*CS512 Computer Vision Assignment 4*

*Guillermo Lopez-Areal*

*Illinois Institute of Technology*

**

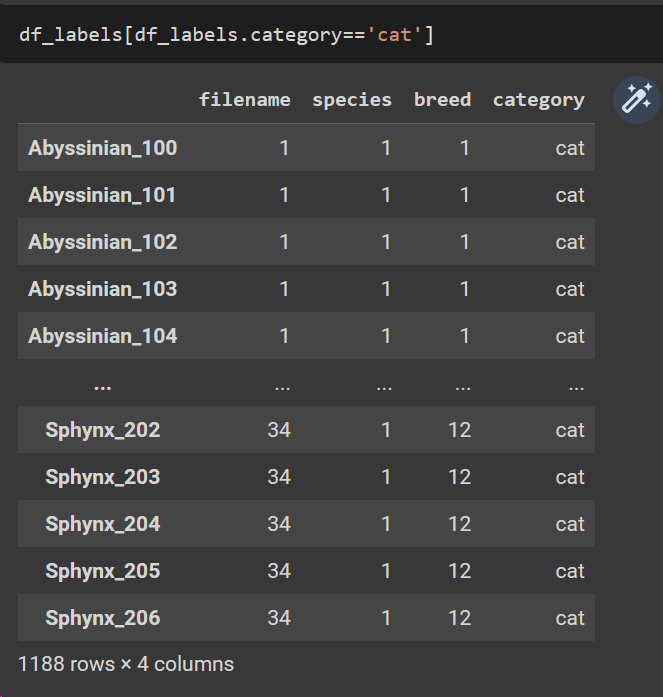
Text

Description automatically generated with low confidence

Text

Description automatically generated  Text

Description automatically generated

Text

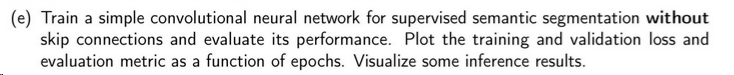
Description automatically generated

A picture containing text

Description automatically generated

Text

Description automatically generated



|  |  |
| --- | --- |
| Text  Description automatically generated |  |
|  |  |
|  |  |



Text

Description automatically generated

We have added six skip connections, one between each decoder layer and its corresponding encoder layer. We concatenate the feature maps from the encoder layer to the feature maps from the decoder layer using the concatenate layer. This helps the decoder layer to recover the fine details lost during the downsampling in the encoder layer. Specifically, we added connections from the output of each encoder layer to a corresponding decoder layer with the same spatial resolution. This enables the decoder layers to better reconstruct the fine details of the input image that were lost during the downsampling operations in the encoder layers.

Chart, line chart

Description automatically generatedHere are the results :

A dog standing on grass

Description automatically generated with low confidenceA picture containing text

Description automatically generated



Text

Description automatically generated

Text

Description automatically generated

|  |
| --- |
|  |
|  |
|  |
|  |
|  |



To run this notebook you might have to add the archive yolo3\_one\_file\_to\_detect\_them\_all\_modified.py, which will be located in the data folder.



To get the bounding box for each object we access the xml files provided when cloning the git repository. In those archives we can access the bounding box of every image in the repository.

Text

Description automatically generated

Here is the method to extract the ground truth coordinates and the results are shown below :

Text

Description automatically generated with medium confidence

A picture containing text, dog, mammal, black

Description automatically generated A cat in a hammock

Description automatically generated with medium confidenceGraphical user interface

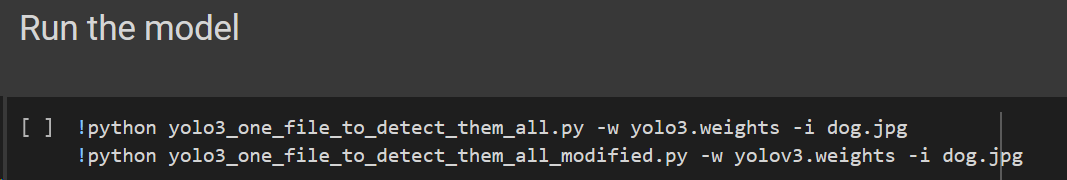
Description automatically generated

The implementation of this first part is in the final of the notebook.



Graphical user interface, website

Description automatically generatedText

Description automatically generated 



Graphical user interface, application

Description automatically generated

Graphical user interface, text, website

Description automatically generated

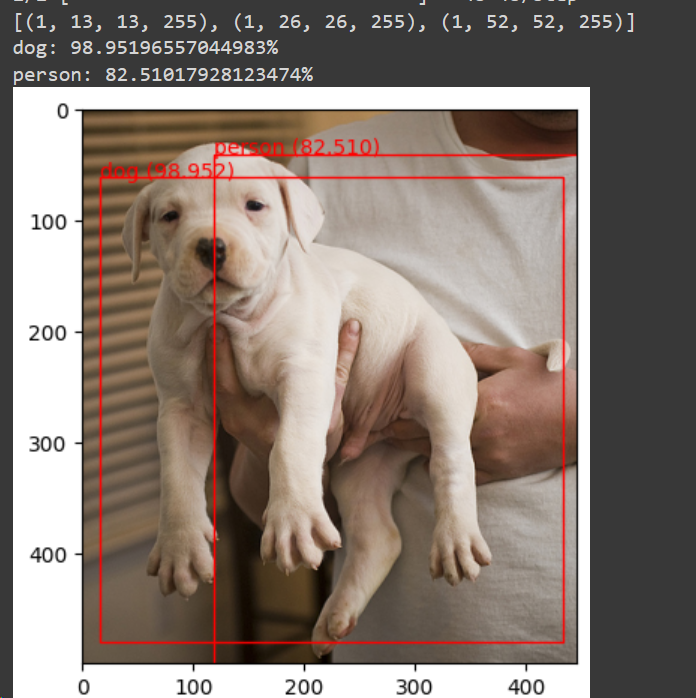


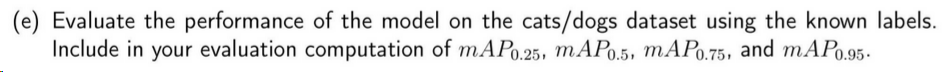
Multiple methods are implemented in the notebook, so I will just comment the results and some of the key methods.

Text

Description automatically generatedHere we first define the input shape of the model to be 416x416 pixels, we then load an image from a file and prepare it for input to the model. Using the yolov3.predict we obtain predictions from the YOLOv3 model.

We then print the shape of the output predictions to verify that the expected shape is obtained, define the anchor boxes used by the YOLOv3 model for object detection. Iterate over the output predictions from the model and use the decode\_netout function to decode the output and extract bounding boxes for detected objects. Apply non-maximal suppression (NMS) to remove overlapping bounding boxes and retain only the most confident detections. Use the draw\_boxes\_plot function to draw the final set of bounding boxes on the original image and save the resulting image.





For this part, for this part first we extracted all the XML data in order to get the ground truth of the bounding boxes. Next, we got the images paths so we were able to get the images. Then, we preprocess those images in order to give it as input to the model and then we obtain the predicted bounding boxes. Now that we have both predictions on ground truth and we can compute the mean average precision for the different IOU (Intersection Over Union) thresholds. The results that we got from the mean average precision are not correct, but I believe the path to get to the right answer is pretty close.

The methods for this part are pretty big, so I will not be including them in the report, they are in the notebook though.