Mathematical Review

DaciaSandero

1. Introduction

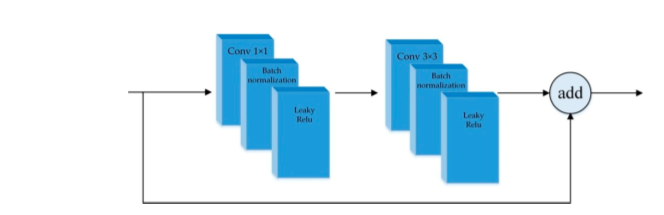
In computer vision field, the target detection is one of the main research point.Target

detection can be used to determine the location and the category of the targets. It can be used in military and also in civil areas , and even in image segmentation,Intelligent surveillance, autonomous driving and intelligent transportation.The rapid development of GPU hardware helped the progress of deep learning. A huge number of algortimhs based on deep learning are used in our daily lives , such as, pedestrian, face and vehicle detection.The convolutional neural network or CNN, is a network with many layers and is used to extract features based on invariance of regional statistic in images. The convolutionl network can learn the featrues which have to be dectected and its performance can be improved gradually , by training it.

2.YOLO v3

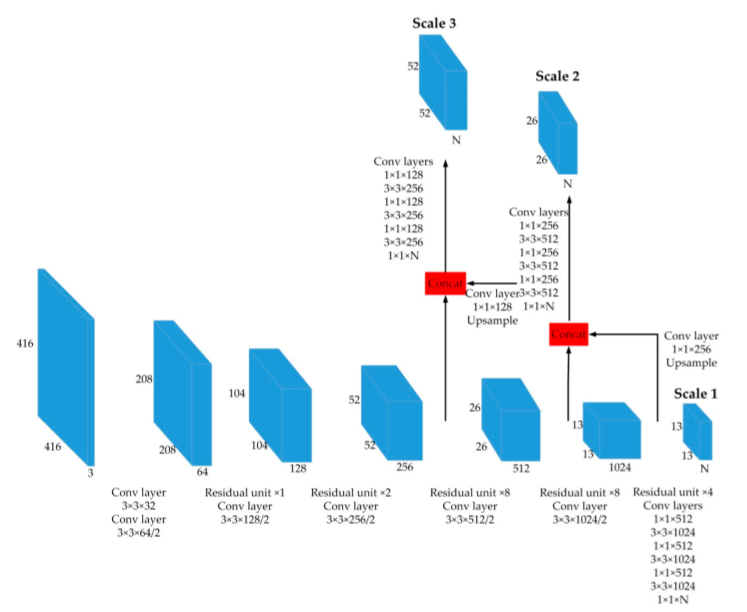
You Only Look Once(YOLO) is the first stage of a target detection algorithm which

has been developed in 3 stages.YOLO v3 is an end to end algorithm and can detect the category of the target and its location directly.YOLO v3 uses successive 3X3 and 1X1 convolutional layers and draws and idea of the residual networks, to perform features extraction YOLO v3 uses 5 residual blocks and each of them is made of multiple “residual units”.

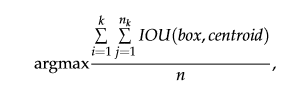


Using residual units , the network can become deeper and the gradient fading can be avoided.

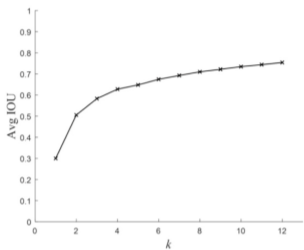
The input is downsampled by five times and the algorithm predicts the target in the last 3 downsampled layers. There are three scales for YOLO v3 to detect the targets.At scale three, the feature map downsampled by 8x is used to detect small targets. At scale two, the feature map downsapled by 16x is used to detect medium-sized targets.At scale 1, the feature map downsampled by 32x is used to detect big targets.Feature fusion is used to detect targets because small features map provides deep semantic information and large feature map provides finer information of the targets.To perform feature fusion, YOLO v3 resizes the feature maps of the deeper layer by upsampling .Then feature maps at different scales will have the same size.YOLO v3 merges the feature from the earlier layer both the features from deeper layer by concatenation.So YOLO v3 has good performance to detect both large and small targets.



