Carnegie Mellon University

Cozmo Partial Cube Detector

By: Tyler Johnson

Cognitive Robotics 15-694

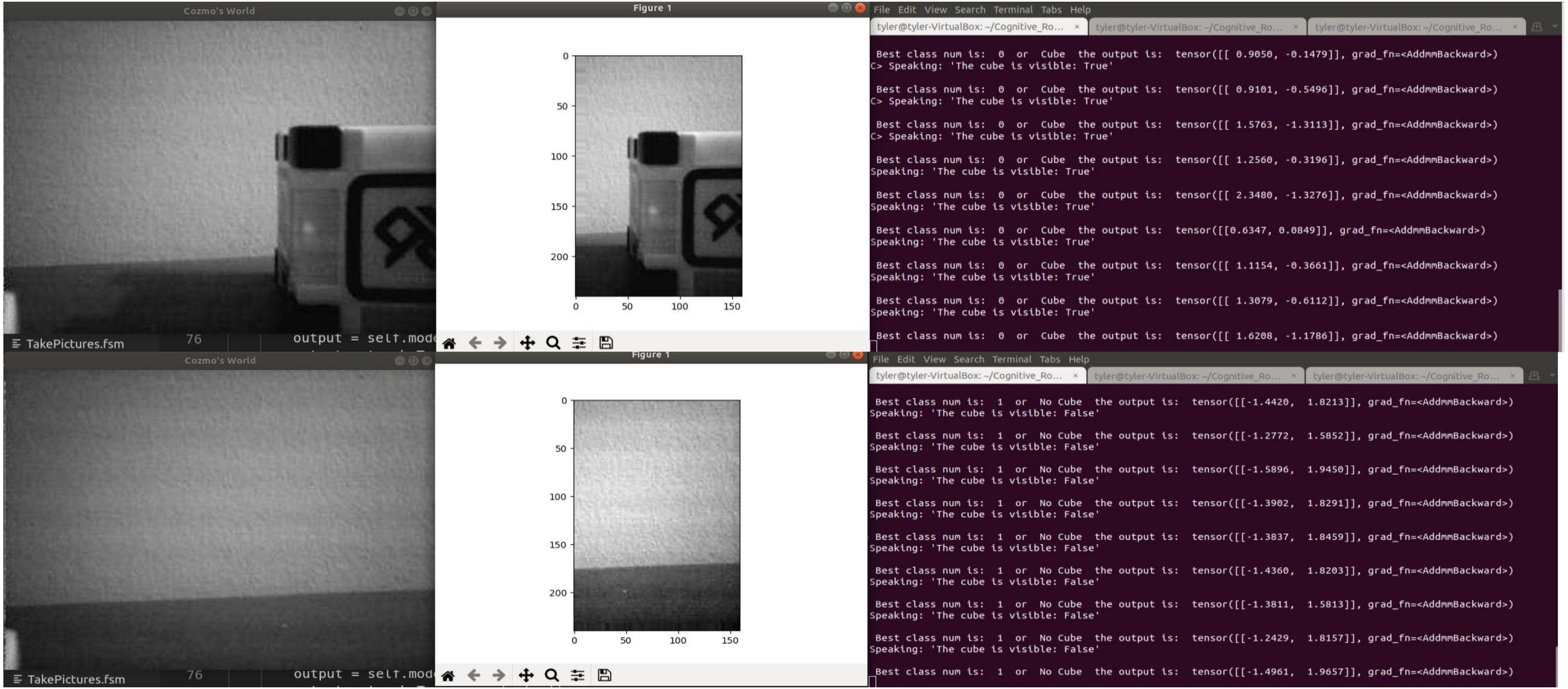
Teaching cozmo to see cubes:

- Utilizing a convolutional neural network train cozmo to recognize cube images from no cube images.
- Training set consisted of many unique right-side images with varying viewpoint, composition, and lighting.
- Extended number of training samples with random cropping to increase stability of model.
- Load best trained model to classify cozmo’s current captured view and make judgements based on presence of cubes.

Current CNN Architecture:

- Two convolution layers, each having 32 kernels of 5*5 dimensions
- With stride: 1, max pooling: 4, and padding: 2
- Batch normalization used after convolution layers to keep instances consistent with current batch.

