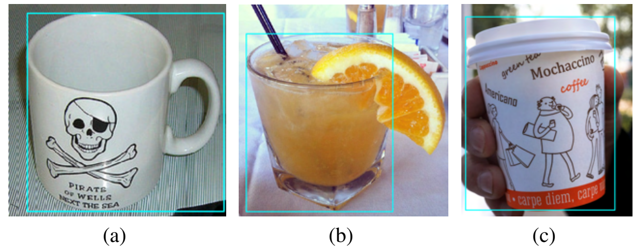


Figure 5. Positive examples of *cup*

Figure 6. Negative examples of *cup*

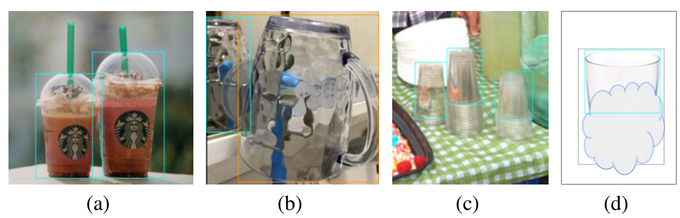


Figure 7. Additional instructions for *cup*

3. Negative examples, see Fig. 6

   (a) glass with a leg (wine glass, etc.)
   (b) canister mug
   (c) jar with a lid such as a jam jar
   (d) wide-mouth bowl such as a soup mug
   (e) painting or illustration

4. Additional instructions, see Fig. 7

   (a) Do not enclose straws and spoons inserted in cups.
   (b) Annotate cups reflected in a mirror or glass.
   (c) Do not separately annotate stacked cups if you cannot clearly identify the boundary.
   (d) If you cannot tell if a glass has a leg or not due to occlusion, annotate it as a cup.

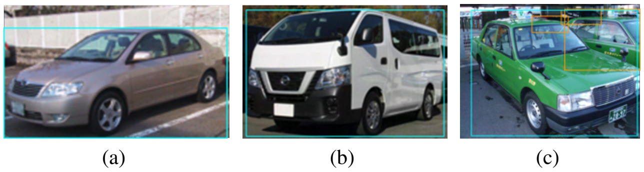


Figure 8. Positive examples of *car*

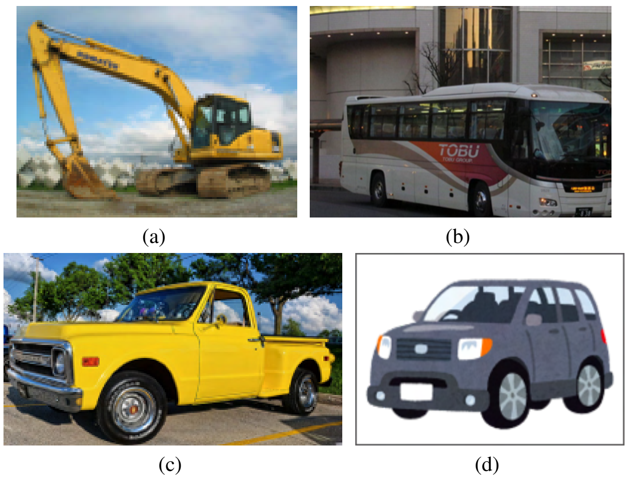


Figure 9. Negative examples of *car*

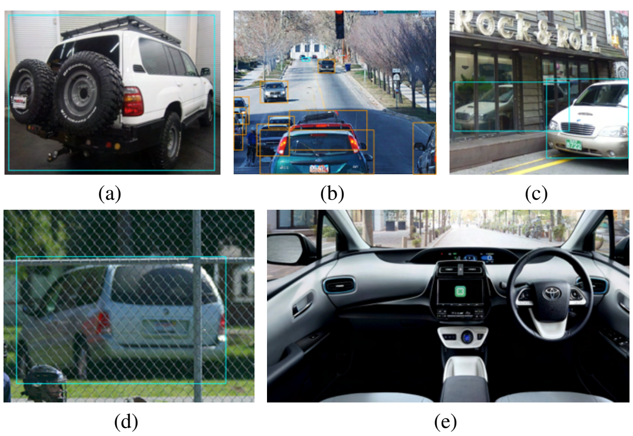


Figure 10. Additional instructions for *car*

1.1.3 Car

1. Definition: A vehicle that can be identified as a passenger car (which is a vehicle with a capacity of 10 people or less for the purpose of moving).

