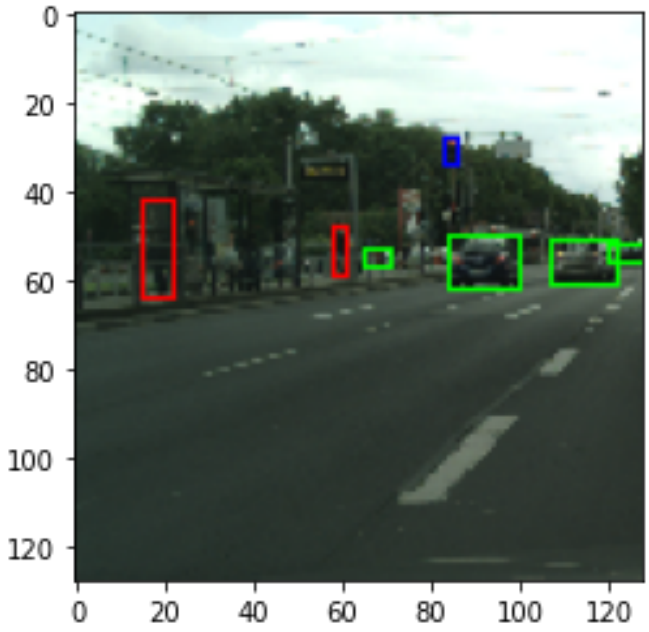
YOLO Results

By : Ruchi Gupte, Ishani Mhatre

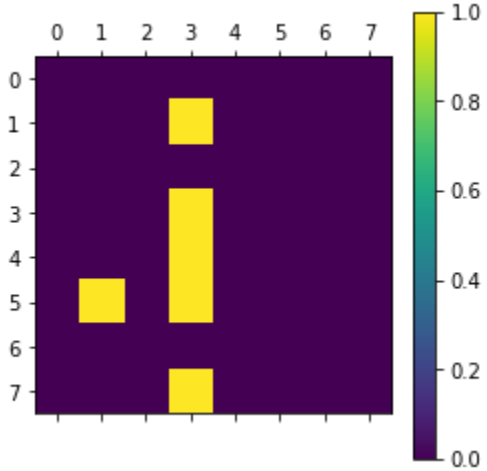
Image with ground truth bounding boxes visualized taking i=1000 image



Each channel of the processed labels:

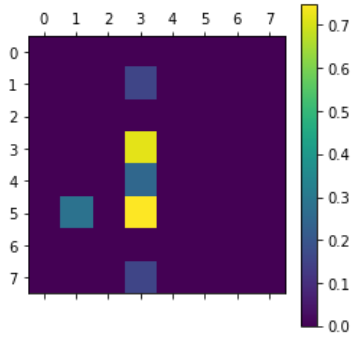
Channel 1: Pr(Objectness)

```
[[0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]]
```



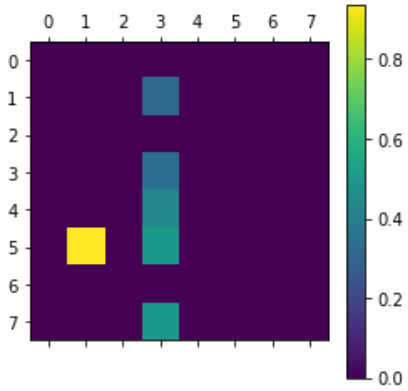
Channel 2: Bounding Box (x)

```
[[0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.15625 0.      0.      0.      0.     ]
 [0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.71875 0.      0.      0.      0.     ]
 [0.      0.      0.      0.25    0.      0.      0.      0.     ]
 [0.      0.28125 0.      0.75    0.      0.      0.      0.     ]
 [0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.15625 0.      0.      0.      0.     ]]
```



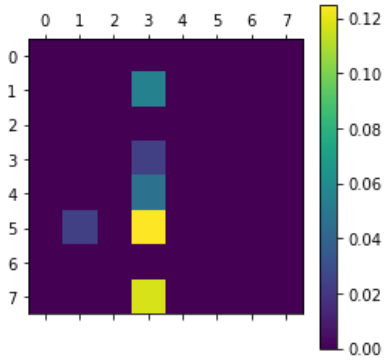
Channel 3: Bounding Box (y)

```
[[0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.3125  0.      0.      0.      0.     ]
 [0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.34375 0.      0.      0.      0.     ]
 [0.      0.      0.      0.4375  0.      0.      0.      0.     ]
 [0.      0.9375  0.      0.5     0.      0.      0.      0.     ]
 [0.      0.      0.      0.      0.      0.      0.      0.     ]
 [0.      0.      0.      0.5     0.      0.      0.      0.     ]]
```



