International Research Journal of Engineering and Management Studies (IRJEMS)
Volume: 03 Issue: 03 | March -2019 e-ISSN: 2395-0126

Object Detection Using Deep Learning

Chillapalli Venkata Kalyani, Choppara Divya, Gudipati Vasavi, Chintalachervu Arundhathi

Vasireddy *Venkatadri* Institute of Technology, Nambur, Guntur District, Andhra Pradesh

---

Abstract – *Object detection made an evolutionary record in the field of deep learning. One of the applications of deep learning is object detection. Years goes on, the importance of object detection would be more and more popular. As there are so many techniques for object detection, we select the Mask R-CNN as the best method. Because Mask R-CNN given the better accuracy. The output of object detection is not only object label and bounding box but also mask for each object. The advantage of Mask is pixel wise segmentation for each object in an image. In Mask R-CNN we are using instance segmentation algorithm for representing the mask. Mask R-CNN is very simple to train and running at 5fps.*

***Key Words*: Deep Learning, Faster RCNN, Mask RCNN, Instance Segmentation.**

1. INTRODUCTION

Object detection using deep Learning is a very popular technique. Deep learning had wider applications in the field of object detection. There are so many deep Learning methods starting from CNN to Faster RCNN each having an improvement during its evolution. In our project we introduce a completely different deep learning approach called Mask R-CNN, and in this paper we will demonstrate clearly about why we chose the method Mask R-CNN among all traditional methods. Our main aim is to detect objects in an image and video with high degree of accuracy. In order to build our design we use COCO (common objects data set) dataset for training the images. For testing purpose we can chose any real time images. The main theme of our project is to detect the objects in an image with high accuracy. Instead of just detecting the objects in an image our new approach will also detects the objects in the video also. In this method we use an instance segmentation algorithm to get the mask for each instance. Our project better suited for an applications including security, face detection, Vehicle detection, people counting, online images and manufacturing industries.

2. RELATED WORK

The problem of object detection had solved by using many methods we will discuss here one by one.

2.1 Convolutional Neural Network (CNN)

Convolutional neural network is the first deep Learning technique to detect the object in an image. This was introduced by Yann LeCun. He is a French computer scientist working primarily in the field of machine Learning, computer vision, mobile robotics and computer vision using CNN, invented in the year of 1994 and is a founding father of convolutional nets [1].

2.1.1 Drawbacks of CNN

A CNN can only tell you the class of the objects but not where they are located. Actually it is possible to regress bounding boxes directly from a CNN but that can only happen for one object at a time. If multiple objects are in the visual field then the CNN bounding box regression can't work well due to interference. The problem with CNN is that the objects of interest may have different spatial locations within the image and different aspect ratios. Therefore, you would have to select a huge number of regions and this could computationally blow up. Therefore, algorithm like R-CNN has been developed to find these occurrences and find them fast.

2.2 Regional Convolutional Neural Network

To bypass the problem of selecting a huge number of regions, Ross Girshick et al proposed a new method called regional convolution neural network(R-CNN). Ross Girshick was a research scientist at face book AI Research (FAIR), working on computer vision and machine learning. He had received the 2017 PAMI young Researcher award and is well-known for developing the R-CNN (Region-based convolutional Neural Network) approach to object detection [2]. Where they used Selective search to extract just 2000 regions from the image and he called them region proposals.
Therefore, now instead of trying to classify a huge number of regions, we can just work with 2000 regions. These 2000 regions proposals are generated using the selective search algorithm [3].

2.2.1 Drawbacks of R-CNN

1. It still takes a huge amount of time to train the network as you would have to classify 2000 regions proposals per image.
2. It can't be implemented real time as it takes around 47 seconds for each test image.
3. The selective search algorithm is a fixed algorithm
   Therefore, no learning is happening at that stage. This could lead to the generation of bad candidate region proposals [3].
4. Slow at test time: need to run full forward pass of CNN for each region proposal.

5. SVM s and regressors are post-hoc: CNN features not updated in response to SVMs and regressors.
6. Complex multistage training pipeline. [7]

To overcome these drawbacks next deep learning technique called Fast R- CNN had comes into picture.

2.3 Fast R-CNN

Fast R-CNN algorithm was invented by Ross Girshick. He is a research scientist at face book AI Research (FAIR), working on computer vision and machine learning [4].Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy[5].

2.3.1 Fast R-CNN results

[6]

Method	R-CNN	Fast R-CNN
Training time	84 hours	9.5 hours
(speed up)	1x	8.8x
Test time per image	4.7 seconds	0.32 seconds
(speed up)	1x	146x

2.3.2 Drawbacks of Fast R-CNN

1. Testing time don't include region proposals [7].
2. The performance of Fast R-CNN during testing time, including region proposals region proposals slows down the algorithm significantly when compared to not using region proposals. Therefore, region proposals become bottlenecks in Fast R-CNN algorithm affecting its performance [3].