2. Positive examples, see Fig. 8
   (a) passenger car
   (b) minivan
   (c) taxi

3. Negative examples, see Fig. 9

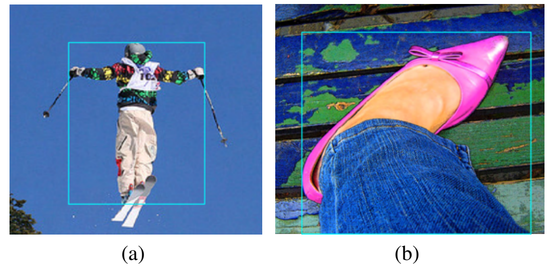


Figure 11. Positive examples of *person*

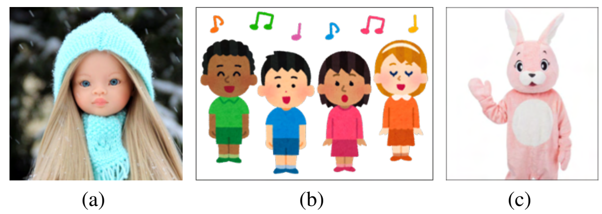


Figure 12. Negative examples of *person*

   (a) heavy equipment
   (b) bus
   (c) truck
   (d) painting or illustration
   (e) toy or model

4. Additional instructions, see Fig. 10
   (a) Enclose equipment that is rarely removed from the car, such as a rear spare tire. Do not enclose equipment that can be easily removed.
   (b) Annotate a car that is small, hidden, or partially cut off from the image, if it is identifiable.
   (c) Annotate a car reflected in a mirror or glass / a car shown on TV or poster.
   (d) Annotate a car through a fence or glass.
   (e) Do not annotate car parts and interiors taken from inside the car.

1.1.4 Person

1. Definition: human, or a part of the human body

2. Positive examples, see Fig. 11
   (a) a real person
   (b) a part of the human body

3. Negative examples, see Fig. 12
   (a) doll or statue
   (b) painting or illustration
   (c) full-body costume

4. Additional instructions, see Fig. 13

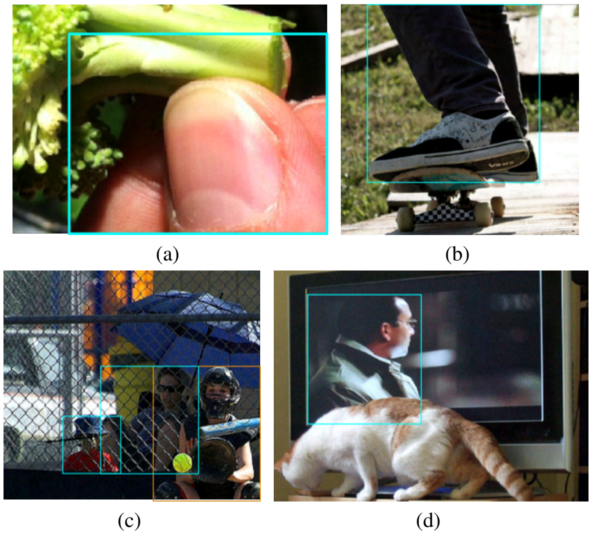


Figure 13. Additional instructions for *person*

(a) Annotate a person that is small, blurred, or only showing a part of body, if it is identifiable.
(b) Enclose clothes, shoes, and hat worn by a person. Do not enclose bags and items held in hands.
(c) Annotate a person through a fence or glass.
(d) Annotate a person reflected in a mirror or glass / a person shown on TV or poster.

1.1.5 Chair

The same as described in the main paper.

1.2. Open Images guideline

For Google Open Images dataset, we have the following general instructions.

- Annotate small sized target if it is identifiable. However, if there are numerous targets in one image, annotate 10–15 targets per category. This number is only a recommendation and can be increased or decreased at the discretion of the annotator.

- Refer to Fig. 1 for examples of the proper enclosure of bounding boxes in various cases.

- If the target to be annotated is not shown in the image at all, annotate a small bounding box at the bottom right of the image with a skip tag.

The instructions related to the five categories involved in this study are shown below. The format is changed from COCO guideline because they are not created together. For example, in Open Images guideline it does not include a "definition" part.

