A Dataset for Developing and Benchmarking Active Vision

Phil Ammirato, Patrick Poirson, Eunbyung Park, Jana Kosecka, and Alexander C. Berg

Experiment Presentation

Presenters: Xingyi Zhou, Yajie Niu

Dataset Overview

- Dense images collection of indoor scenes

- Aligned high quality depth image.

- Bounding box and labels for object instances

- Images are connected by movement pointers

Dataset Tour

- See demo

Active Vision

- The paper used the REINFORCE algorithm for action prediction, with a reward of class scores.

- Alternative: The object score is highly related to object size, we can test simply moving forward to it, by first in-place rotating to centralize the object and then moving forward.

Active Vision - Experiment 1

- Idea: find the goal object and move towards it

- Motivation: test a simple approach on this dataset and see how it works

- Based on the intuition that when a person wants to pick up an object which is in sight, he usually catches the object with his eyes and then walk towards it.
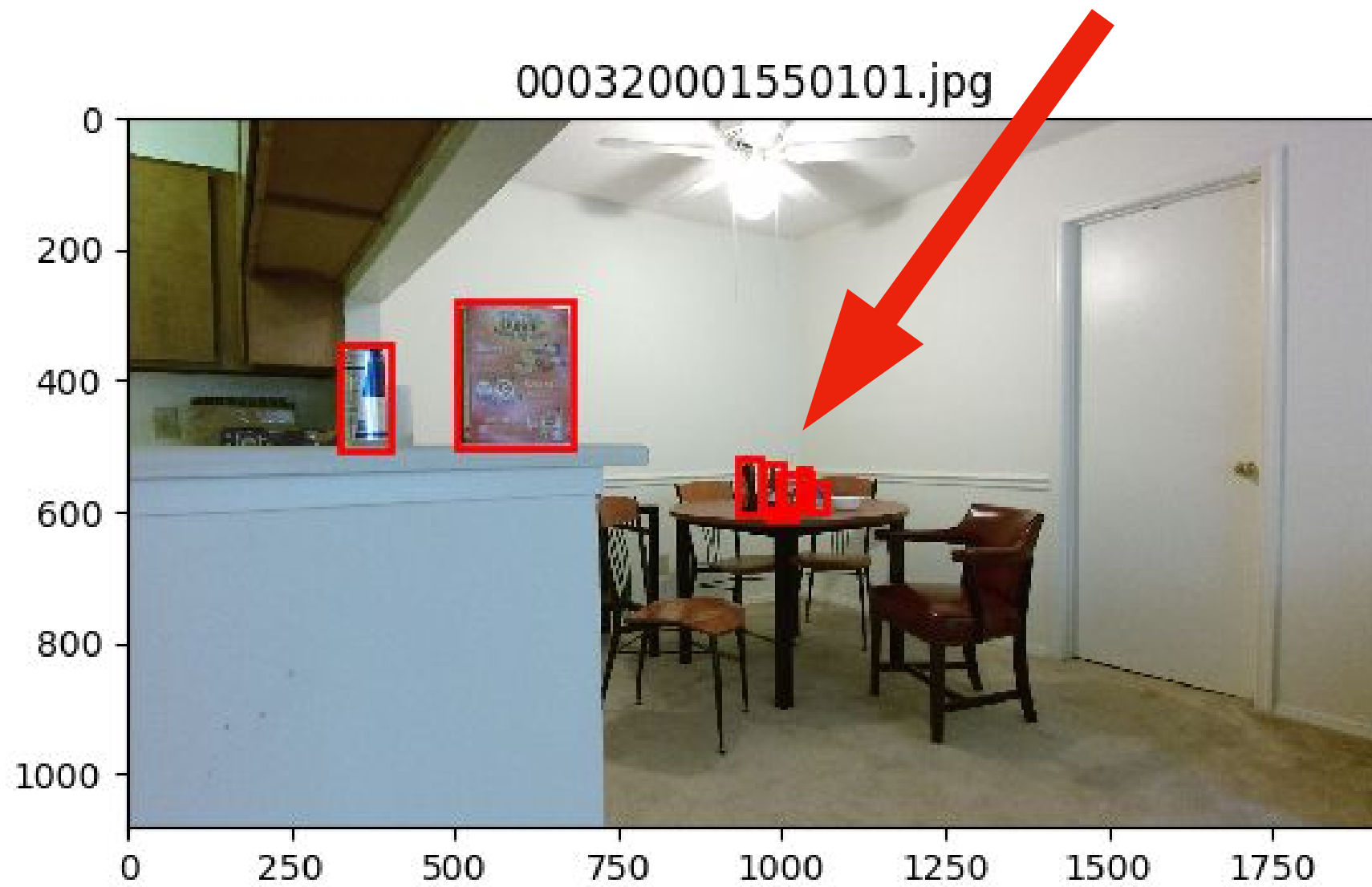
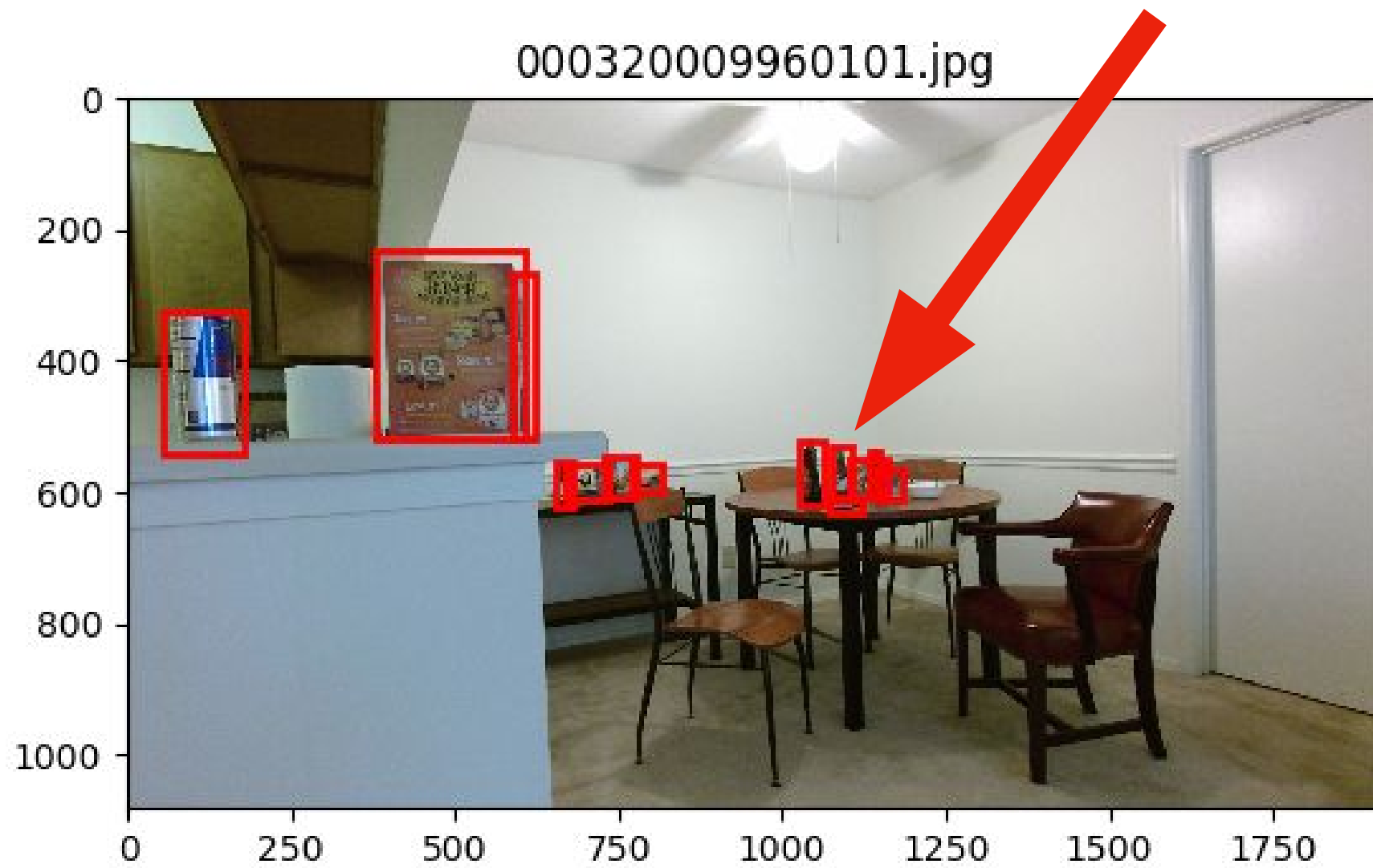**Step 0:**