2.4 Faster R-CNN

In the middle 2015, team at Microsoft Research composed of Shaoqing Ren, Kaiming. He, Ross Girshick and jian sun found a way to make the region proposal step almost cost free through an architecture they named as Faster R-CNN [8].

2.4.1 Faster R-CNN results

As We can compare the Faster R-CNN with R-CNN and Fast R-CNN , the test time per image is very less in the faster R-CNN so we can conclude that Fast R-CNN have more processing speed when compared to remaining.

method	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	0.2 seconds
speedup	1x	25x	250x
mAP(VOC 2007)	66.0	66.9	66.9

2.4.2 Drawbacks of Faster R-CNN:

Faster R-CNN was not designed for pixel-to-pixel alignment between networks inputs and outputs. To overcome the drawbacks of traditional methods like CNN,R-CNN, Fast R-CNN and Faster R-CNN, new method have been introduced called Mask R-CNN.

3. METHODOLOGY

3.1 Mask R-CNN

To overcome the drawbacks what we discussed earlier, a new method called Mask R-CNN was introduced. The approach looked at here while simultaneously generating a high-quality segmentation mask for each instance is efficient enough to detect objects in an image. This method, named Mask R-CNN by addition of order to predict an object mask in parallel with the existing branch for bounding box recognition extends Faster R-CNN. Mask R-CNN, running at 5 fps is relatively simple to train and adds only a small overhead to Faster R-CNN [9].

3.2 Mask R-CNN architecture

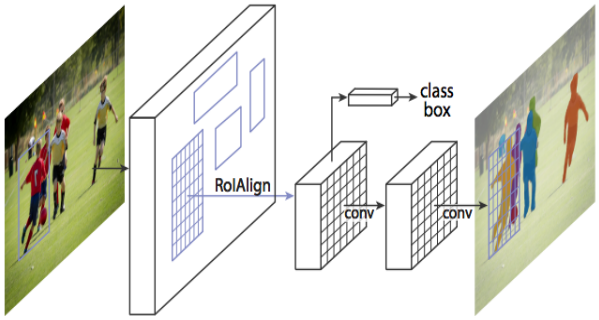

Fig .1Mask R-CNN

Mask R-CNN (regional convolutional neural network) is a two stage framework: the first stage scans the image and generates *proposals* (areas likely to contain an object). And the second stage classifies the proposals and generates bounding boxes and masks [10]. The mask branch is small FCN applied to each RoI, predicting a segmentation mask in a pixel-to-pixel manner. Mask R-CNN is simple to implement and train given the Faster R-CNN framework, which facilitates a wide range of flexible architecture designs. Additionally, the mask branch adds a small computational overhead, enabling a fast system and rapid experimentation.

Fig 3.3 Instance segmentation

We perform instance segmentation for Mask R-CNN algorithm. In instance segmentation the goal is to detect specific objects in an image and create a mask around the object of interest of interest. Instance segmentation can also be thought as object detection where the output is a mask instead of just bounding box. Unlike semantic segmentation, which tries to categorize each pixel in the image, instance segmentation does not aim to label every pixel in the image [11].

3.3 ROI-align

In each ROI bin, the value of four regularly sampled locations are computed directly through bilinear interpolation. Thus avoid the misaligned problem. The ROI align is reported to have ~3 points improvement in AP in trainval35k [12].

3.4 Working of Mask R-CNN [13]

General steps for the approach of Mask R-CNN

1. **Backbone model**:

A convolutional neural network that serves as a feature extractor. For example, it will turn a 1024x1024x3 image into a 32x32x2048 feature map that serves as input for the next layers.

2. **Region proposal Network (RPN):**

Using regions defined as many as 200k anchor boxes, the RPN scans each region and predicts whether or not an object is present. One of the great advantages of the RPN is that does not scans the actual image, the network scans the feature map, making it much faster.

3. **Region of Interest Classification and Bounding Box:**

In this step algorithm takes the regions of interest proposed by the RPN as inputs and outputs a classification (softmax) and bounding box (regressor).

4. **Segmentation Masks**

In the final step, the algorithm such that positive ROI regions are taken in as inputs and 28x28 pixel masks with float values are generated as outputs for the objects. During interference, these masks are scaled up.

3.5 Loss function

The multi-task loss function of Mask R-CNN combines the loss of classification and segmentation mask [14].

L(total) = L(class) + L(box) + L(mask)

4. RESULTS AND DISCUSSIONS

4.1 COCO data set

For training the images we used COCO data set. COCO stands for common objects in context. The COCO is an excellent object detection dataset with 80 classes, 80k training images and 40k validation images. COCO is large scale object detection, segmentation and captioning dataset COCO has several features [15].

- Object segmentation
- Recognition in context
- Super pixel stuff segmentation
- 330k images (>200k labeled)
- 1.5 m object instances

- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250k people with key points

4.2 Result

For testing the Mask R-CNN, we can give any images with in the bounding labeled range of COCO dataset.

