Journal: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications
LINK
Group Members:
Sawan Kumar: MCA'23
Hemant Singh: MCA'23

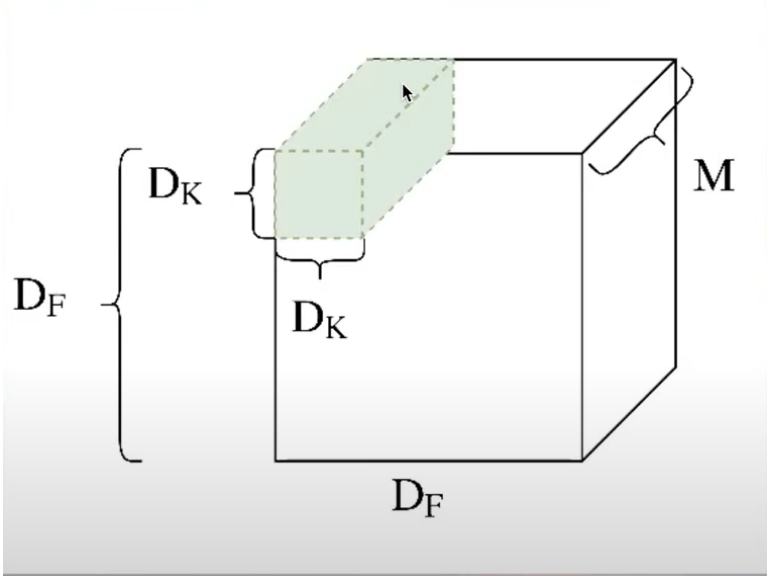
(This document contains a single question (Q1) with its four parts (a), (b), (c), and (d).)
(Total marks: 14)

Q1. Considering the following figures where:

$D_F$ = spatial width of input = spatial height of output
$D_K$ = spatial width of filter = spatial height of filter
M = number of channels in input = number of channels in filter
N = number of filters

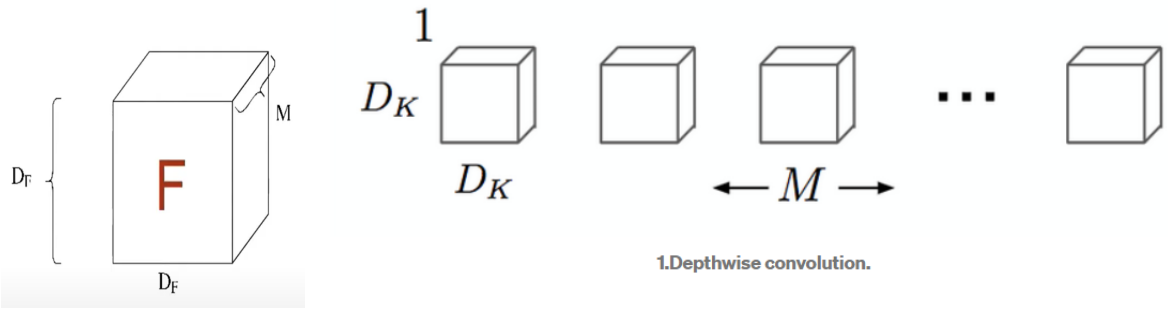
(a) Find the computation cost in each of the three cases while stating the approach used :

Case 1: Standard convolution filter with N filters is applied. [2 marks]



