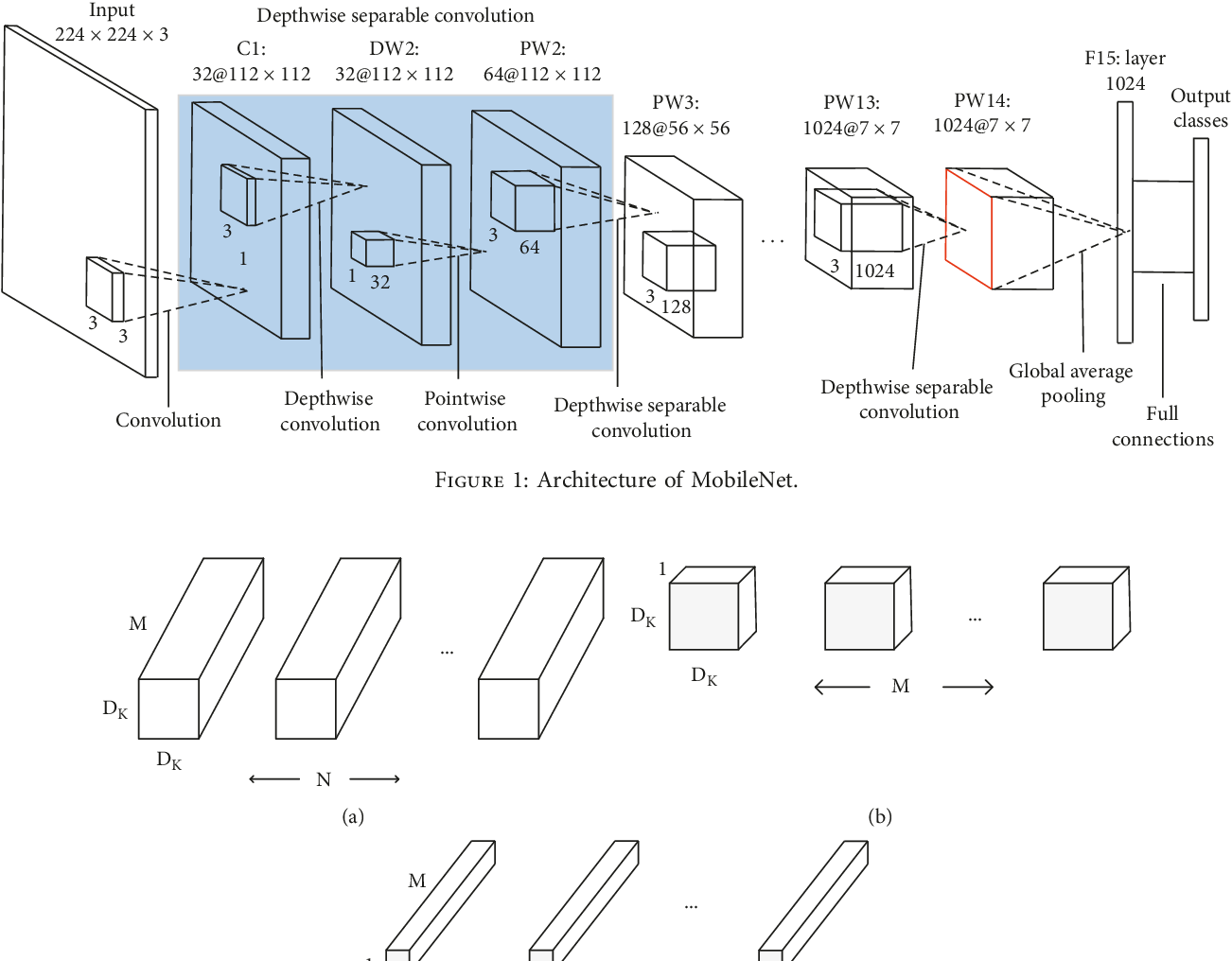
**Mobilenet v2 Architecture**

* MobileNet-v2 is a convolutional neural network that is 92 layers deep. The network has pretrained version which trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

# It’s one of the applications of Depth-wise Separable Convolutional Neural Networks

****

# Diagram Description automatically generatedUnderstanding Normal Convolution operation

* Suppose there is an input data of size **Df x Df x M**, where **Df x Df** can be the image size and **M** is the number of channels (3 for an RGB image). Suppose there are **N** filters/kernels of size **Dk x Dk x M**. If a normal convolution operation is done, then, the output size will be **Dp x Dp x N**.
* *The number of multiplications in 1 convolution operation = size of filter =* ***Dk x Dk x M***
* *the total number of multiplications become* ***N x Dp x Dp x******(Multiplications per convolution)***
* ***Total no of multiplications = N x Dp2 x Dk2 x M***

# Depth-Wise Separable Convolutions

* This process is broken down into 2 operations :
* Depth-wise convolutions
* Point-wise convolutions

**1-DEPTH WISE CONVOLUTION :**

**Graphical user interface

Description automatically generated**In ***depth-wise operation***, convolution is applied to a **single channel** at a time unlike standard CNN’s in which it is done for all the M channels. So here the filters/kernels will be of size **Dk x Dk x 1**. Given there are M channels in the input data, then M such filters are required. Output will be of size **Dp x Dp x M**.

Cost of this operation:

* A single convolution operation require **Dk x Dk** multiplications.
* Since the filter are slided by **Dp x Dp** times across all the M channels, the total number of multiplications is equal to **M x Dp x Dp x Dk x Dk**
* So for depth wise convolution operation, **Total no of multiplications = M x Dk2 x Dp2**

**2-POINT WISE CONVOLUTION :**

In **point-wise operation**, a **1×1 convolution** operation is applied on the M channels. So the filter size for this operation will be **1 x 1 x M**. Say we use N such filters, the output size becomes **Dp x Dp x N**.

Diagram

Description automatically generated

Cost of this operation:

* *A single convolution operation require* ***1 x M*** *multiplications.*
* Since the filter is being slided by **Dp x Dp** times, *the total number of multiplications is equal* ***to M x Dp x Dp x (no. of filters)***
* So for point wise convolution operation, ***Total no of multiplications = M x Dp2 x N***

**Therefore, for overall operation:**

* *Total multiplications = Depth wise conv. multiplications + Point wise conv. multiplications*
* *Total multiplications* ***= M \* Dk2 \* Dp2 + M \* Dp2 \* N = M \* Dp2 \* (Dk2 + n)***
* So for depth wise separable convolution operation, *Total no of multiplications* ***= M x Dp2 x (Dk2 + N)***

# Comparison between the complexities of these types of convolution operations

|  |  |
| --- | --- |
| **Type of Convolution** | **Complexity** |
| Standard | N x Dp2 x Dg2 x M |
| Depth wise separable | M x Dp2 x (Dk2 + N) |

***Complexity of depth wise separable convolutions***

***-------------------------------------------------- = RATIO ( R )***

***Complexity of standard convolution***

***Ratio(R) = 1/N + 1/Dk2***

As an example, consider N = 100 and Dk = 512. Then the ratio R = 0.010004

This means that the depth wise separable convolution network, in this example, performs 100 times lesser multiplications as compared to a standard constitutional neural network.

**This implies that we can deploy faster convolution neural network models without losing much of the accuracy.**