object is here

000320001540101.jpg

Action to take: rotate to the left

Step 1:

000320001550101.jpg

Action to take: move forward

Step 2:
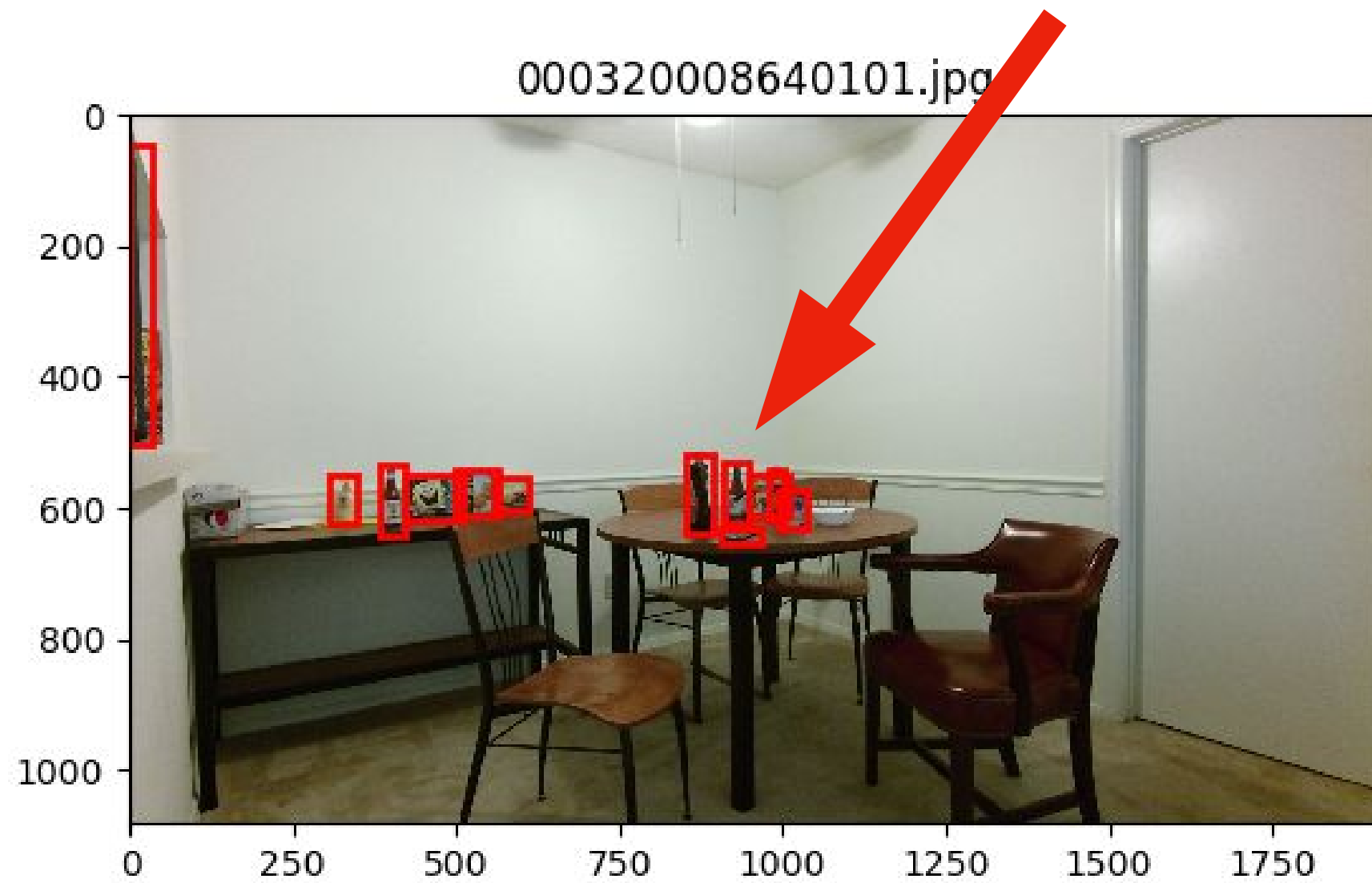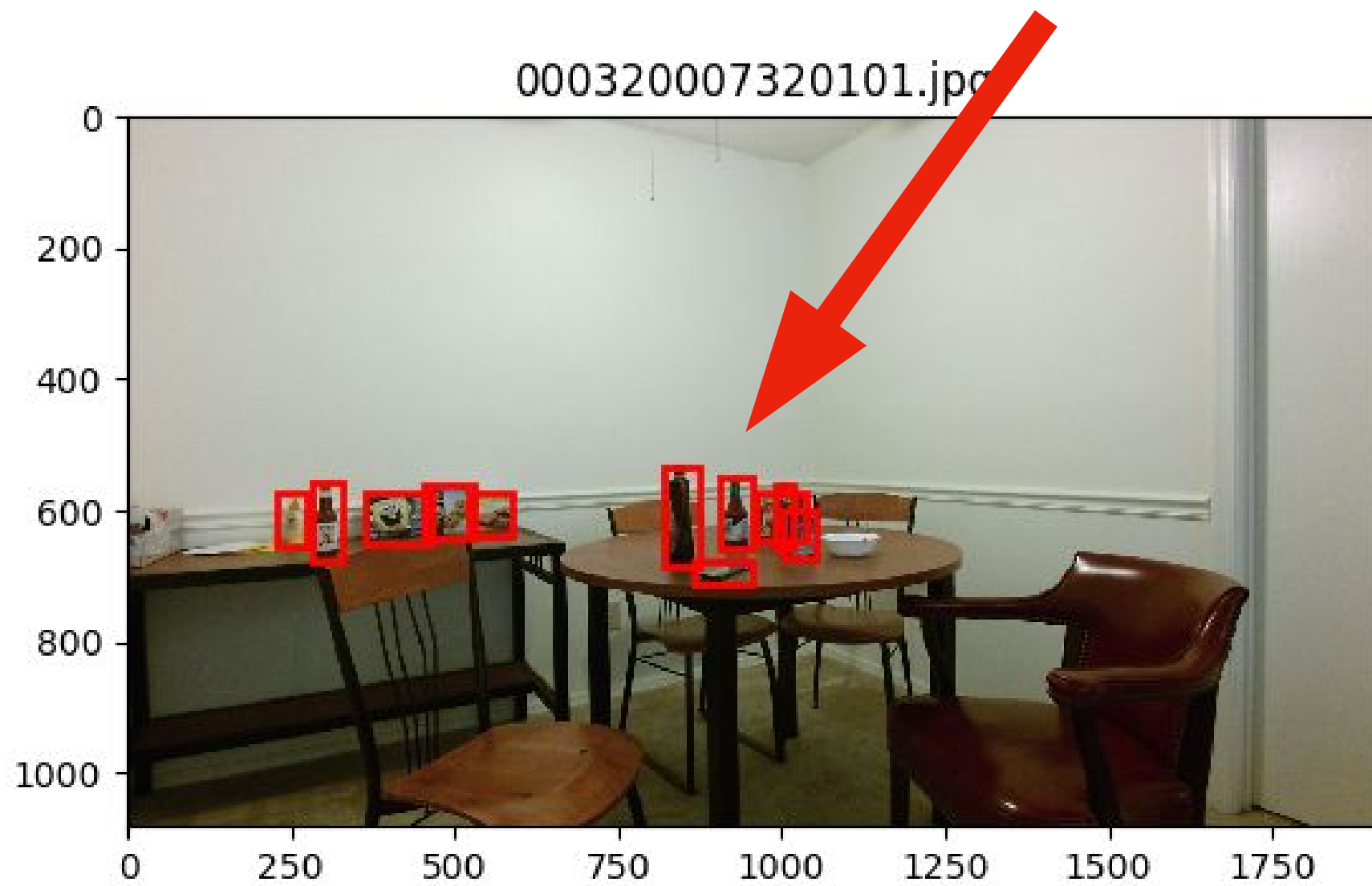
000320009960101.jpg