Results of initial Training:


Cozmo's World
Figure 1
0
50
100
150
200
0
50
100
150
TakePictures.fsm
76
output = self.mod
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Ro...
tyler@tyler-VirtualBox: ~/Cognitive_Ro...
tyler@tyler-VirtualBox: ~/Cognitive_Ro...

```
 Best class num is:  0  or  Cube  the output is:  tensor([[ 0.9050, -0.1479]], grad_fn=<AddmmBackward>)
C> Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 0.9101, -0.5496]], grad_fn=<AddmmBackward>)
C> Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 1.5763, -1.3113]], grad_fn=<AddmmBackward>)
C> Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 1.2560, -0.3196]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 2.3480, -1.3276]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[0.6347, 0.0849]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 1.1154, -0.3661]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 1.3079, -0.6112]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube  the output is:  tensor([[ 1.6208, -1.1786]], grad_fn=<AddmmBackward>)
```

```
 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.4420,  1.8213]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.2772,  1.5852]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.5896,  1.9450]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.3902,  1.8291]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.3837,  1.8459]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.4360,  1.8203]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.3811,  1.5813]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.2429,  1.8157]], grad_fn=<AddmmBackward>)
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube  the output is:  tensor([[-1.4961,  1.9657]], grad_fn=<AddmmBackward>)
```

The cube detector works well for simple cases, and instances that look like training set. Has much larger output values for correct class compared to the other.

Results of Right/Left cube detector:

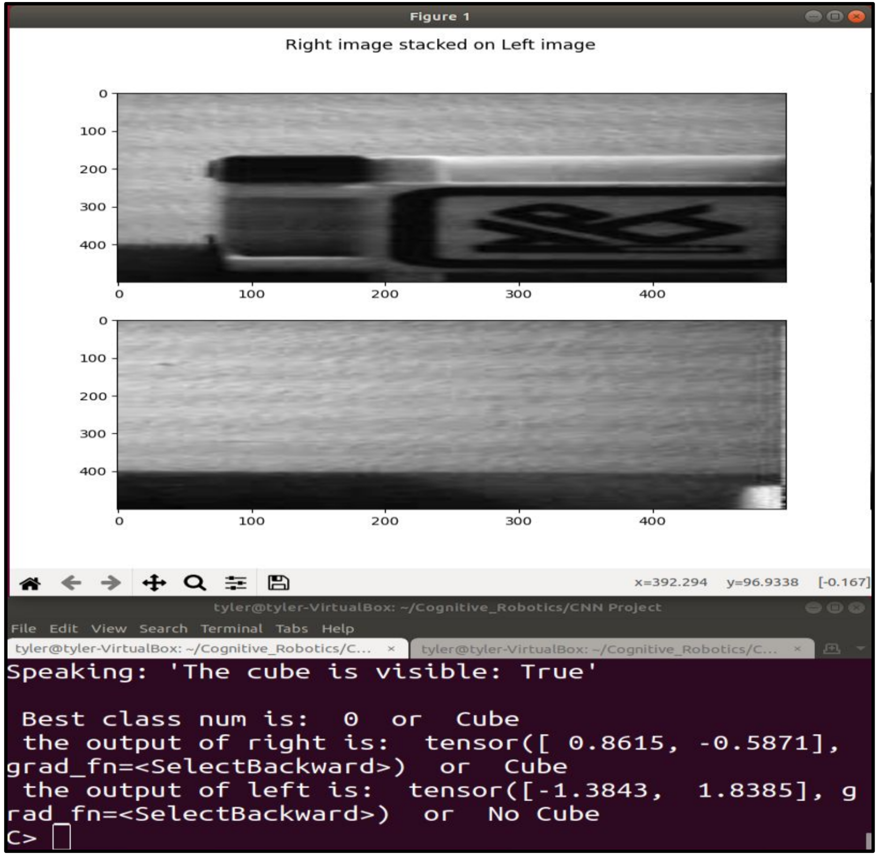


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
x=392.294 y=96.9338 [-0.167]
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/C...
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/C...

```
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube
 the output of right is:  tensor([ 0.8615, -0.5871],
grad_fn=<SelectBackward>)  or  Cube
 the output of left is:  tensor([-1.3843,  1.8385], g
rad_fn=<SelectBackward>)  or  No Cube
C>
```

