1 Task - MobileNets V1 - V3

MobileNets apply a particular method called depthwise separable convolution. Checking out the papers (this was introduced in the first) could be of great help.

The task

With this technique a lot of computations are spared, especially when working with large images and many feature maps.

We here have an image X, over three channels (RGB for instance), and want to start its travel through a MobileNet.

Given :

$$X = [X_1, X_2, X_3]$$

$$X_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 2 & 1 \end{bmatrix},\ X_2 = \begin{bmatrix} 2 & 1 & 1 & 2 \\ 1 & 0 & 2 & 2 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix},\ X_3 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 2 & 1 \end{bmatrix}$$

$$F_{D1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ F_{D2} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},\ F_{D3} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

$$F_P = \begin{bmatrix} 2 & 1 & 1 \end{bmatrix}$$

$$Activation\ h - swish[x] = x\frac{ReLU6(x+3)}{6}$$

$$With\ stride = 1,\ and\ bias\ b = 0$$

$$Where$$

$$X_1 \star F_{D1} = \begin{bmatrix} 2 & 0 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 3 \end{bmatrix},\ X_2 \star F_{D2} = \begin{bmatrix} 3 & 1 & 3 \\ 3 & 1 & 3 \\ 1 & 2 & 4 \end{bmatrix}$$

Find :

1) *The number of operations needed to produce* 10 *feature maps from X using*
a) 2x2x3 *kernels (conventional convolution)*
b) depthwise seperable convolution with 2x2 *kernels for detphwise convolution,* (1)
and one 1x1x3 *kernel for pointwise convolution.*

2) *One feature map of X applying the filters F with depthwise seperable convolution*