Action to take: move forward

Step 3:

000320008640101.jpg

Action to take: move forward

Step 4:

000320007320101.jpg

Action to take: move forward

Step 5:

000320006720101.jpg

Action to take: move forward

Step 6:

000320006000101.jpg
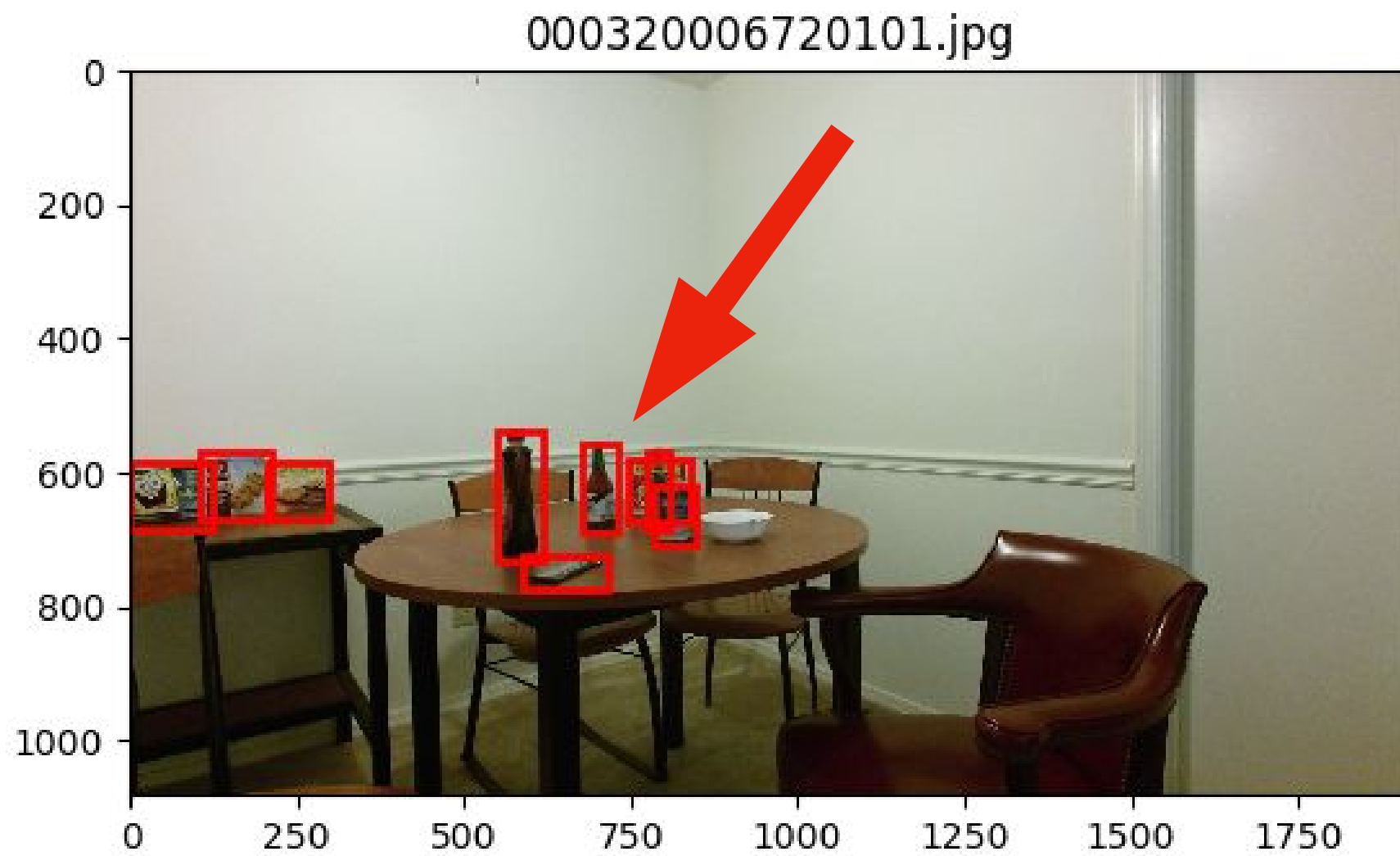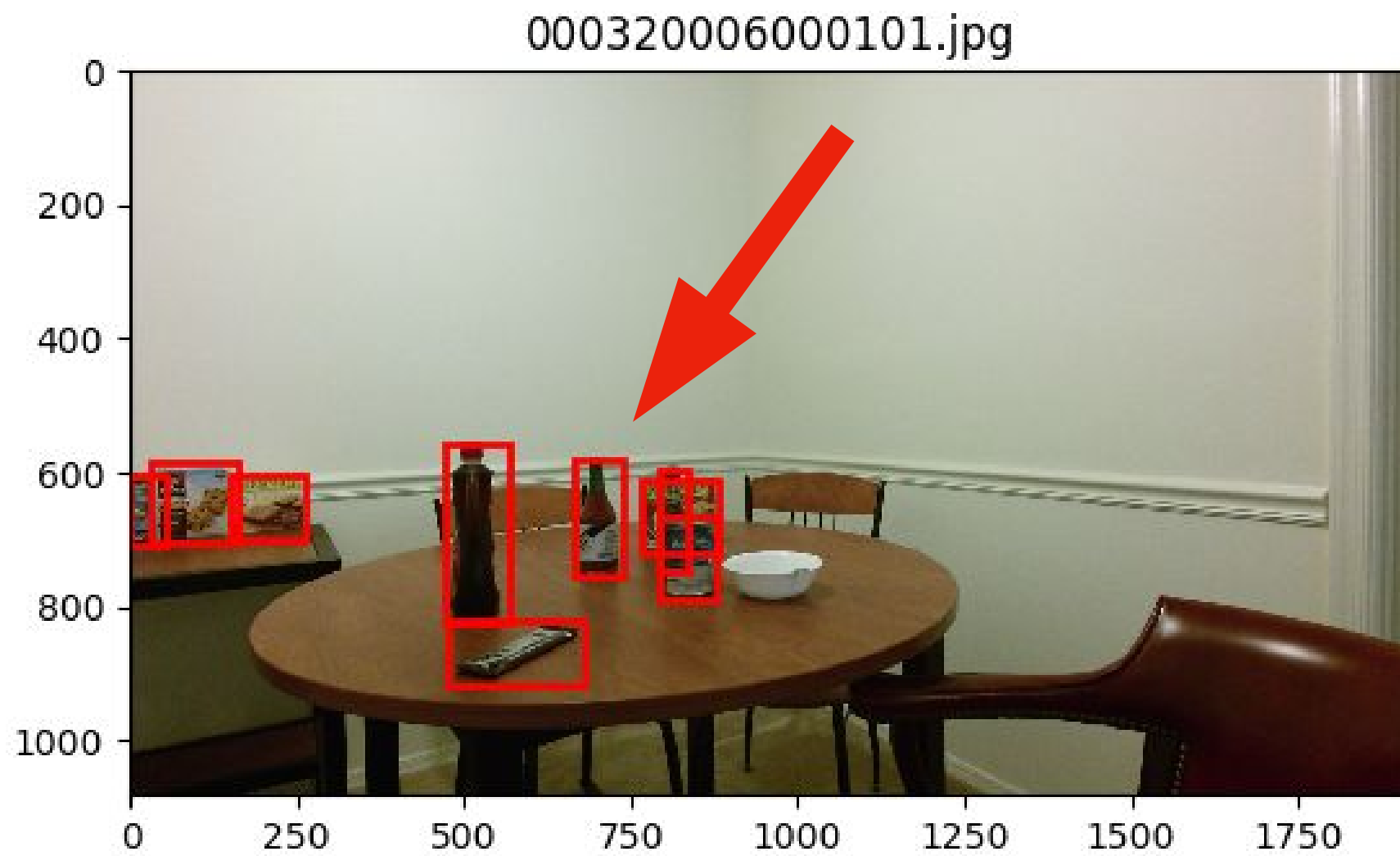
Action to take: move forward