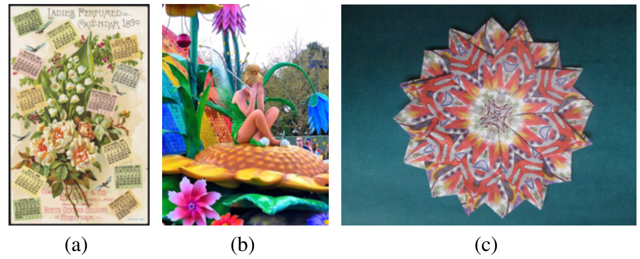


Figure 14. Negative examples of *flower*

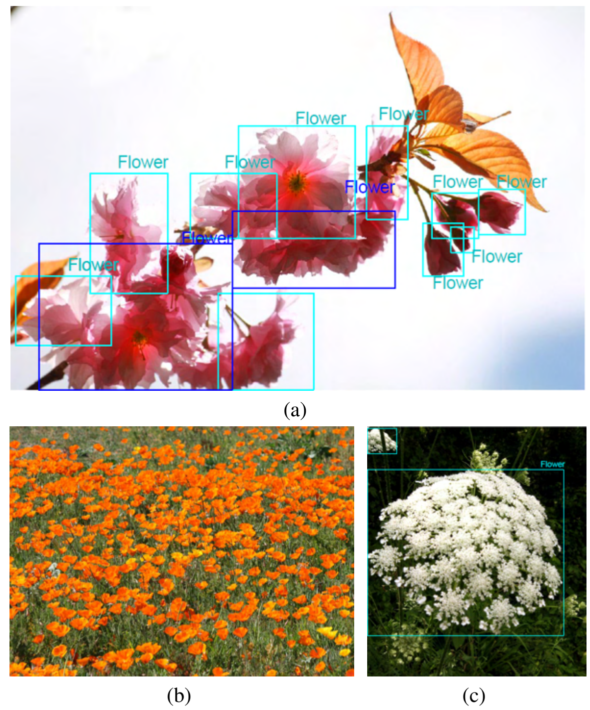


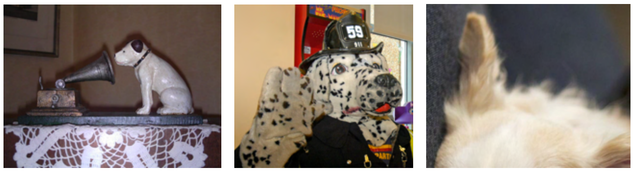
Figure 15. Additional instructions for *flower*

1.2.1 Flower

1. Positive examples
   (a) real flowers

2. Negative examples, see Fig. 14
   (a) painting or illustration
   (b) artificial flower
   (c) paper flower
   (d) buds that have not yet bloomed

3. Additional instructions, see Fig. 15
   (a) If flowers are overlapped and the boundary is unclear, please infer the boundary. However, if overlapped flowers are more than five and the boundary is unclear, enclose the entire area with a single bounding box.



(a) (b) (c)

Figure 16. Negative examples of *dog*

(a) (b) (c)

Figure 17. Negative examples of *car*

   (b) If flowers are clustered all over the image, enclose the entire image with a single bounding box.
   (c) If multiple small flowers are crowded and look like one big flower, enclose them with a single bounding box.

1.2.2 Dog

1. Positive examples
   (a) a real dog
   (b) a dog that is small or partially cut off from the image
2. Negative examples, see Fig. 16
   (a) dog decoration
   (b) dog costume / plush toy
   (c) an animal that is difficult to identify
   (d) painting or illustration
   (e) a dog in poster / mirror
3. Additional instructions
   (a) Enclose the clothes, hat, and collar that are worn by the dog.
   (b) Do not enclose the string that is connected to the collar.

1.2.3 Car

1. Positive examples
   (a) a vehicle that can be identified as a passenger car (which is a vehicle with a capacity of 10 people or less for the purpose of moving)
   (b) a car that is small or partially cut off from the image

(a) (b) (c)

Figure 18. Negative examples of *person*

   (c) racing car
2. Negative examples, see Fig. 17
   (a) model / toy / CG car
   (b) a car shown on TV / poster / illustration
   (c) a vehicle that cannot be identified as a passenger car
   (d) a heavy equipment / bus / truck that is not a passenger car
3. Additional instructions
   (a) Enclose equipment that is rarely removed from the car, such as a rear spare tire.
   (b) Do not enclose equipment that can be easily removed such as surfboards, bicycles, etc.