YOLO v3 introduced the idea of anchor boxes used in Faster R-CNN. Those boxes are a set of initial candidate boxes with a fixed width and length.The coice of the initial anchor boxes will directly affect the accuracy of detection and its speed .Insted of choosing anchor boxes by hand , YOLOv3 runs K-means clustering on the dataset to find good priors automatically.The clusters generated by K-means can reflect the distribution of the samples in each dataset, which can make it easier for the network to get good predictions.AvgIOU is a metric of target clustering analysis and its objective function of clustering is:



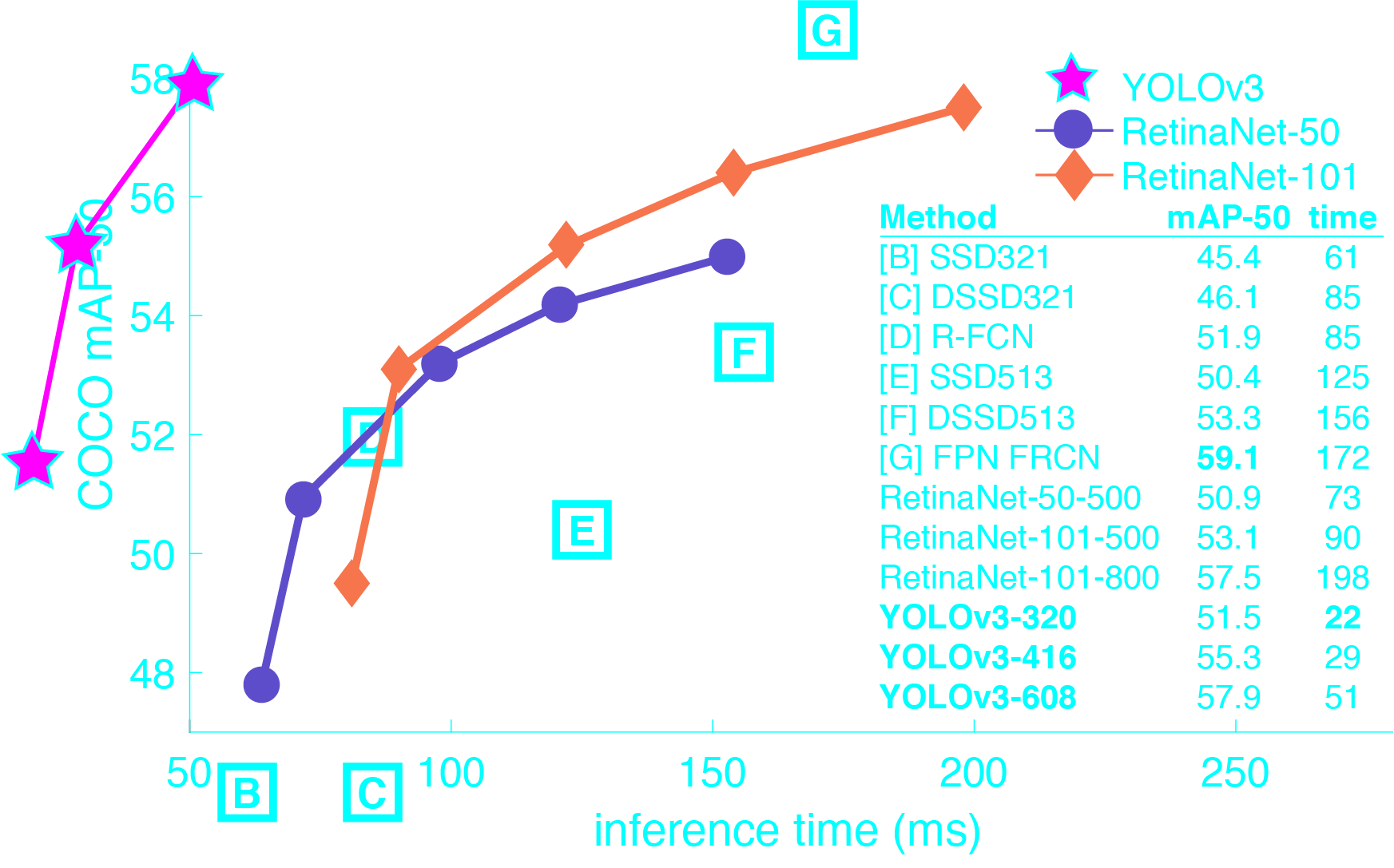
The “box” is the sample , “centroid” is the center of the cluster . represent the numbers of samples in the k-th cluster center, “k” is the total number of samples,”n” is the number is clusters and “IOU(box,centroid)” is the intersection over Union of the clusters and the samples.After applying the K-means clustering the results shows that the increase of k make the objective function more stable(AvgIOU).