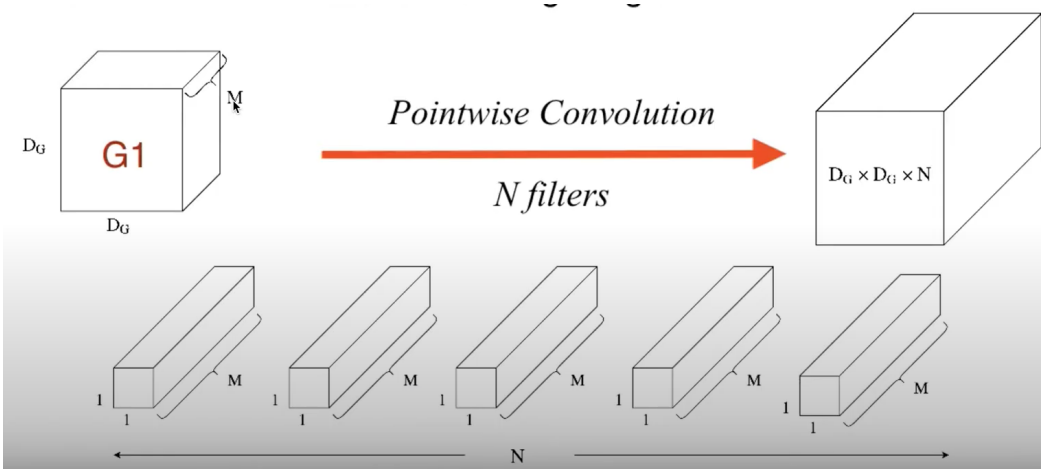

Answer: Cost = $D_K.D_K.D_F.D_F.M.N$
Since we follow the same padding, the height and the width of output remains the same as that of input. Each filter is applied throughout all the channels in input. Cost of single layer in output = $D_K.D_K.D_F.D_F.M$. and number of layers in output = N which is equal to number of filters used.

Case 2: Depthwise convolution is applied: [2 marks]




Answer: Cost = $D_K.D_K.D_F.D_F.M$
Here, Each filter has only one channel. Different filters act on different channels of input. Same padding is maintained. First filter acts on channel 1 , the second filter acts on channel 2 and so on.

Case 3: Pointwise convolution is applied on result obtained after applying depthwise convolution: [2 marks]




Answer: Cost = $D_F.D_F.M.N$
Single filter is applied throughout all the input channels. Same padding is maintained, though valid padding would also produce the same result as height and width of filter = 1. Total channels in output = N which is equal to the number of filters used.

(b) Using the results in previous three cases show which one between depth-wise separable and standard convolution is better. [3 marks]

Answer: (Depth-wise separable cost) / (standard cost)
=((depth-wise cost) + (pointwise cost))/(standard convolution cost)
= $((D_K.D_K.D_F.D_F.M) + (D_F.D_F.M.N)) / (D_K.D_K.D_F.D_F.M.N)$

= $1/N + 1/(D_k.D_K)$

Clearly, the cost of depth-wise separable is a very small fraction of the cost of standard convolution. Hence depth-wise separable is more efficient cost wise as it massively reduces the number of computations.

(C) Go through the following table and fill the blanks A and B. [2 + 2 marks]

'Conv' means convolution
's2' means stride = 2
's1' means stride = 1
'dw' means depth-wise

Table 1. MobileNet Body Architecture

| Type / Stride | Filter Shape | Input Size |
|---|---|---|
| Conv / s2 | $3 \times 3 \times 3 \times 32$ | $224 \times 224 \times 3$ |
| Conv dw / s1 | $3 \times 3 \times 32$ dw | $112 \times 112 \times 32$ |
| Conv / s1 | $1 \times 1 \times 32 \times 64$ | $112 \times 112 \times 32$ |
| Conv dw / s2 | $3 \times 3 \times 64$ dw | **Blank A** |
| Conv / s1 | $1 \times 1 \times 64 \times 128$ | **Blank B** |
| Conv dw / s1 | $3 \times 3 \times 128$ dw | $56 \times 56 \times 128$ |
| Conv / s1 | $1 \times 1 \times 128 \times 128$ | $56 \times 56 \times 128$ |

Answers:

Blank A: = 112*112*64
Explanation:
Considering Row 3:
Input: Filter:
112* 112*32 1*1*32*64
Convolution , stride = 1
Output:
"Same" convolution is applied:
Here pointwise convolution is applied:
No. of filters used = 64
Hence, output size = $N_H$*$N_W$*32 = 112*112*64

Blank B = 56*56*64
Explanation: Using the value obtained in Blank A which becomes input to calculate Blank B

Considering Row 4:
Input: Filter:
112* 112*64 3*3*64
Convolution: dept-wise , stride = 2
Output:
“Same” convolution is applied:
Therefore, N + 2P – F + 1 = N
P = (F-1)/2 = (3-1)/2 = 1
$N_H = N_W$ = FLOOR [($N_H$ + 2P – F ) / S + 1]= FLOOR[(112 + 2*1 – 3)/2 + 1]
= FLOOR[111/2 + 1] = FLOOR[55.5 + 1] = FLOOR[56.5] = 56
No. of filters used = 64
Hence, output size = $N_H$*$N_W$*64 = 56*56*64

(d) Despite the fact that depth-wise separable compromises on accuracy, why is it still preferred ? [1 mark]

Answer: With slight or negligible decrease in accuracy we get a very sharp decrease in the cost so we trade off between cost and accuracy. It is acceptable to have a bit less accuracy if in return our cost falls by a large magnitude. We prefer cost decrement to accuracy till a certain acceptable limit.