Correctly Detected cube on right half of image and not on the left

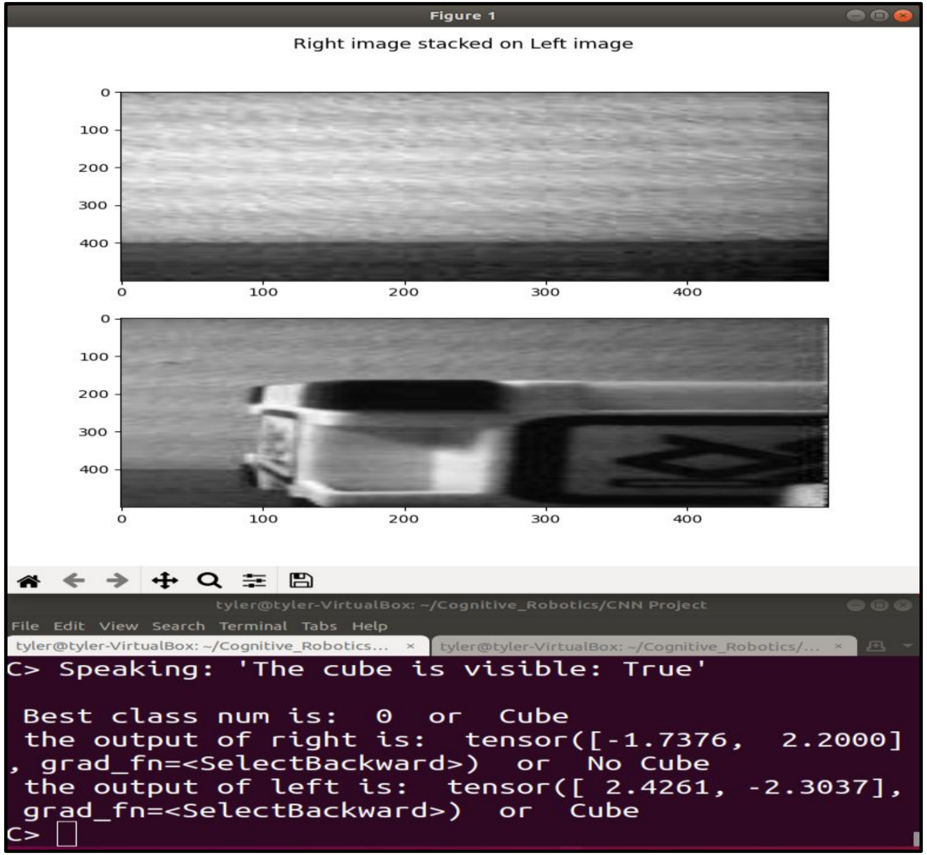


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics...
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/...

```
C> Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube
 the output of right is:  tensor([-1.7376,  2.2000]
, grad_fn=<SelectBackward>)  or  No Cube
 the output of left is:  tensor([ 2.4261, -2.3037],
 grad_fn=<SelectBackward>)  or  Cube
C>
```

Correctly detected cube on left half of image and not on the right

Results of Right/Left cube detector:

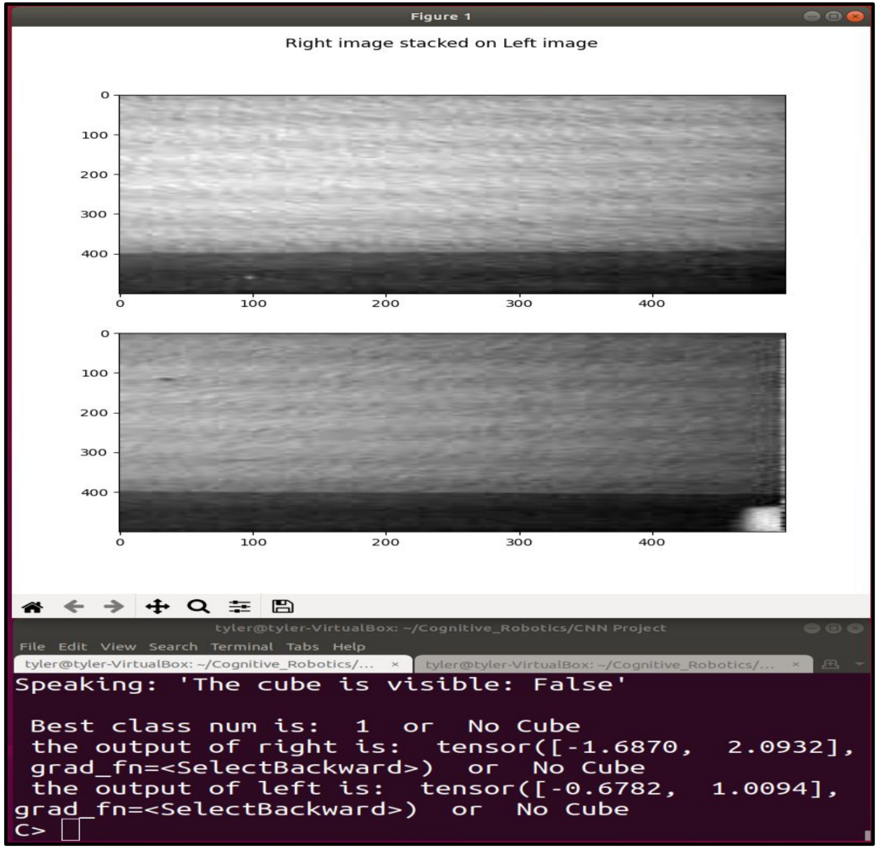


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/...
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/...

```
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube
 the output of right is:  tensor([-1.6870,  2.0932],
 grad_fn=<SelectBackward>)  or  No Cube
 the output of left is:  tensor([-0.6782,  1.0094],
grad_fn=<SelectBackward>)  or  No Cube
C>
```

Correctly did not detect a cube on either side

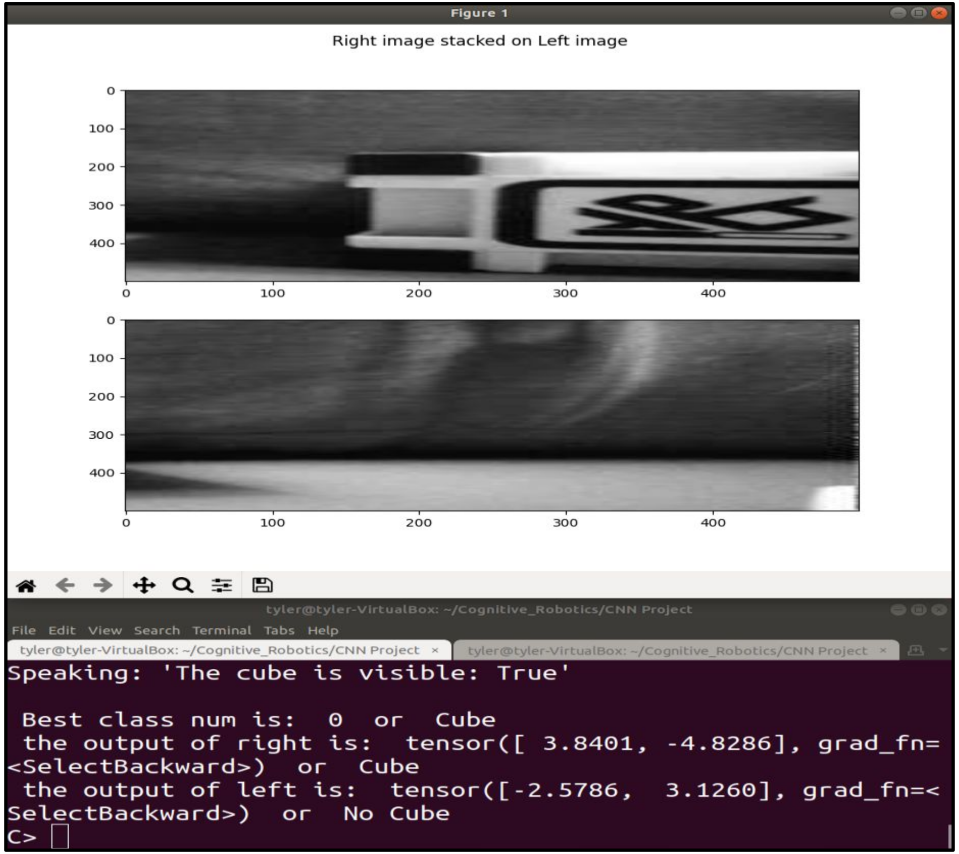


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project

```
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube
 the output of right is:  tensor([ 3.8401, -4.8286], grad_fn=
<SelectBackward>)  or  Cube
 the output of left is:  tensor([-2.5786,  3.1260], grad_fn=<
SelectBackward>)  or  No Cube
C>
```