1.2.4 Person

1. Positive examples
   (a) a real person that shows more than 20% of the whole body
   (b) a person that is small or partially cut off from the image
2. Negative examples, see Fig. 18
   (a) a person in newspaper
   (b) statue, doll, painting, or illustration
   (c) a body part less than 20% of the whole body
   (d) a person in poster / mirror
3. Additional instructions
   (a) Enclose clothes, shoes, and hat that are worn by the person.
   (b) Do not enclose items in the person's hands.

1.2.5 Building

1. Positive examples
   (a) a real building, which is a construction that consists of roofs, pillars, and walls for people to work and live.

(a)

(b)

(c)

Figure 19. Negative examples of *building*

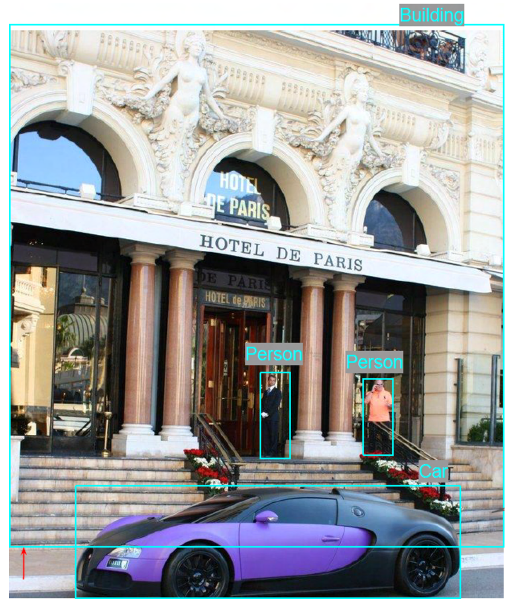
(b) a building that is small or partially cut off from the image

2. Negative examples, see Fig. 19

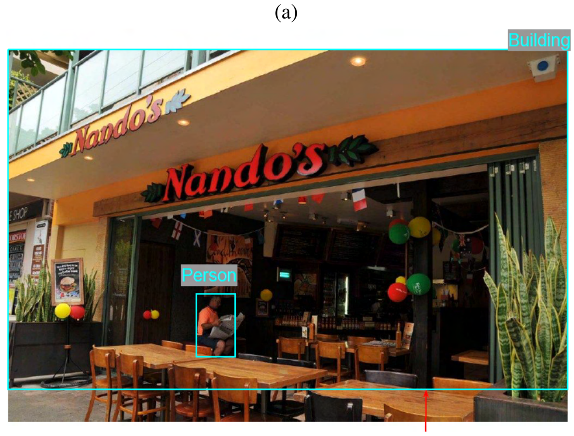
   (a) a scene taken from inside the building
   (b) model, painting, illustration, or poster
   (c) mobile home

3. Additional instructions, see Fig. 20

   (a) Enclose walls, roofs, pillars, and other parts that are integrated with the building (e.g., stairs and slopes in front of the entrance).
   (b) Do not enclose independent fences, stairs, rain shelters, and other items that are not integrated with the building.



(a)



(b)

Figure 20. Additional instructions for *building*

2. Qualitative results

This section displays some results of detected objects in both COCO and Open Images.

Figure 21, 22, and 23 show some results on COCO. The models trained using our annotations achieve better detection accuracy, especially for small objects, as shown in Fig. 21. Figure 22 also shows mutually occluded persons can be precisely detected by reannotating the training data. On the other hand, even the reannotated training data cannot let the models detect all small objects, as shown in Fig. 23. Figure 23 also shows that detecting some object such as a small traffic light and mutually occluded chairs is challenging for the reaanotated data.

Figure 24, 25, and 26 show some results on Open Images. The models trained using our annotations achieve better detection accuracy, especially for grouped objects, as shown in Fig. 24. Figure 25 also shows crowded persons and flowers can be precisely detected by reannotating the training data. On the other hand, detecting all flowers separately is still challenging for reannotated Open Images, as shown in Fig. 26. Figure 26 also shows that detecting some object such as even a large cars and flowers and grouped flowers and buildings is challenging for the reanotated data.

Qualitatively, we found that the results of the reannotated datasets, both COCO and Open Images, are generally more acceptable than the original one. This conclusion seems counter-intuitive when the quantitative results in the main paper showed a performance drop on the reannotated COCO. However, it is actually possible when the task difficulty is increased in the reannotated COCO dataset.

