SSD

Single shot MultiBox Detector

SSD deep learning object detection Model.

Researchers have used different deep Neural Networks such as VGG, ResNet or MobileNet. as feature extractor or object classifier for SSD.

People have implemented SSD under different deep learning platforms such as Caffe, pyTorch or TensorFlow.

	SSD	Yolo V2
Object classification model	MobileNet	Darknet-19
pre-trained Dataset	COCO	COCO
deep learning software platform	Tensorflow	DarkNet
Opencv ReadNet function call	readNetFromTensorflow	readNetFromDarknet
Config file	frozen_interference_graph.pb	yolov2.cfg
weight file	ssd_mobilenet_v1_coco_2017_11_17.pbtxt	yolov2.weights
opencv forward function return	(1, 1, 100, 7)	(845, 85)
Representation of Bounding Box	(left, top, right, bottom)	(x-center, y-center, width, height)
Representation of Class Name	90	80

Value of score Indicates the confidence/probability of detected objects

SSD attains a better balance betn swiftness & precision

SSD runs a convolutional N/w on i/p img only one time

SSD :- SSD Algorithm in the regards of object detection, by using SSD, we only need to take one single shot to detect multiple objects within the image.

Regional proposal network (RPN) based approaches such as R-CNN series that need two shots, one for generating region proposals, one for detecting objects of each proposals. Thus, SSD is must faster compan with two-shot RPN-based approaches.

The SSD detector differs from other single shot detectors due to the usage of multiple layers that provide fine accuracy on object with different scales.

The SSD normally starts with a VGG on Resnet pre-trained model that is converted to a fully Convolution NN.

Then we attach some extra conv layers, which will actually help to handle bigger objects.

SSD architecture can in principle be used with any deep Network base model. after img is passed on the VGG Network, some conv layers are added producing feature maps of size 19x19, 10x10, 5x5, 3x3, 1x1, together with 38x38 feature map produced by VGG's conv4_3, are the feature maps which will be used to predict bounding Boxes. Conv4_3 is responsible to detect smallest object while conv11_2 is responsible for the biggest objects.

Single short - object localization & classification is done in single forward pass of Network

Multibox - Technique for bounding box regression.

Detector - classify the detected objects.

Page No.
Date

The architecture of SSD is build based on VGG-16 architecture, but instead of fully connected layers we use set of Auxiliary convolutional layers from Conv6 layer onwards.

The reason of using VGG-16 is fundamental N/w to its high quality image classification & transfer learning to improve results. using Auxiliary conv. layers we can extract features at multiple scales & progressively decrease the size at each architecture

Ex -

we have image with few horses. we have divided our image into the set of grids. Then we make couple of rectangles of different aspects ~~a ratio~~ around these grids. Then we apply convolution in those boxes to find if there is an object or not in those grids.

If here one of the black horse is closer to the camera in the img, so the rectangle will draw is unable to identify if that is ~~horse~~ is horse or not ? because the rectangle does not have any features that are identifiying to horses.

The architecture of SSD, in each step after conv6 layer the size of images get reduced substantially. Then every operation we discussed on making grids are finding objects on those grids appiles in every single step of convolution going from back to front of N/w. The classifier are appiled ~~to~~ detect ~~single step~~ objects ~~too~~., so since the object become smaller in each steps, they get easily identified.

The SSD algorithm also knows how to go back from one convolution to another. It not only learns to go forward but backword also.

working Mechanism :-

To train our algorithm, we need a training set that contains images with objects & those objects must have bounding boxes on them. the algorithm learn how to put rectangle on the object & where to put. Unlike in CNN, we don't only predict if there is an object in the images or not we also need to predict where the image the object is.

usage

SSD can be used with Kalman filter for vehicle tracking + detection in ~~an~~ · ~~auton~~. autuonomus vehicles

RCNN

Researchers improved CNN for object localization & detection & called this architecture R-CNN (Region -CNN). The output of RCNN is the image with rectangular boxes surrounding for the objects in an image ~~with~~ as well as class of that objects

steps on how R-CNN works :

① Scan the input images for possible object using algo. Called selective search & generate around 2000 region proposals,

② Run CNN over each of these region proposals.

③ Take output of each CNN & feed into
- svm ~~to~~ classify region
- A linear regressor to tighten bounding box of object if such object exists.

Input img
↓
Region of Intrest (ROI) from a proposal method
↓
Weaped image Regions
↓
forward each region through ConuNet
~~class~~ ↓
classify region with SVM (Bounding box reg, SVM)

Although R-CNN made a lot of progress over traditional CNN for object localization, detection & classification it still seems a little problem for achieving this in real time. problems

1) Training data is very difficult to handle & very long
2) Training happens in two stages (region proposal & classification)
3) N/w is slow when dealing with Non-training data.

To improve R-CNN there are also algorithm called fast-RCNN, Faster-RCNN gives more acurrate result for object detection but bit slower for real time detection. so SSD comes into play.

MAP – (mean Average Precision)

mobileNet & SSD info

There are two types of deep Neural Network here. Base Network & detection Network. MobileNet, VGG-Net, LeNet & all ~~are~~ base Networks. Base Network provide high level features for classification or detection. If you use a fully connected layer at the end of this Networks, you have classification.

But you can remove fully connected layer & replace it with detection Network like SSD, faster R-CNN & so on.

Page No.
Date

SSD use of last Convolutional layer on base Network for detection task. MobileNet just like other base N/W use of Convolution to produce high level features.

Receptive - field

It is a region in the input space that a perticular CNN feature is looking at. A receptive field of a feature can be described by its center location & its size.

SSD Basic Idea

we can use shallow (a little depth) layers to predict small objects & deeper layers to predict big objects, as small objects don't need bigger layers to predict big objects, as small object don't need bigger receptive fields & bigger receiptive fields can be confusing for small object.

SSD

SSD was release at the end of Nov 2016 & reached new records in terms of performance & precision for object detection task.