Correctly classified cube against different background

Common Misclassifications and Problems:

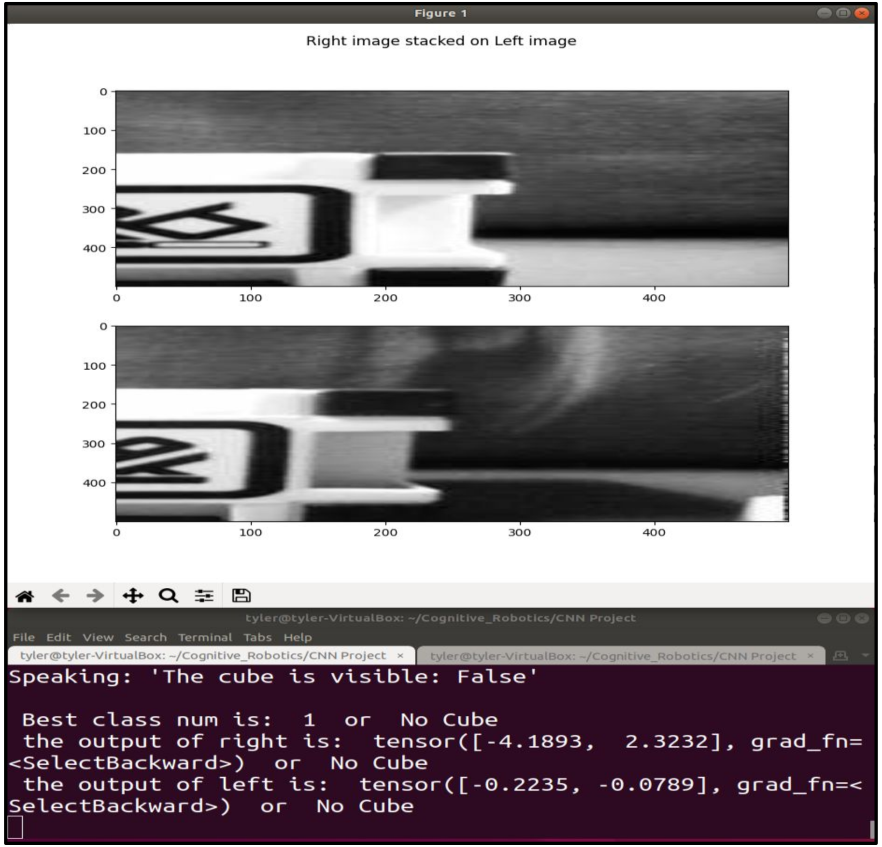


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project ×
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project ×

```
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube
 the output of right is:  tensor([-4.1893,  2.3232], grad_fn=
<SelectBackward>)  or  No Cube
 the output of left is:  tensor([-0.2235, -0.0789], grad_fn=<
SelectBackward>)  or  No Cube
```

If cube is too close to center and not near right side the images do not look like right-cube training images.