Input image:

Fig 2 Input Image

Output for the given input image as follows

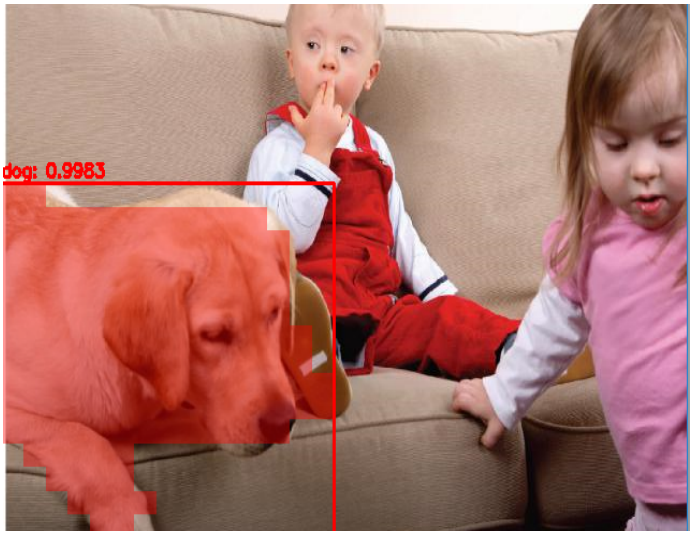

Fig 1.Output Image1

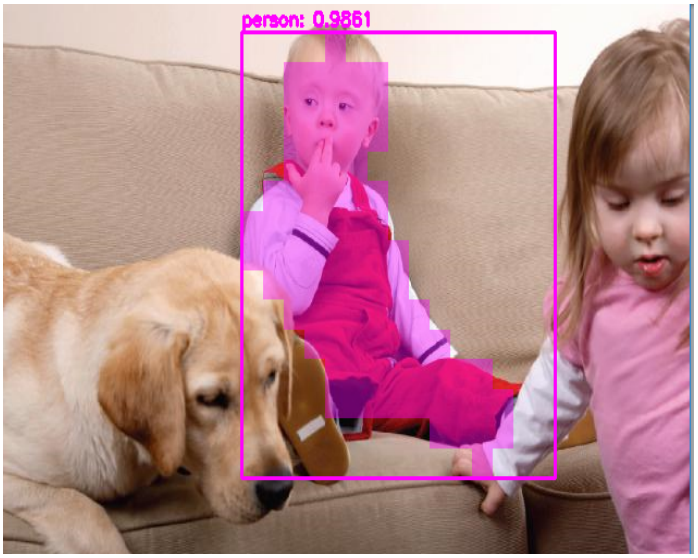

Fig 2 Output Image 2

Fig 3 Output Image 3

Fig Mask R-CNN output

As shown in the output, along with class label and bounding box we also get the mask output. Mask having different colors to differentiate the each object in an image. Object with high score nearer to 1 made mask R-CNN attractive compared to other techniques. One of the advantages of Mask R-CNN is that we can get the output with high degree of accuracy. Our model, not only detects each objects in an image but also detect the objects in the video.

5. CONCLUSION

Hence Deep Learning had made its mark in the field of object detection. In our paper we proved that, though having so many object detection algorithms the reason for preferring the Mask R-CNN is its extraordinary features. The Advantage of mask R-CNN is mask output with

different mask colors such that we can easily differentiate the objects in the foreground and objects in the background using instance segmentation algorithm. The COCO data set for training the model. So many real time applications starting from security to vehicle detection everywhere we can observe the tremendous growth in the deep learning approach. Our model not only detects objects in the image but also detects each every object present in the video with high degree of accuracy.

6. REFERENCES

[1] Convolutional Nets and CIFAR-10: An Interview with Yann LeCun

[2] Ross Girshick, "rbg's home page"

[3] Rohit Gandhi, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms", july 9,2018

[4]https://docs.microsoft.com/en-us/cognitive-toolkit/object-detection-using-fast-r-cnn

[5] https://arxiv.org/abs/1504.08083

[6] spatial localization and detection by justin johnson, andrej karpathy, Fei-Fei Li

[7]http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

[8] a-brief-history-of-cnns-in-image-segmentation-from-r-cnn-to-mask-r-cnn-34ea83205de4

[9] kirti bakshi , "mask-r-cnn-for-object-detection-and-instance-segmentation-on-keras-and-tensorflow", july 1, 2018

[10] waleed abdulla,"Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow", Mar 20, 2018

[11] sunita nayak, "Deep learning based Object Detection and Instance Segmentation using Mask R-CNN in OpenCV (Python / C++) | Learn OpenCV"

[12] sheng hu, "ROI-pooling and ROI-align",Mar 30, 2017

[13] Gabriel Garza, "Mask R-CNN for Ship Detection & Segmentation",jan 7

[14] Lilian Weng,"Object Detection for Dummies Part 3: R-CNN Family" dec 31, 2017

[15] http://cocodataset.org/#home,"COCO - Common Objects in Context"