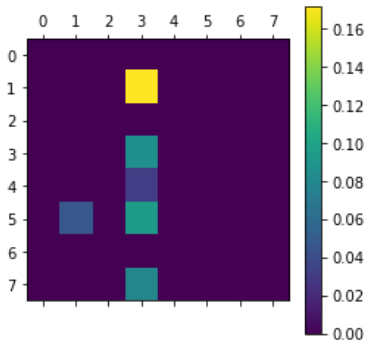
Channel 4: Bounding Box (w)

```
[[0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.0546875 0.        0.        0.        0.       ]
 [0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.0234375 0.        0.        0.        0.       ]
 [0.        0.        0.        0.046875  0.        0.        0.        0.       ]
 [0.        0.0234375 0.        0.125     0.        0.        0.        0.       ]
 [0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.1171875 0.        0.        0.        0.       ]]
```



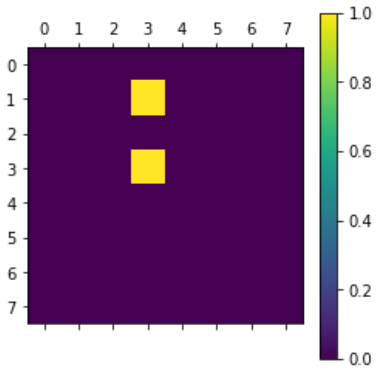
Channel 5: Bounding Box (h)

```
[[0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.171875  0.        0.        0.        0.       ]
 [0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.0859375 0.        0.        0.        0.       ]
 [0.        0.        0.        0.03125   0.        0.        0.        0.       ]
 [0.        0.046875  0.        0.09375   0.        0.        0.        0.       ]
 [0.        0.        0.        0.        0.        0.        0.        0.       ]
 [0.        0.        0.        0.078125  0.        0.        0.        0.       ]]
```



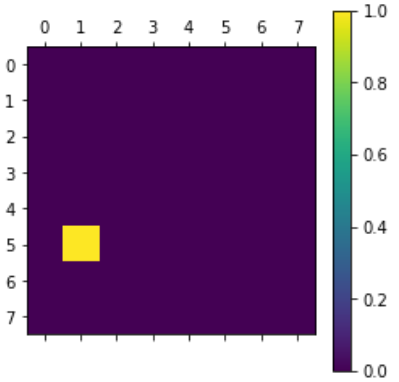
Channel 6: P(Class=Pedestrian)

```
[[0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]]
```



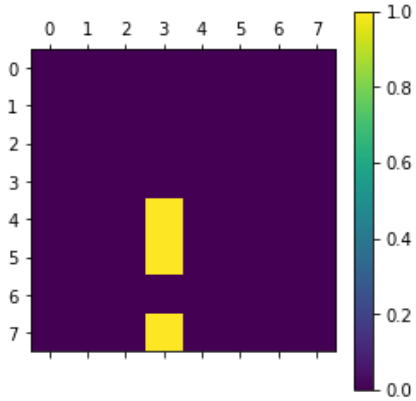
Channel 7: P(Class=Traffic Light)

```
[[0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]]
```

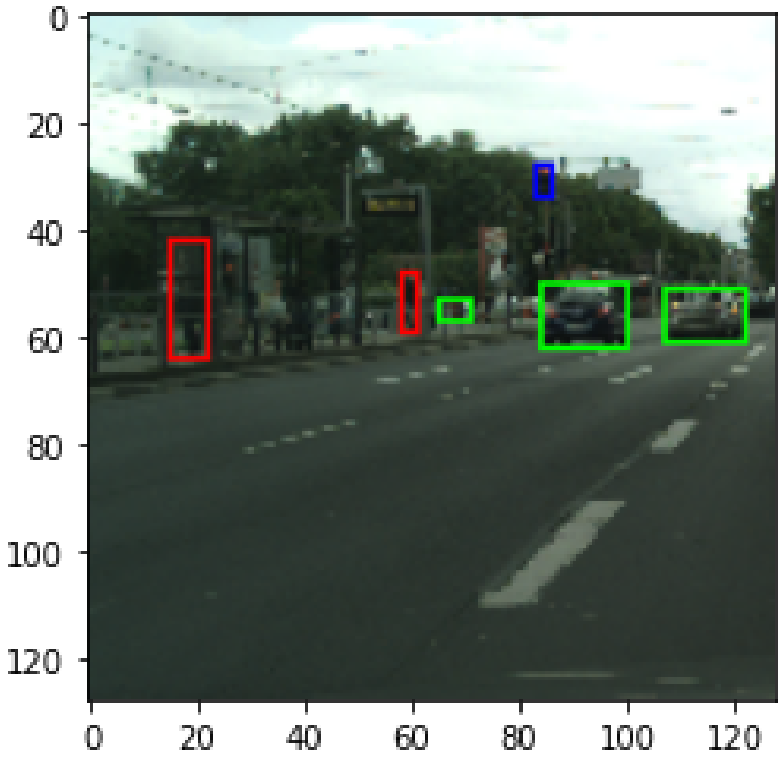


Channel 8: P(Class=Car)

```
[[0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0.]]
```



Reconstructing the processed 8*8*8 labels back to original data format and plotting bounding box over image:



One of the bounding boxes of a car was not recovered as it was in the same grid cell as another. All other objects were successfully detected.