Step 7:

000320006120101.jpg
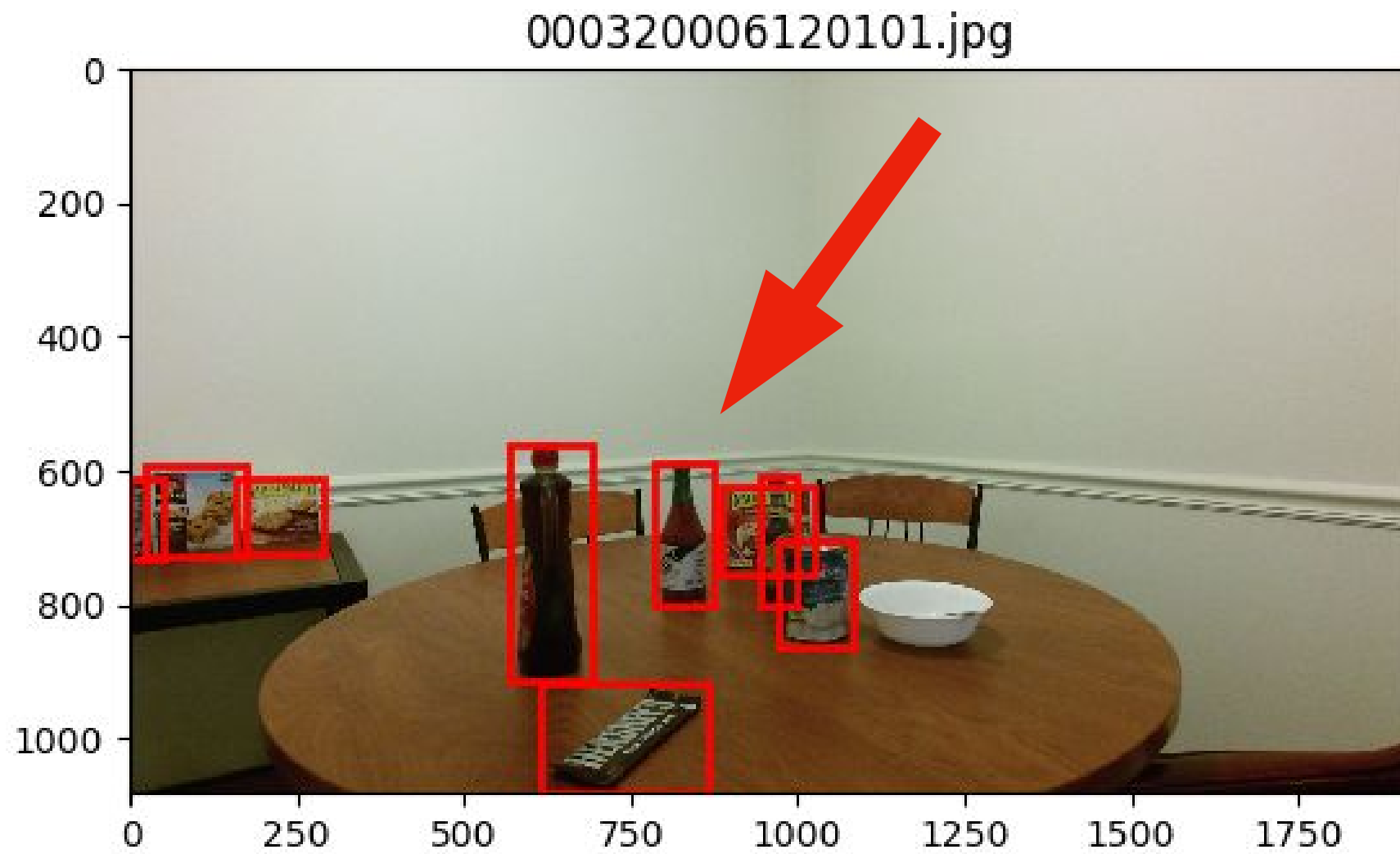
Action to take: move forward

Can't move forward anymore.