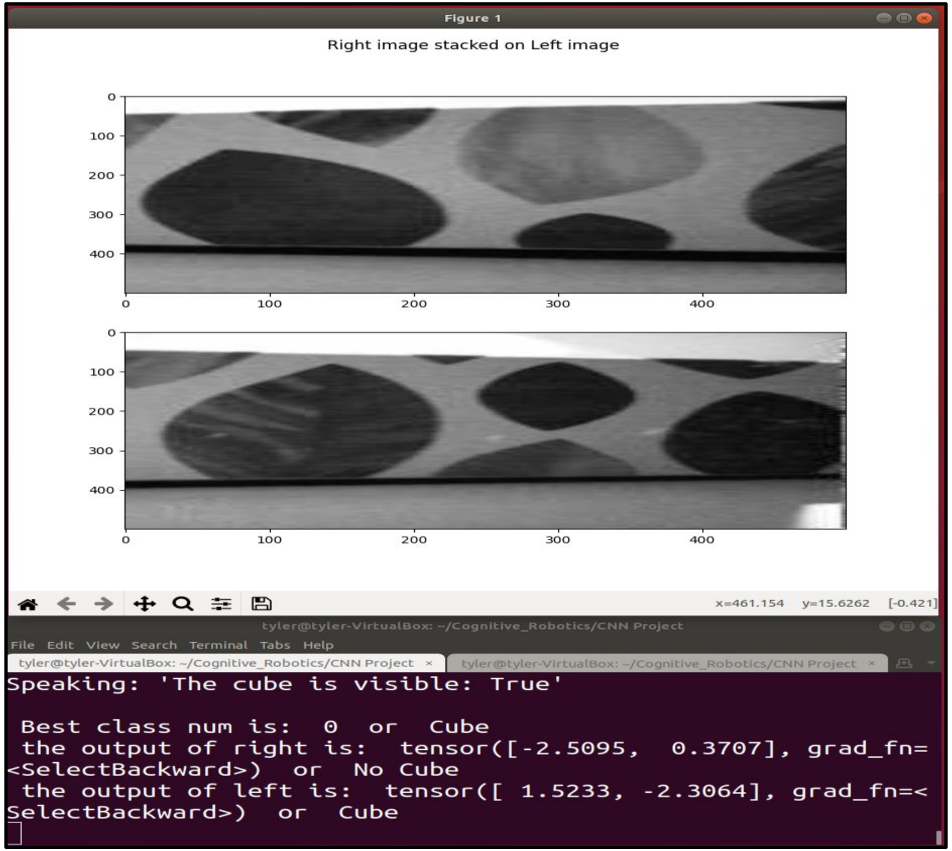


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
0
100
200
300
400
x=461.154 y=15.6262 [-0.421]
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project ×
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project ×

```
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube
 the output of right is:  tensor([-2.5095,  0.3707], grad_fn=
<SelectBackward>)  or  No Cube
 the output of left is:  tensor([ 1.5233, -2.3064], grad_fn=<
SelectBackward>)  or  Cube
```

Edge of objects can look like cube and give false-positives

Common Misclassifications and Problems:

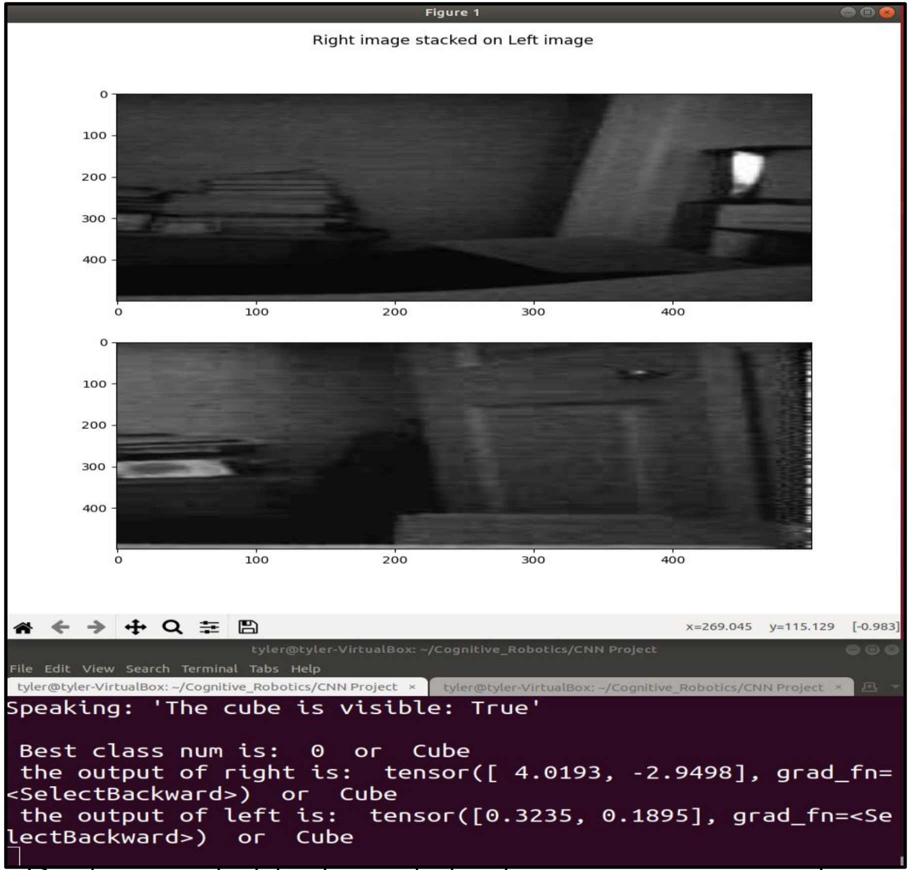


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
x=269.045 y=115.129 [-0.983]
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project

```
Speaking: 'The cube is visible: True'

 Best class num is:  0  or  Cube
 the output of right is:  tensor([ 4.0193, -2.9498], grad_fn=
<SelectBackward>)  or  Cube
 the output of left is:  tensor([0.3235, 0.1895], grad_fn=<Se
lectBackward>)  or  Cube
```

Background objects and shadows can appear cube like to the model.