For i = 1000:

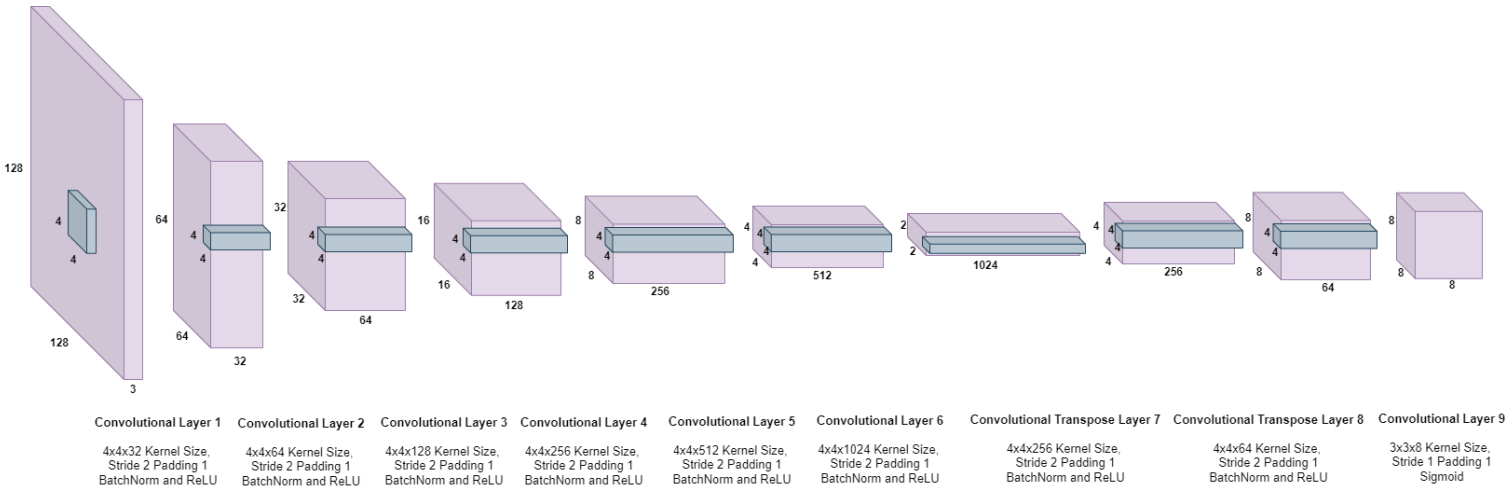
```
raw_labels[i]
```

```
array([[  2.,  84.,  50., 100.,  62.],
       [  2., 120.,  52., 128.,  56.],
       [  2., 107.,  51., 122.,  61.],
       [  0.,  15.,  42.,  22.,  64.],
       [  1.,  83.,  28.,  86.,  34.],
       [  0.,  58.,  48.,  61.,  59.],
       [  2.,  65.,  53.,  71.,  57.]])
```

```
reconstructed_labels[i]
```

```
array([[  0.,  15.,  42.,  22.,  64.],
       [  0.,  58.,  48.,  61.,  59.],
       [  2.,  65.,  53.,  71.,  57.],
       [  1.,  83.,  28.,  86.,  34.],
       [  2.,  84.,  50., 100.,  62.],
       [  2., 107.,  51., 122.,  61.]])
```

Block diagram of YOLO architecture:

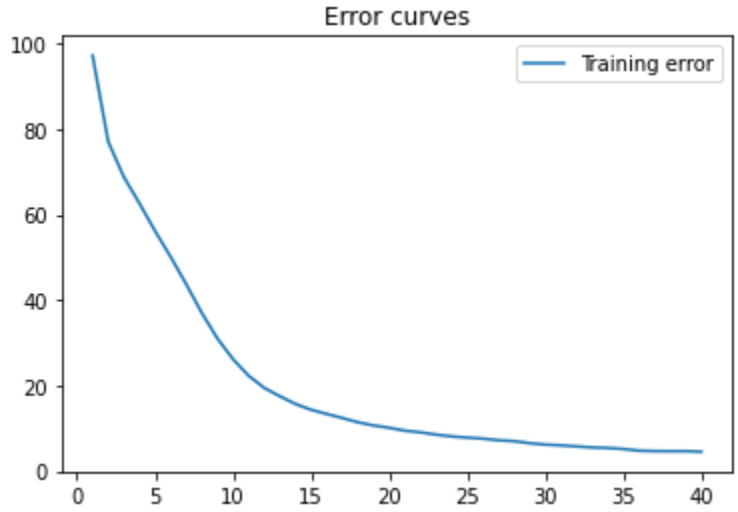


Deviations made:

A sigmoid layer was added at the end of the 9th Convolutional Layer. Since the model can generate values as a probability, while the Probability of Objectness are all between 0 and 1, a sigmoid layer helps constrain the results and improves the performance of the model.