Problem: can't go around the obstacle on the way

000320006120101.jpg

new object

obstacle

where
we
are
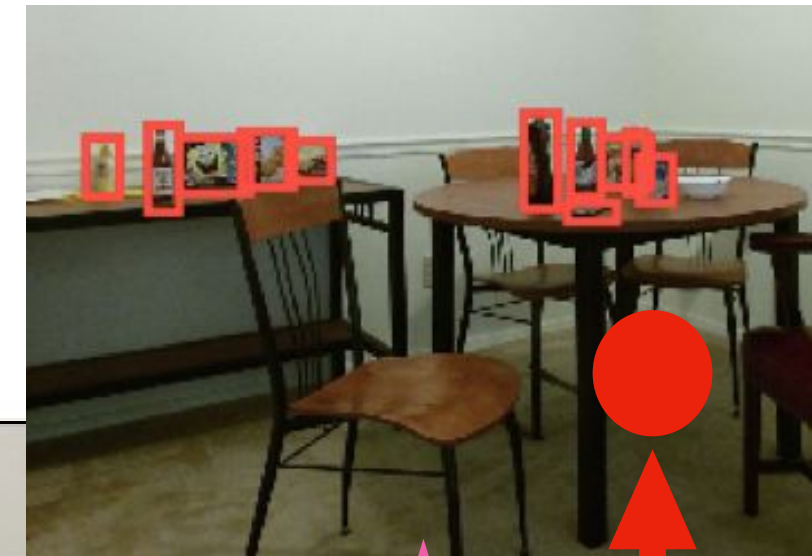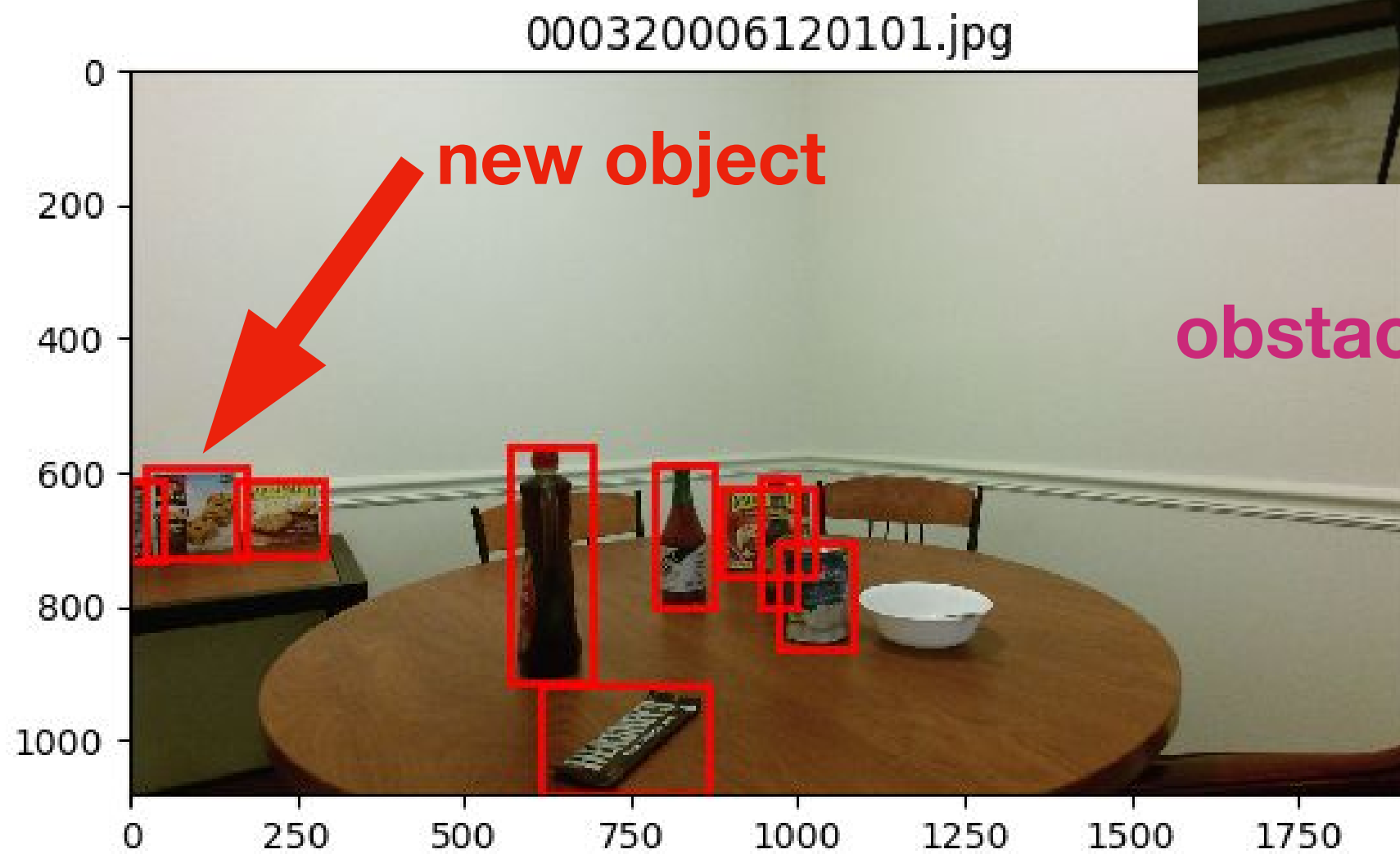
Action to take: rotate to the left

Problem: unexpected position change when rotating

Step 0:
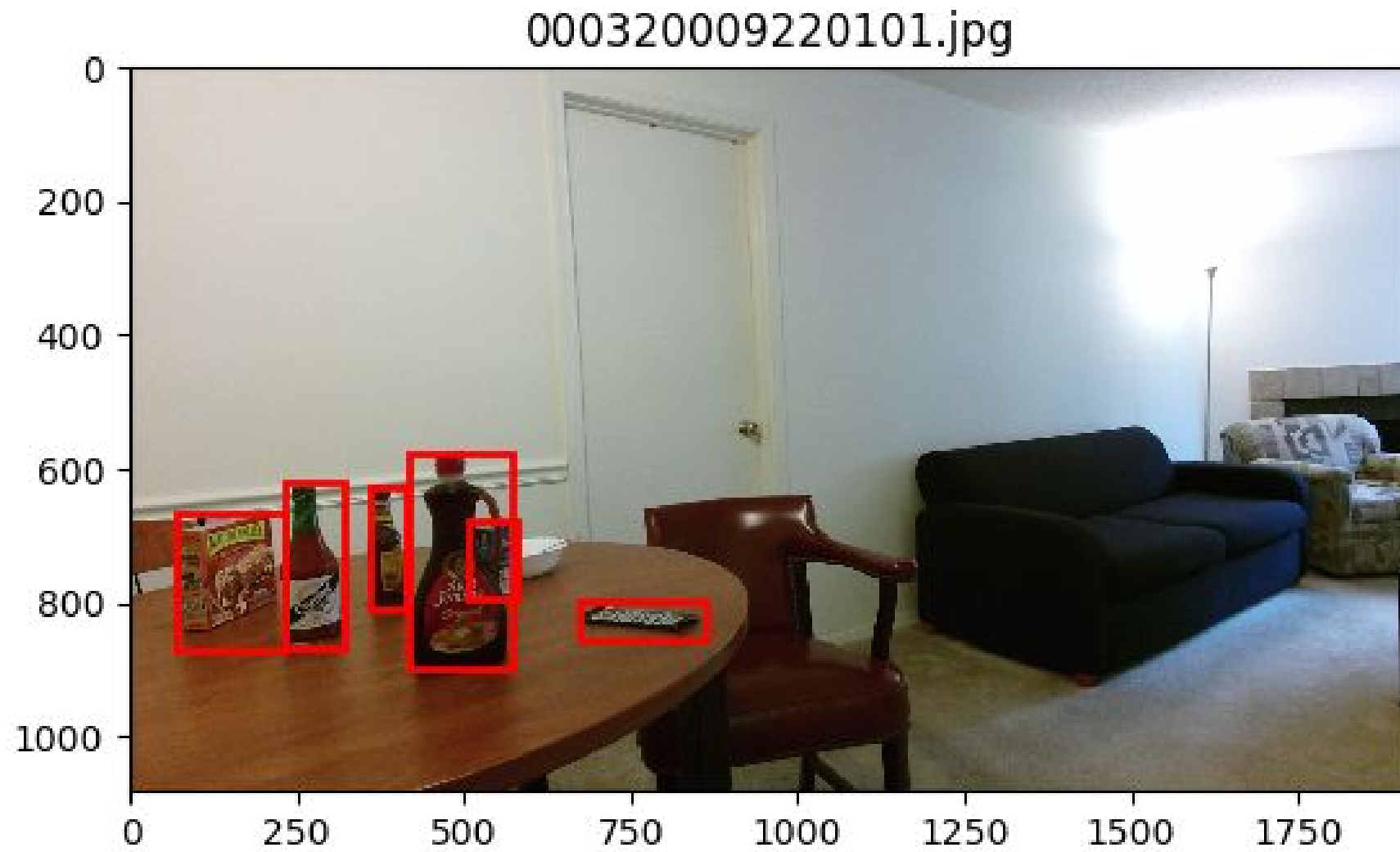
Action to take: rotate to the left

