Interesting Posts

- Top Competitive Programmers
- Unsolved Problems in Algorithms

Machine Learning (ML) Tutorial

SSD MobileNetV1 architecture

Machine Learning (ML)

is one of the many deep convolution models
us. In this article, we have dived deep into what is
what makes it special amongst other convolution
work architectures, Single-Shot multibox Detection (
ow **MobileNet V1 SSD** came into being and its
re.

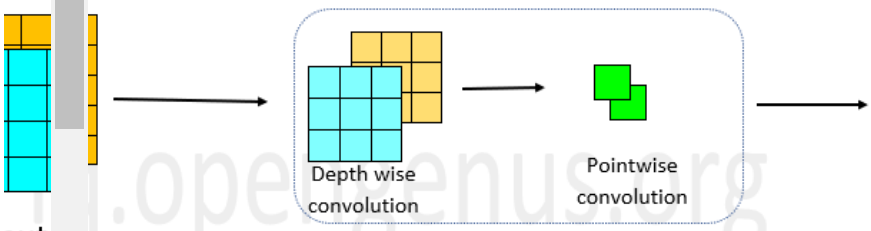
ents:

ileNet

an architechture model of the convolution neural network (CNN) focuses on Image Classification for mobile applications. Rather standard convolution layers, it uses **Depth wise separable layers**. What makes this model stand out is that its architechture s the omputational cost and very low computational power is needed or ap y transfer learning.

leNet V1 architecture

Net is an adaptation of the MobileNet model.



put

ove age depicts the depth wise separable convolution. In mobileNet co lution box in the given image that consists of depthwise and ise nvolutions is **repeated 13 times** after the initial convolution layer ble ow gives its detailed architecture.

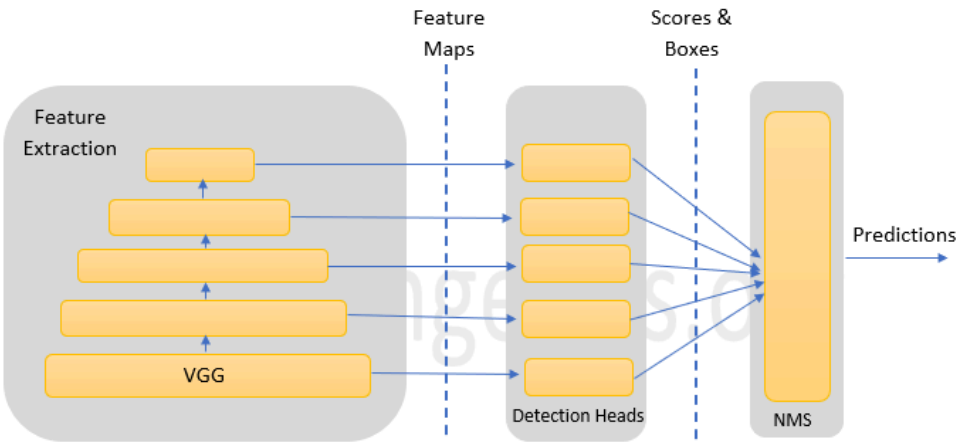
T E/STRIDE	FILTER SHAPE	INPUT SIZE
nv/s2	3 x 3 x 3 x 32	224 x 224 x 3
C v dw/s1	3 x 3 x 32 dw	112 x 112 x 32
nv/s1	1 x 1 x 32 x 64	112 x 112 x 32
C v dw/s2	3 x 3 x 64 dw	112 x 112 x 64
nv/s1	1 x 1 x 64 x 128	56 x 56 x 128
C v dw/s1	3 x 3 x 128 dw	56 x 56 x 128

Conv/s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw/s1	3 x 3 x 256 dw	28 x 28 x 256
Conv/s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw/s1	3 x 3 x 256 dw	28 x 28 x 256
Conv/s1	1 x 1 x 256 x 512	14 x 14 x 256
Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 512	14 x 14 x 256
Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 512	14 x 14 x 256
Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 512	14 x 14 x 256
Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 512	14 x 14 x 256
Conv dw/s1	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 512	14 x 14 x 256
Conv dw/s2	3 x 3 x 512 dw	14 x 14 x 512
Conv/s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw/s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv/s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
Avg Pool/s1	Pool 7 x 7	7 x 7 x 1024
FC/s1	1024 x 1000	1 x 1 x 1024
Softmax/s1	Classifier	1 x 1 x 1000

In the above table, in convolution layer mentioned as **Conv**, the fourth parameter in the column *'Filter shape'* represents the number of filters for the respective conolution layer.

Detector

Single shot Multibox detector is an algorithm which takes only one shot to detect many objects in the image using multibox. It uses a single deep neural network to achieve this. This detector works at a variety of different scales, so it is able to detect objects of various different sizes/scales in the image.Given below is the architecture of SSD:



Generally, SSD uses an auxillary network for feature extraction. This is also called as base network. In the above image, the algorithm uses VGG to extract feature maps. But the last few layers of VGG like the maxpool, FC and Softmax are omitted and the output of VGG is used as feature maps on which to base detections.