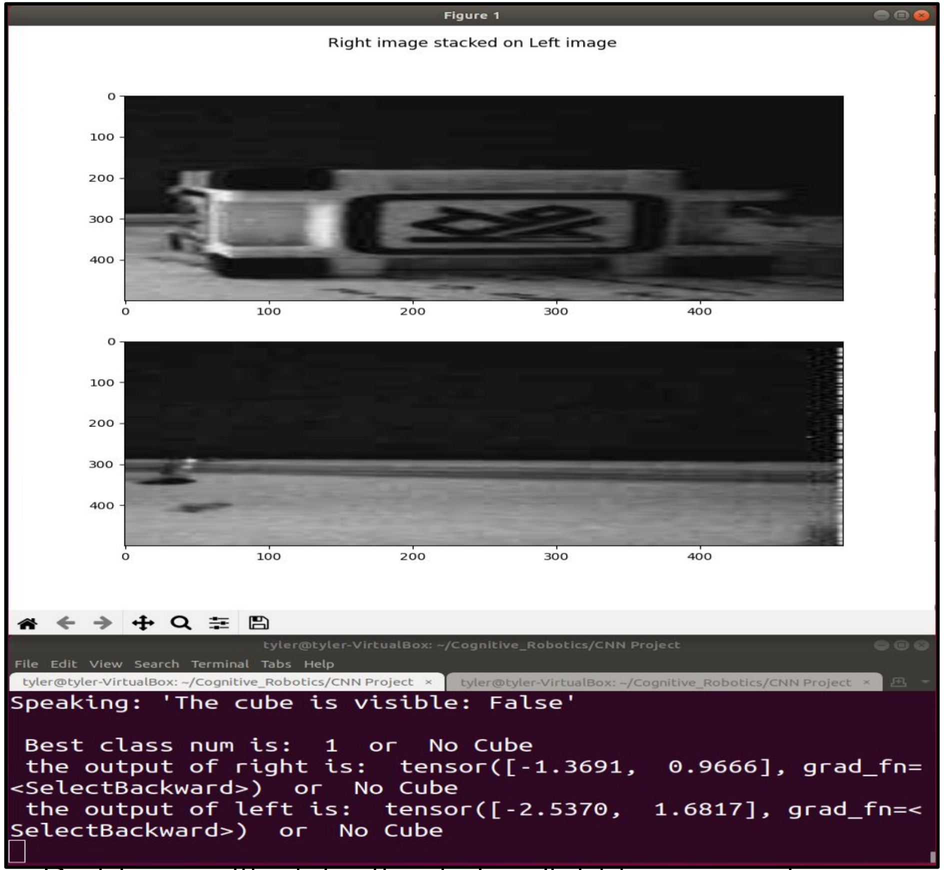


Figure 1
Right image stacked on Left image
0
100
200
300
400
0
100
200
300
400
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
File Edit View Search Terminal Tabs Help
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project
tyler@tyler-VirtualBox: ~/Cognitive_Robotics/CNN Project

```
Speaking: 'The cube is visible: False'

 Best class num is:  1  or  No Cube
 the output of right is:  tensor([-1.3691,  0.9666], grad_fn=
<SelectBackward>)  or  No Cube
 the output of left is:  tensor([-2.5370,  1.6817], grad_fn=<
SelectBackward>)  or  No Cube
```

Problems with detection in low light images or images much brighter than training set.

Future improvements/ Experiments:

- Test different image processing techniques to handle large differences in brightness, from training set, seen in cozmo’s viewpoint: per-instance normalization, preprocessing…
- Test preprocessing (edges / gaussian blurring) methods on training set to see if trains a more accurate model with respect to current model problems.
- Get more training data that handles hard cases such as cube with many different objects, or no cube with many different objects, and background objects.
- Experiment with different neural network architecture designs with varying parameters to find optimal model.