Problem: unexpected position change when rotating

Step 1:

000320009230101.jpg
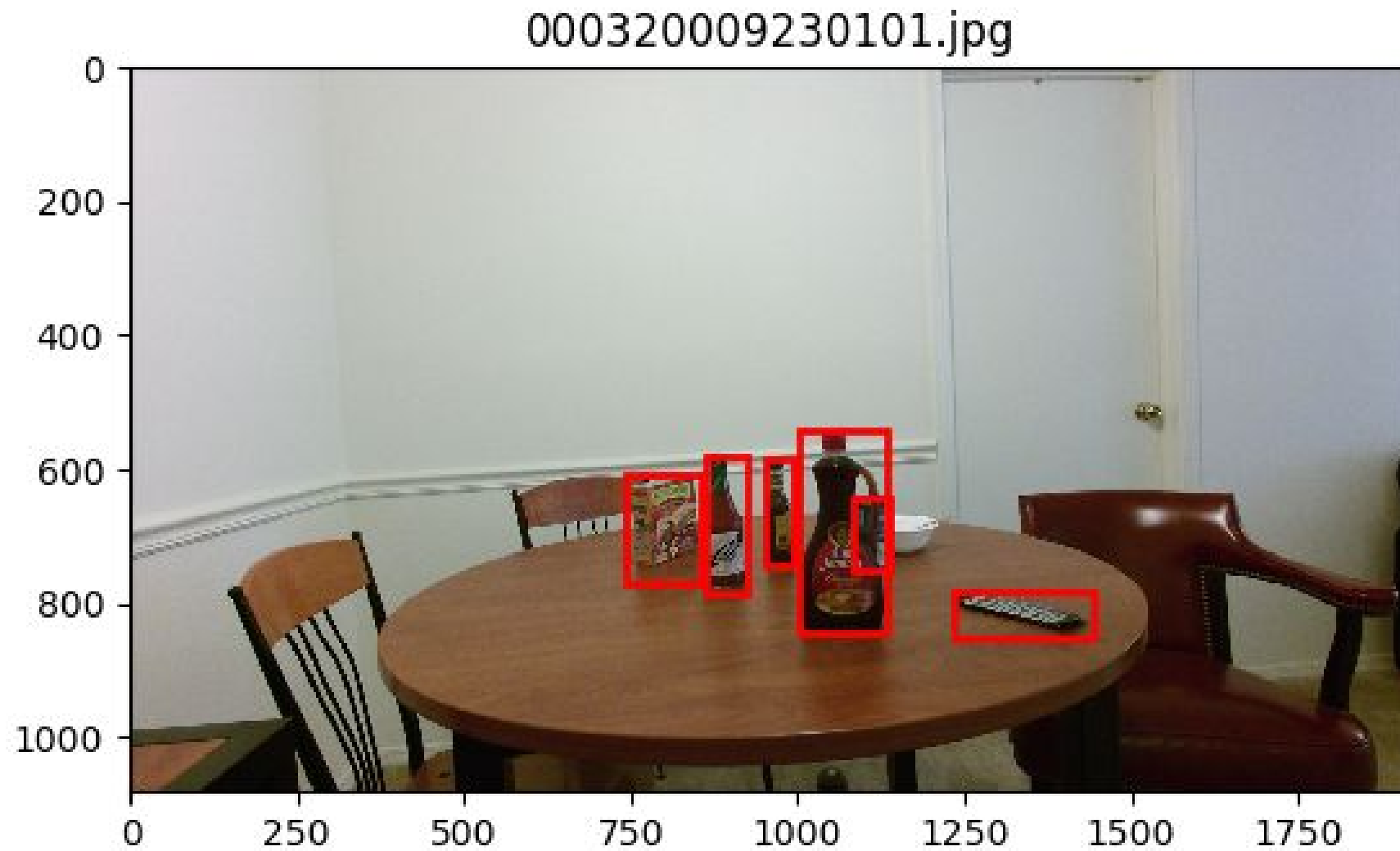
Action to take: rotate to the left

Problem: unexpected position change when rotating

Step 2:

Action to take: rotate to the left

Problem: unexpected position change when rotating

Step 3:

000320009130101.jpg
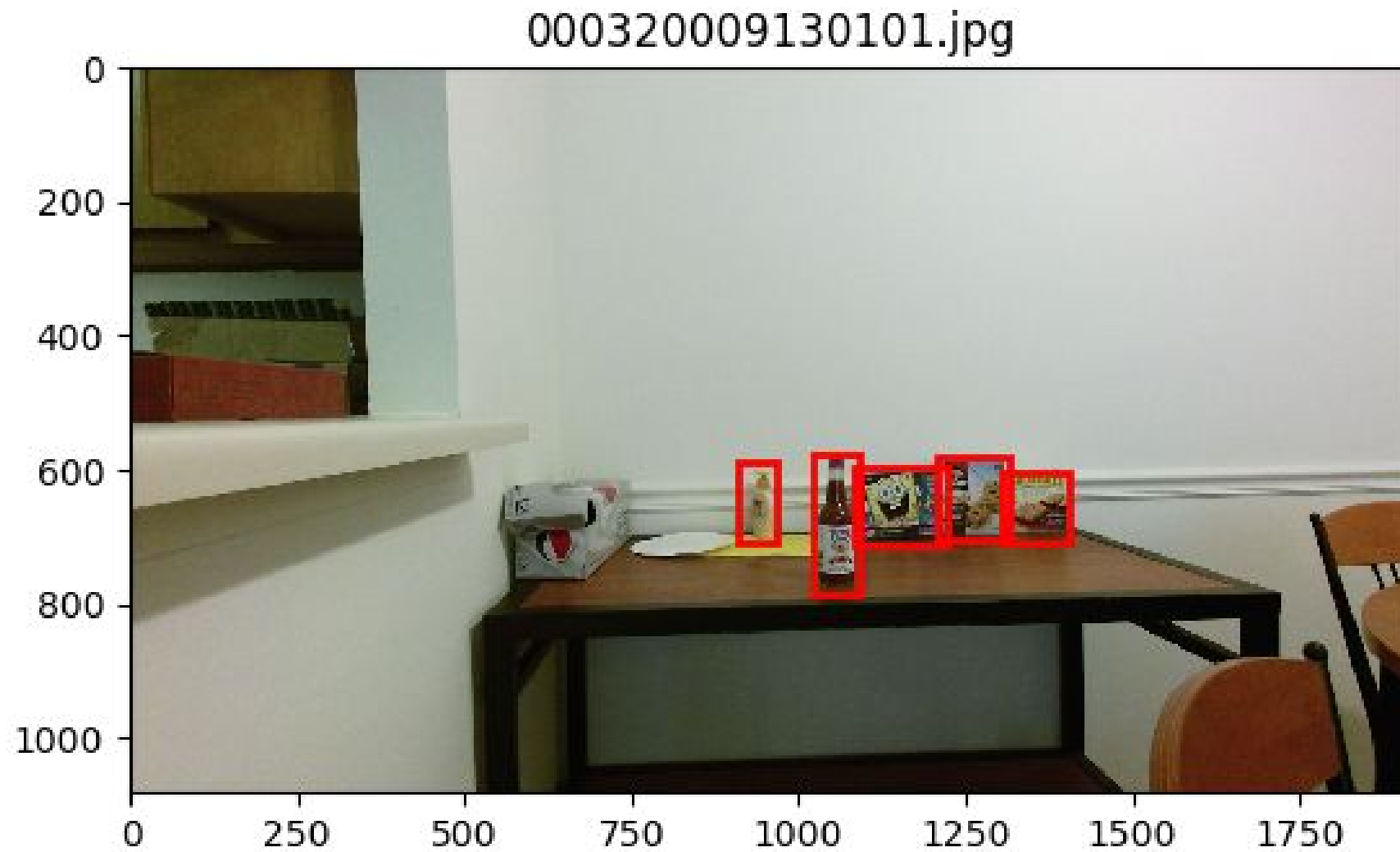
Action to take: rotate to the left