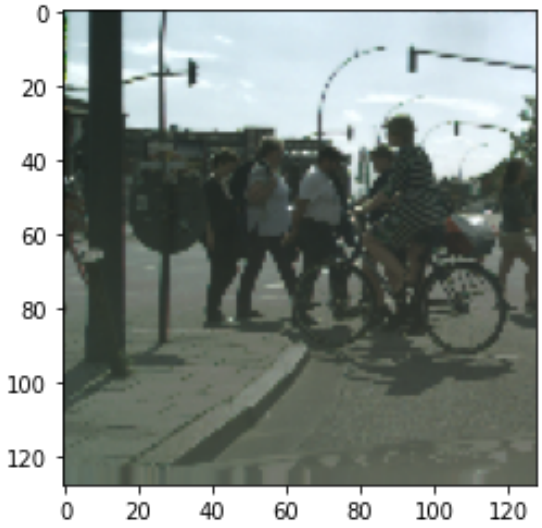
Loss over training:

For number of epochs = 40, learning rate = 0.001

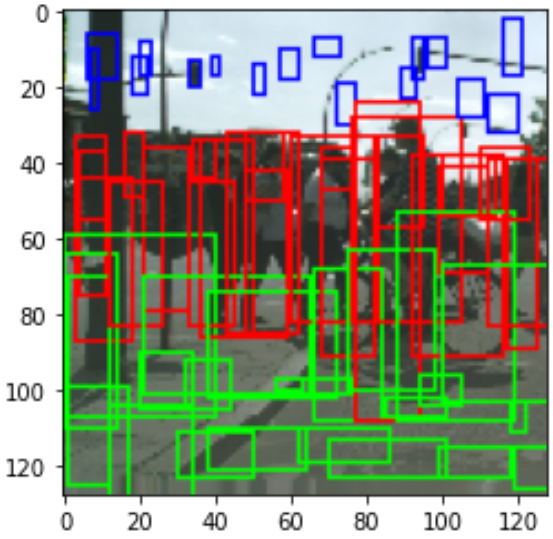


Results

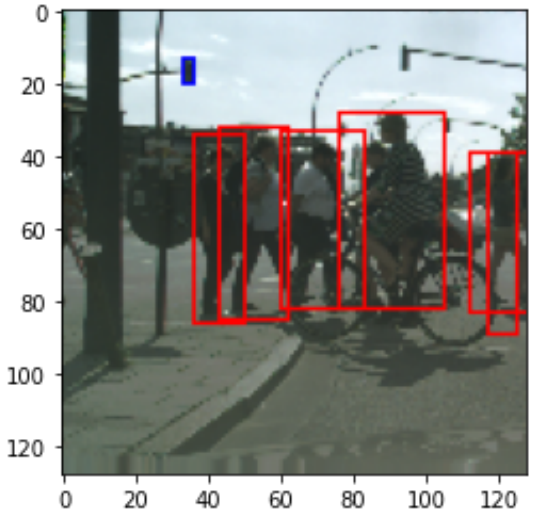
Image



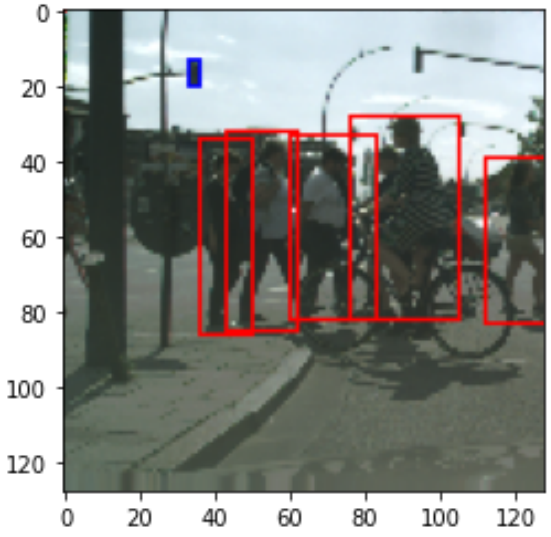
Raw Predictions



After low confidence suppression



After NMS



Final Result