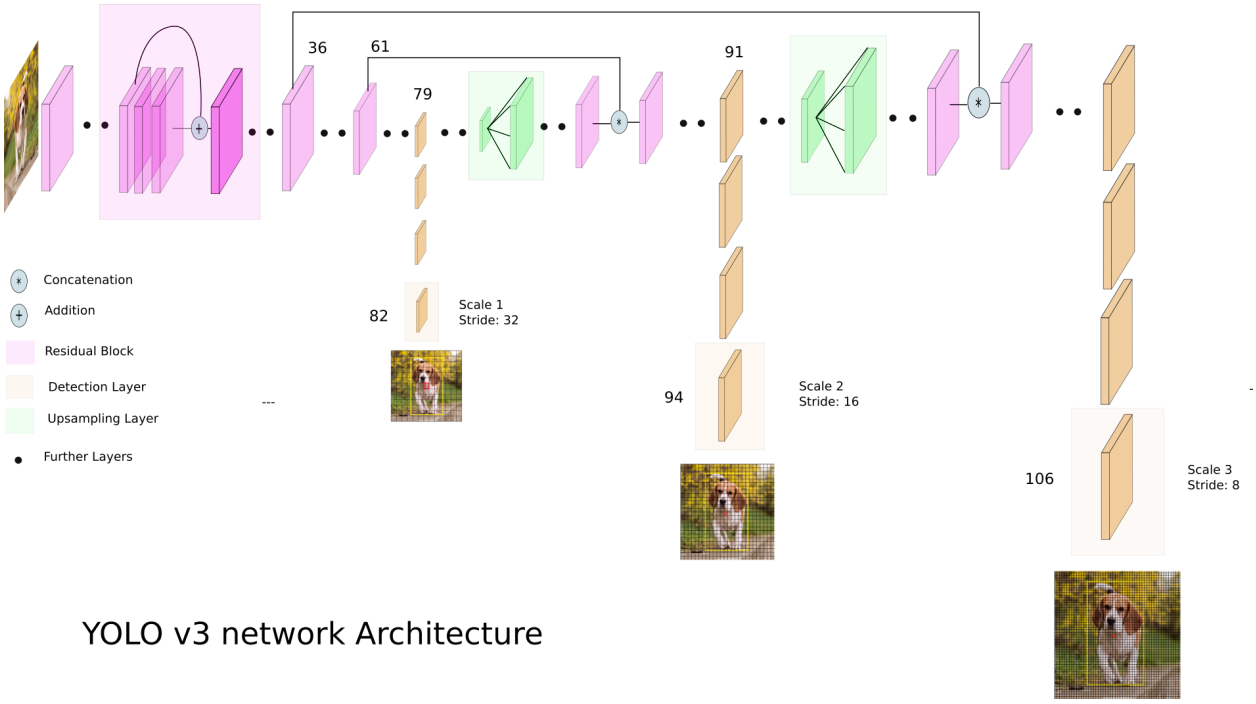
Task: 08

YOLO

YOU ONLY LOOK ONCE -- V3

Architecture:



YOLO v3 network Architecture

1. Yolo uses convolutional layers and yolo version 3 originally consists of 53 CNNs layers which is also called Darknet-53. But for detection tasks original architecture statect 53 more layers that means Yolo v3 uses 106 layers of convolutional neural networks.
2. Darknet framework loaded 106 layers the detection is made at 3 layers: 82 94 & 106.
3. In the architecture of yolo V3 there are some essential elements:
    a. Residual blocks
    b. Skip connections
    c. up-sampling
4. Each CNN layers followed by:
    a. Batch normalization
    b. Leaky ReLU
5. There are no pooling layers but instead additional CNN layers with stride 2 are used to down sample feature maps because use of additional CNN layers to down sample feature maps reveals low level features that the pooling layer just

excluded in a result it helps to improve in detection of small objects & prevent loss of low level features.

6. How does input to network
   Input is batch of images (n, 416, 416, 3)
   [where n-- images, 416 --width & height(height and width may any number which can divisible by 32>16>8), 3-- RGB]
   There is no need to resize, they will be resized to network size.
7. Detection at : 82,94,106 layers
   Downsamples input by 32, 16 & 8 at seperate places of network.
8. Network strides:
   a. 32, 16, 8 are called strides at layers 82, 94, 106 they show the outputat 3 seperete level at network is smaller than input image.
   b. Consider if input image is 416*416
      i. strike of image is 32 then size of output is 13*13
         (detect large object)
      ii. strike of image is 16 then size of output is 26*26
         (detect medium object)
      iii. strike of image is 8 then size of output is 52*52
         (detect small object)

Improvements:

The average precision for small objects improved, it is now better than Faster RCNN but Retinanet is still better in this.

1. As MAP increased localization errors decreased.
2. Predictions at different scales or aspect ratios for same object improved because of the addition of feature pyramid like method.

object confidence and class predictions in YOLO v3 are now predicted through logistic regression.