A sudden change of position!

Problem: unexpected position change when rotating

Step 4:

000320009140101.jpg
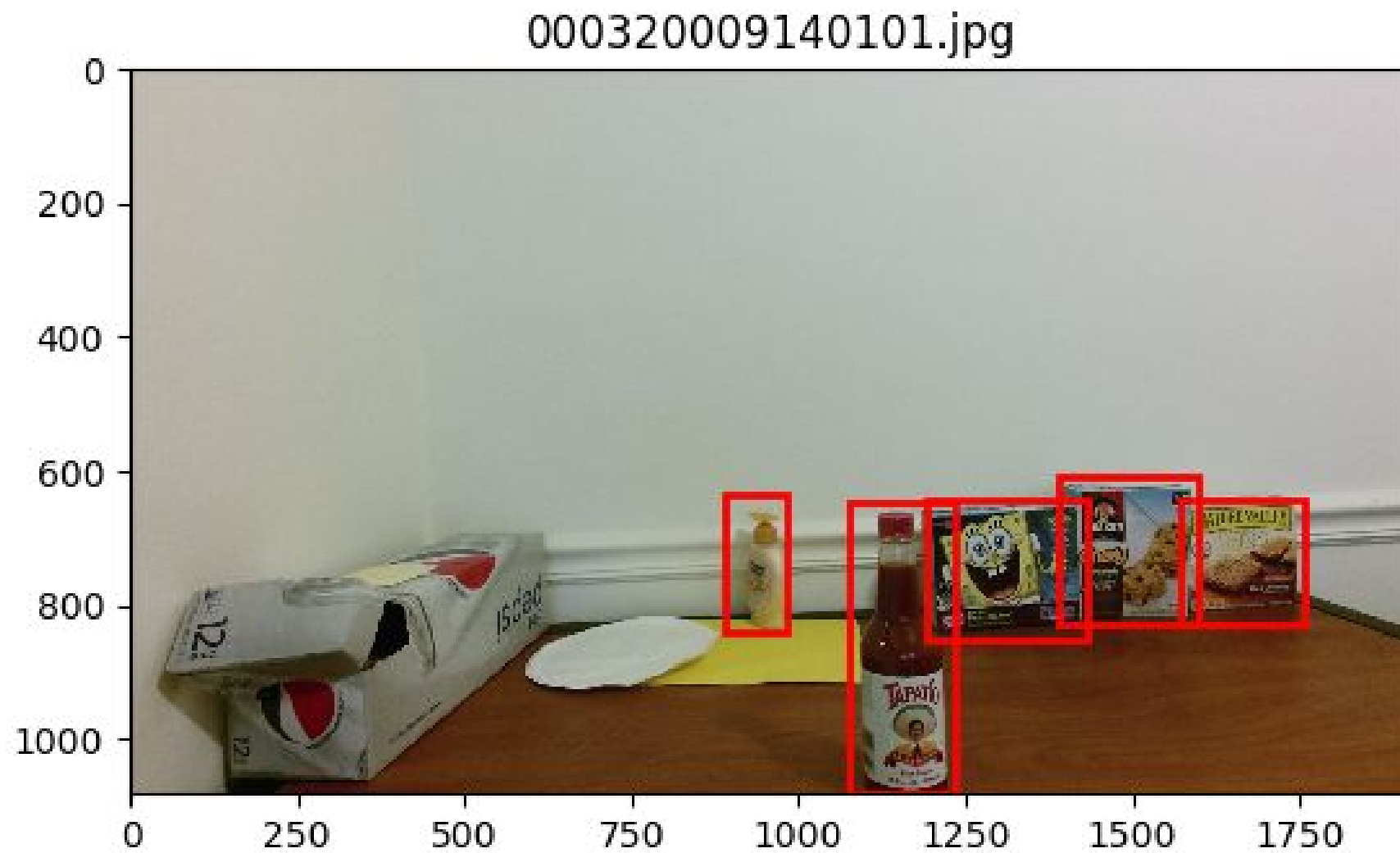
Action to take: rotate to the left