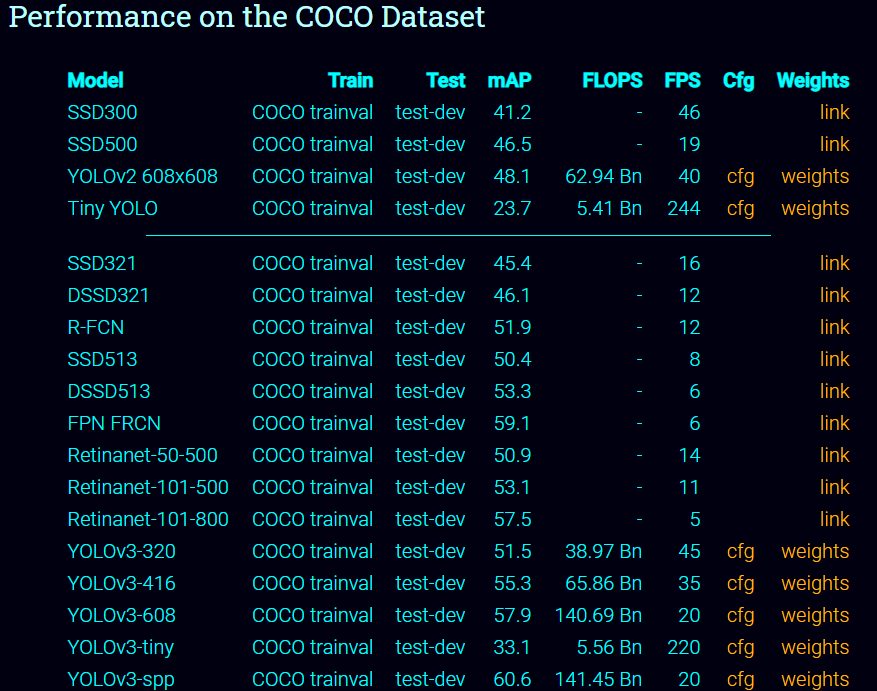


After the generation of cluster center , YOLO v3 divides up them evenly across scales.Because the clusters are arbitrarily allocated , this may place some clusters at inappropriate scales.

3.Comparison with other detectors

YOLO v3 is fast and also accurate ,which is the most important part. In mAP measured at .5 IOU the algorithm is as accurate as Focal Loss but around 4 times faster. The following results are based on the COCO dataset:





YOLO v3 is a good detector.It’s fast and accurate .Is not quite great on the COCO average AP between .5 and .95 IOU metric , but it perform outstanting on the old metric of .5 IOU.

4.Conclusion

With the rapid development of technologies such as smart phones, autopilot cars , and drones , more and more embedded devices have been endowed with computer vision function. For example, Autopilot technology require vehicles are able to sense environment, parse scene and react accordingly; Artificial Intelligence (AI) retouching and other functions in the smartphone need to accurately locate the object position in the image; Object detection might be the most common one that is adopted as a basic functional module for scene parsing in embedded applications, and hence it has been the area of increasing interest

5.Reffrence

1.Ju M. ,Luo H. ,Wang Z. ,Hui B., Chang Z. , (2019),“The Application of Improved YOLO V3 in Multi-Scale Target Detection”.

2.Redmon J. ,Farhadi A. ,(2018) , “YOLOv3: An Incremental Improvement”

3.Mao Q. , Sun H. , Liu Y. ,Jia R. , (2019) , “Mini-Yolov3:Real-Time Object Detector for Embedded Applications”

4.Tian Y. , Yang G. , Wang Z. , Wang H. , Li E. , Liang Z. , (2019) , “Apple detection during different growth stages in orchards using the improved YOLO-V3 model”

5.Zhang Y. , Shen Y. , Zhang J. , (2019) , “An improved tiny-yolov3 pedestrian detection algorithm”