More convolution layers are added in which the intermediate tensors are kept, so that a stack of feature maps with variety of sizes are generated to make detection. Let us assume, that we have a feature layer of size a x b and we have c channels. Then the convolution (mostly 3 x 3) is applied on a x b x c feature layer. So for each location of the objects identified, there are k bounding boxes possible each with a probability score assigned to it.

At last, Non-max supression is used to make sure that there's only one bounded box around an object. Its achieved as folows:

Firstly, all the bounding boxes around the objects that has probability less than a certain threshold (say 0.6). Then of the remaining boxes, the box with the greatest probability factor is looked upon for each and every object and the other boxes except the one with maximum probability factor is supressed. Thus leaving only a single bounded box around a single identified object.

Since in this, all the boxes with non-maximum values are supressed, the method is called **Non-maxima Supression**.

SSD MobileNet V1

technology. Since, SSD is independent of its base network, MobileNet was used as the base network of SSD to tackle this problem.

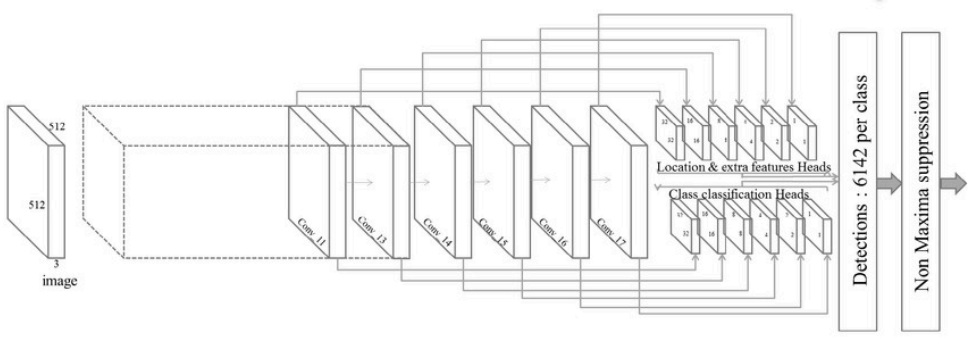
This is known as **MobileNet SSD**.

When MobileNet V1 is used along with SSD, the last few layers such as the FC, Maxpool and Softmax are omitted. So, the outputs from the final convolution layer in the MobileNet is used, along with convolutiong it a few more times to obtain a stack of feature maps.These are then used as inputs for its detection heads. Its architecture can be modified as per required.The table below gives one of its architecture in detail.

TYPE/STRIDE	FILTER SHAPE	INPUT SIZE
Conv/s2	3 x 3 x 3 x 32	300 x 300 x 3
Conv dw/s1	3 x 3 x 32 dw	150 x 150 x 32
Conv/s1	1 x 1 x 32 x 64	150 x 150 x 32
Conv dw/s2	3 x 3 x 64 dw	150 x 150 x 64
Conv/s1	1 x 1 x 64 x 128	75 x 75 x 64
Conv dw/s1	3 x 3 x 128 dw	75 x 75 x 128
Conv/s1	1 x 1 x 128 x 128	75 x 75 x 128
Conv dw/s2	3 x 3 x 128 dw	75 x 75 x 128
Conv/s1	1 x 1 x 128 x 256	38 x 38 x 128
Conv dw/s1	3 x 3 x 256 dw	38 x 38 x 256
Conv/s1	1 x 1 x 256 x 512	38 x 38 x 256
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512

Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv/s2	3 x 3 x 512 x 1024	38 x 38 x 512
Conv/s1	1 x 1 x 1024 x 1024	19 x 19 x 1024
Conv/s1	1 x 1 x 1024 x 256	19 x 19 x 1024
Conv/s2	3 x 3 x 256 x 512	19 x 19 x 256
Conv/s1	1 x 1 x 512 x 128	10 x 10 x 512
Conv/s2	3 x 3 x 128 x 256	10 x 10 x 128
Conv/s1	1 x 1 x 256 x 128	5 x 5 x 256
Conv/s2	3 x 3 x 128 x 256	5 x 5 x 128
Conv/s1	1 x 1 x 256 x 128	3 x 3 x 256
Conv/s1	3 x 3 x 128 x 256	3 x 3 x 128
Conv/s1	1 x 1 x 256 x 128	1 x 1 x 256
Conv/s1	3 x 3 x 128 x 256	1 x 1 x 128

Given below is a pictorial representation of MobileNet V1 based SSD architecture pattern.



By the end of this article at OpenGenus, you will have a clear idea on SSD MobileNet architecture.

Sanjana Babu
Sanjana Babu is an Intern at OpenGenus. She is a Philomath.

Read More

Improved & Reviewed by:

Benjamin QoChuk, PhD

OpenGenus Tech Review Team

Learn Better with Human Guidance

Stepwise Regression

Guide to Deep Learning Model Training and Quantization

See all 636 posts →

decompression

In this article, we will learn about the Lempel Ziv Welch compression and decompression algorithm, a famous compression technique that is widely used in Unix systems and GIF format files

ADITH NAREIN T

ALGORITHMS

Lomuto Partition Scheme

We have explained the Lomuto partition scheme, which is used in the famous Quicksort algorithm. It is an algorithm to partition an array into two parts based on a given condition.

VANSH PRATAP SINGH