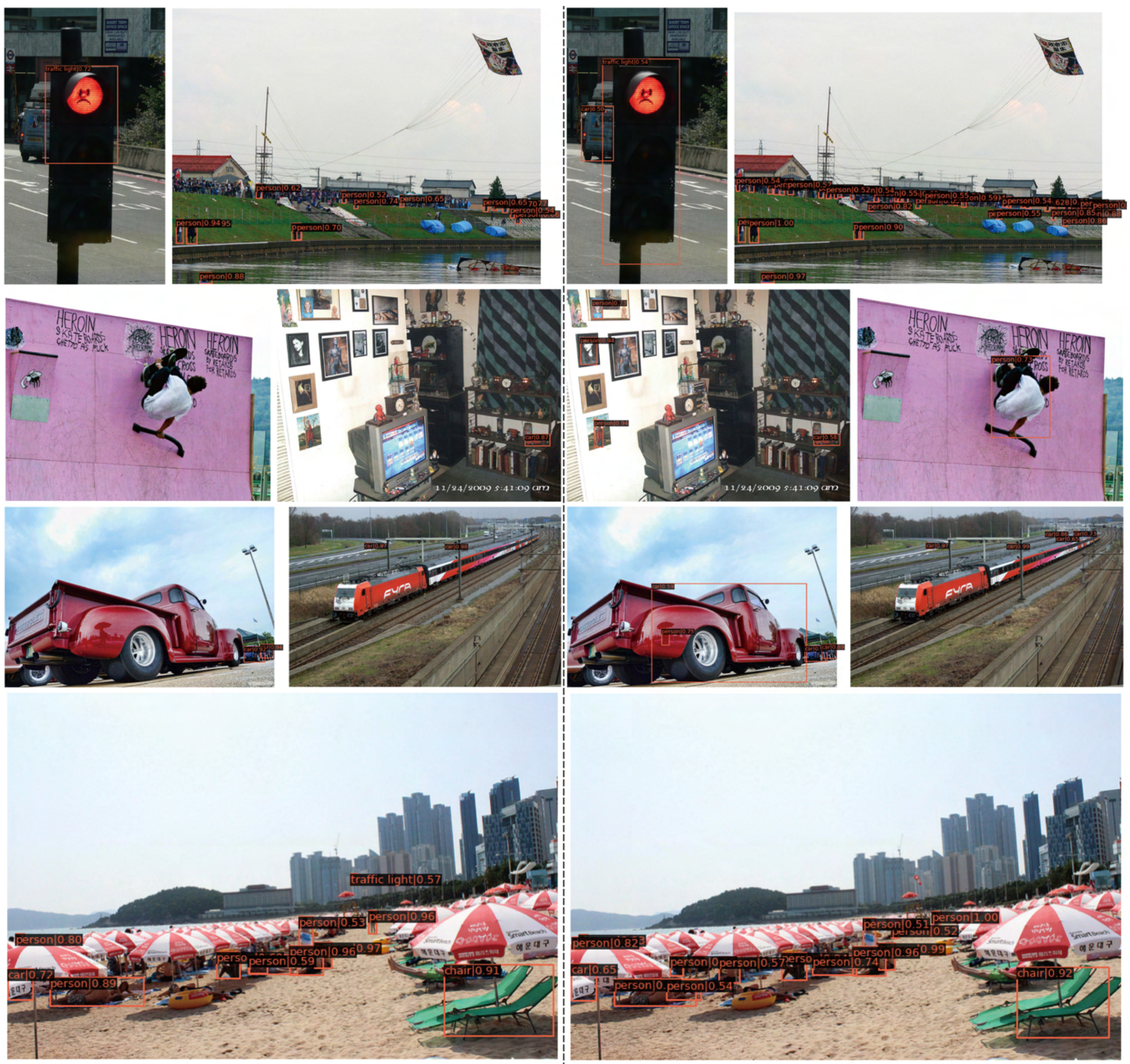


Figure 21. Results of detected objects in COCO. The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where the models trained using our annotations achieve better detections.

Figure 22. Results of detected objects in COCO (contd.). The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where the models trained using our annotations achieve better detections, especially for crowded objects.

Figure 23. Results of detected objects in COCO (contd.). The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where (1) the models trained using our annotations fail to detect some objects and (2) the models trained using either set of annotations fail to detect some objects properly.