Problem: unexpected position change when rotating

Step 5:
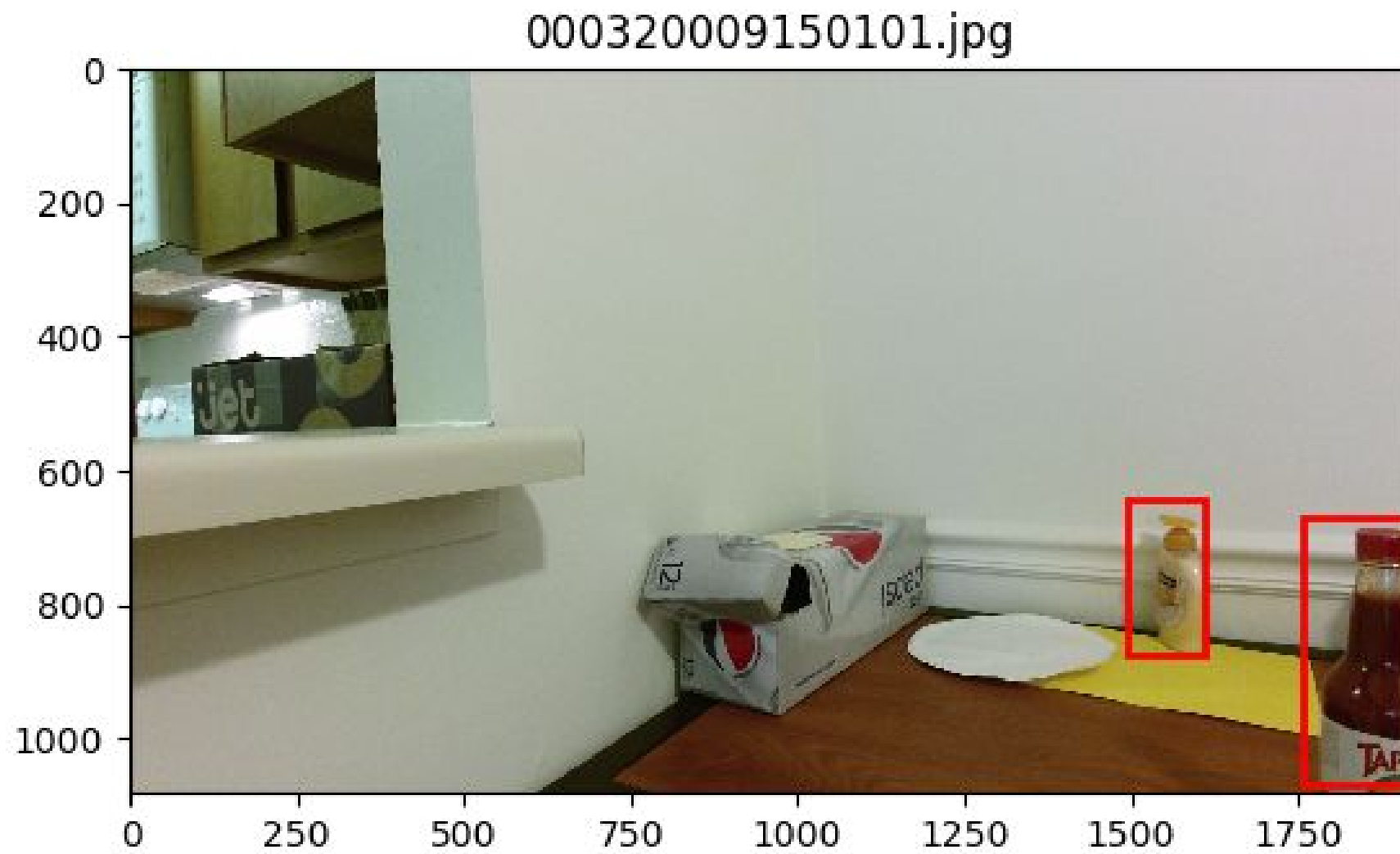
Alternative 1 - Results

- Results

Number of Moves	5	20
Method	Split 1	
REINFORCE	0.45	0.51
Alternative 1	0.330	0.394
Random	0.208	0.251

- Drawbacks

- Can't bypass the obstacle on the way

- Position change due to the dataset

- 'Fine-tuning' at the end to get a higher accuracy score

Active Vision - Supervised

- Alternative 2: Since we have all the object score information in training, we can apply supervised learning guided by the ground truth best movement.

Supervised action classification

Active Vision - Supervised

- Training data generation
  - Each frame is a tuple of (image, bbox, target_object_score)
  - Assign one of the six directions or a stop sign as classification target. Score is discarded in training.

(Image, box, score = 0.4)

Assigned action: rotate clockwise

(score = 0.35)   (score = 0.8)   (score = 0.9)

Supervised - Framework

Stacked RGB + Object Mask

ResNet 18

Prob of 6 + 1 actions
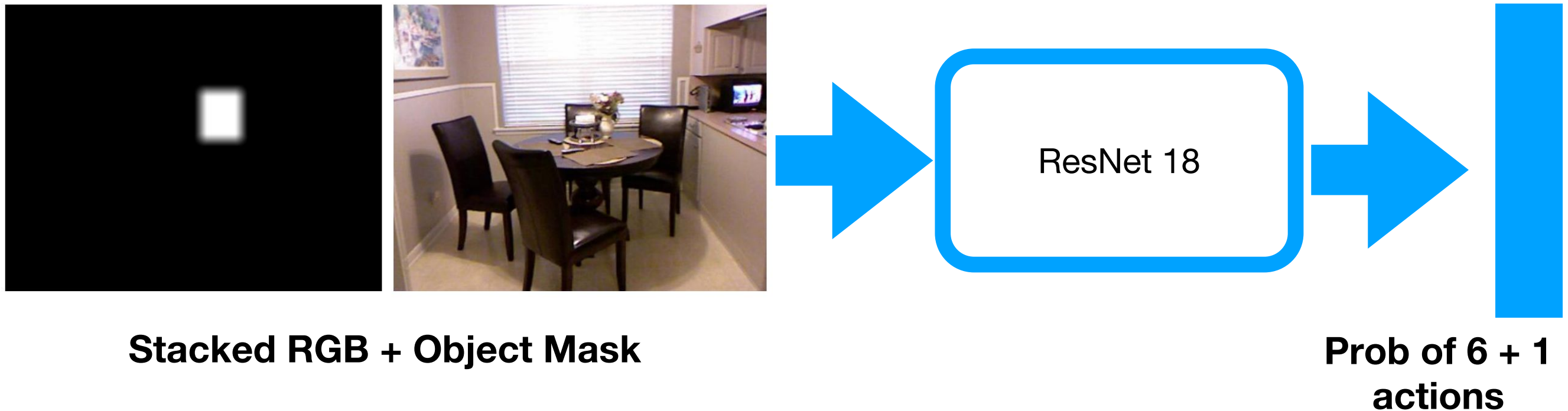
- input is a 4 channel RGB+Mask tensor
- The convolutional weight of the first 3 channel is copied from pretrained resnet
- initialize the conv weight of Mask channel with zero, so in the initial stage the resnet performs exactly the same as 3-channel version.

"What happens if..." Learning to Predict the Effect of Forces in Images. Mottaghi, R.,

Rastegari, M., Gupta, A., & Farhadi, A. ECCV16

Supervised - Results

Number of Moves	5	20
Method	Split 1	
REINFORCE	0.45	0.51
Greedy	0.330	0.394
Random	0.208	0.251
Supervised	0.252	0.304

Problem: The robot is easy to get stuck in a cycle or a deadend.

Code modified from the authors, by Xingyi Zhou
 https://github.com/xingyizhou/deep_active_vision

Supervised - Demos

```
Episodes 1041: object_id 29:
image_id: 816, correct: 0, score: 0.0168
action: 6, image_id: 816, correct: 0, score: 0.0168, bb(56,143,118,199)
action: 6, image_id: 805, correct: 0, score: 0.0429, bb(181,144,233,196)
action: 2, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
action: 6, image_id: 793, correct: 1, score: 0.6805, bb(212,134,251,170)
action: 5, image_id: 804, correct: 0, score: 0.2034, bb(92,127,132,165)
```

Supervised - Demos

001420000980101.jpg

Supervised - Demos

Supervised - Demos

001420000980101.jpg

Supervised - Demos

001420001640101.jpg

Supervised - Demos

```
Episodes 1054: object_id 20:
image_id: 830, correct: 0, score: 0.0068
action: 6, image_id: 830, correct: 0, score: 0.0068, bb(75,118,78,131)
action: 5, image_id: 829, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 2, image_id: 817, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 6, image_id: 818, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 773, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 701, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 653, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 5, image_id: 652, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 4, image_id: 712, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 6, image_id: 713, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 665, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 1, image_id: 701, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 653, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 5, image_id: 652, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 4, image_id: 712, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 6, image_id: 713, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 665, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 1, image_id: 701, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 3, image_id: 653, correct: 0, score: 0.0000, bb(1,1,2,2)
action: 5, image_id: 652, correct: 0, score: 0.0000, bb(1,1,2,2)
```

Supervised - Demos

001420001070101.jpg

Supervised - Demos

001420001060101.jpg

Supervised - Demos

Supervised - Demos

001420001060101.jpg

Supervised - Demos

Conclusion

- Dataset tour

- Experiment 1: moving towards the goal object through a straight line

- Experiment 2: supervised learning given the ground truth best action.

- Active vision is a challenging task and this dataset serves as a useful benchmark for this task.

Thank you!