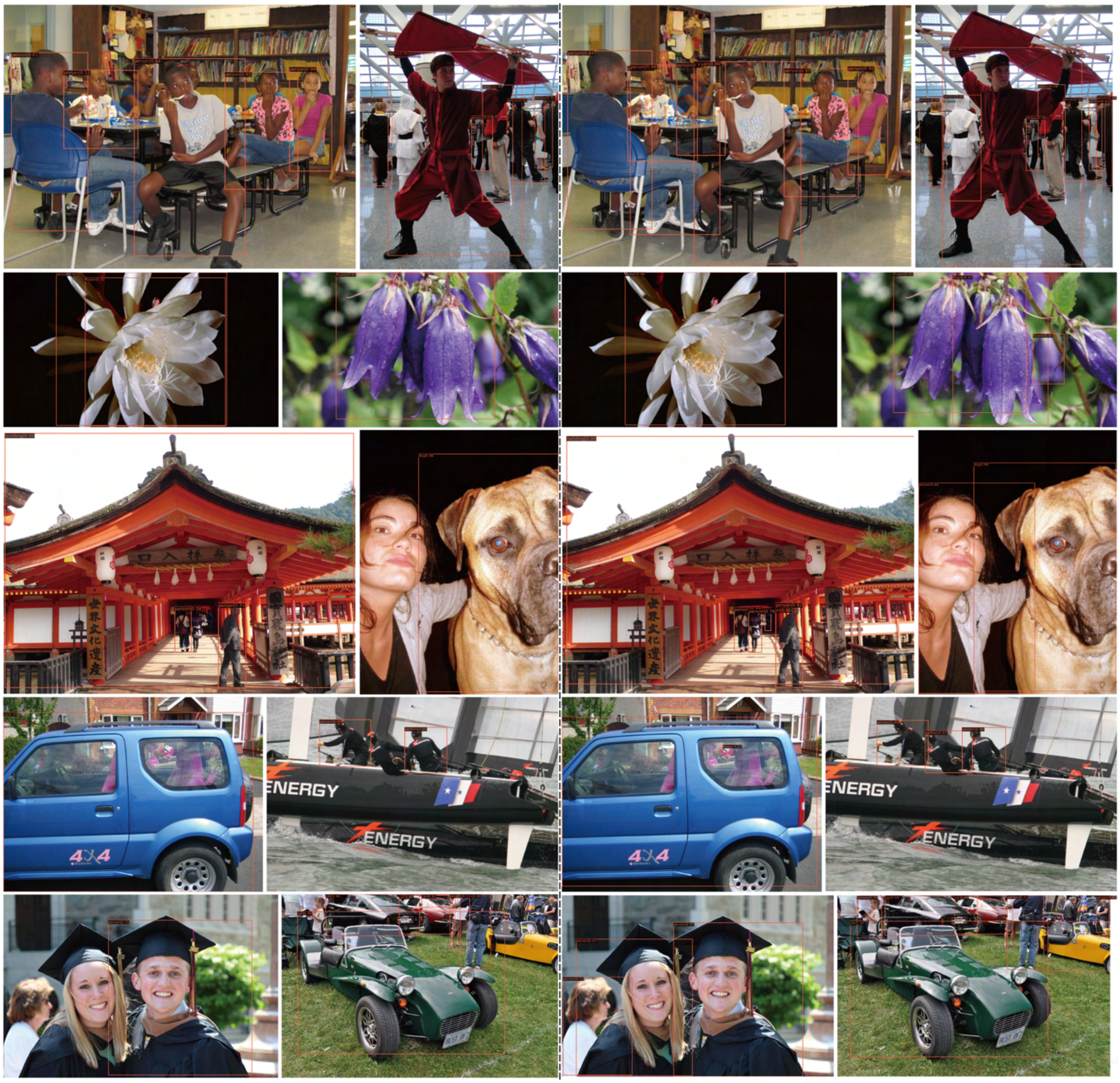


Figure 24. Results of detected objects in Open Images. The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where the models trained using our annotations achieve better detections.

Figure 25. Results of detected objects in Open Images (contd.). The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where the models trained using our annotations achieve better detections, especially for crowded objects.

Figure 26. Results of detected objects in Open Images (contd.). The left side shows predictions from the models trained using the original annotations. The right shows ones from the models trained using our annotations. This figure shows some results where (1) the models trained using our annotations fail to detect some objects and (2) the models trained using either set of annotations fail to detect some objects properly.