MobileNet:

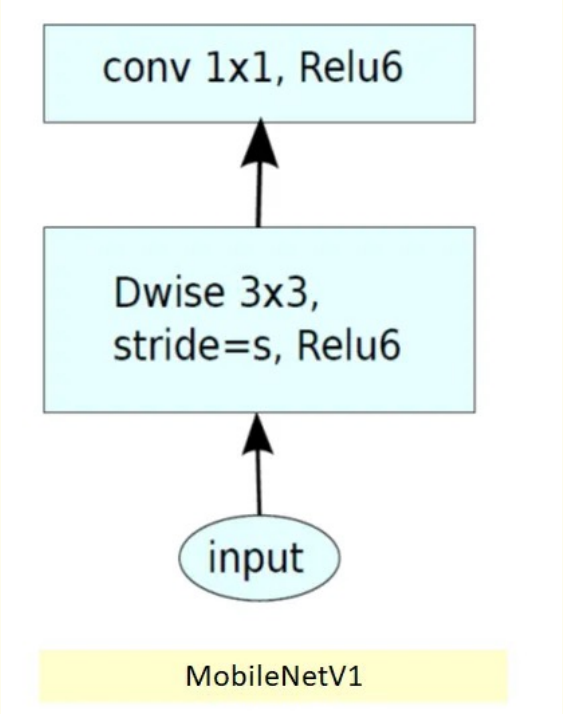
→ Depthwise separable convolution is introduced in v1 that reduces the complexity cost and model size drastically.

→ these models are suitable for mobile devices with low computational power.

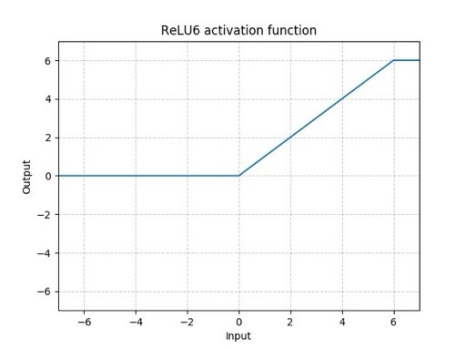
V1:

→ There are two layers:

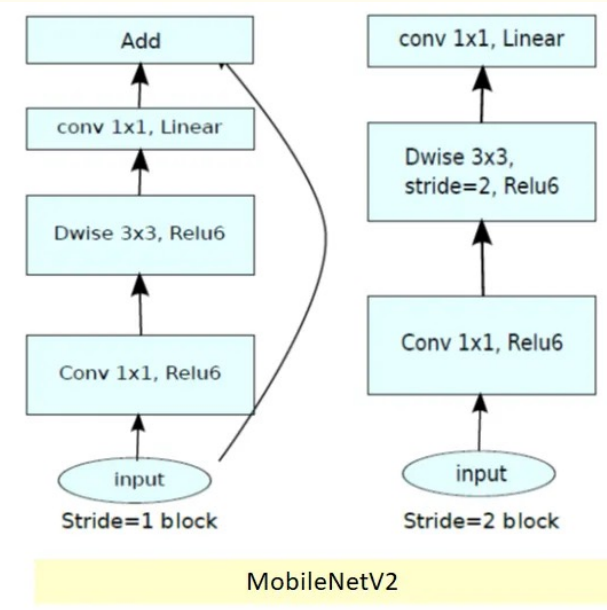
1. **Depthwise Conv:**
   it performs lightweight filtering by applying single convolution filter per input chanel.
2. **Pointwise convolution:**
   (1x1) conv layer responsi-ble for building new feature map through computing linear combinations of the input channels.



→ ReLU6 is used here for its robust-ness with low precision computation.



V2 :



→ In V2 there are 2 types of blocks:

→ residual block with stride = 1

→ residual block with stride = 2

→ first layer:
(1x1) conv with ReLU6 [pointwise]

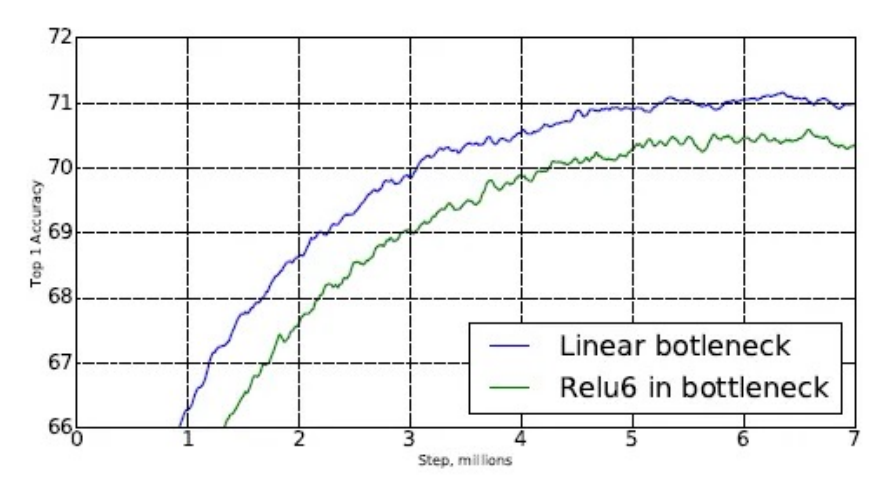
→ second layer:
(3x3) depthwise conv with ReLU6

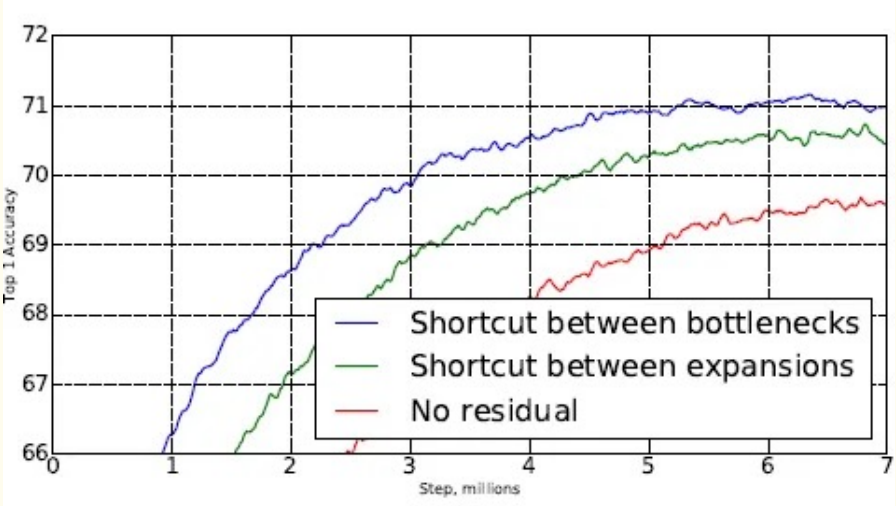
→ third layer:
(1x1) pointwise conv without non-linearity (no ReLU6)

Input	Operator	Output
$h \times w \times k$	1x1 conv2d , ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwise s=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

→ t = expansion factor

Ex: if t=6, internal output = (6×64) = 384 channels.





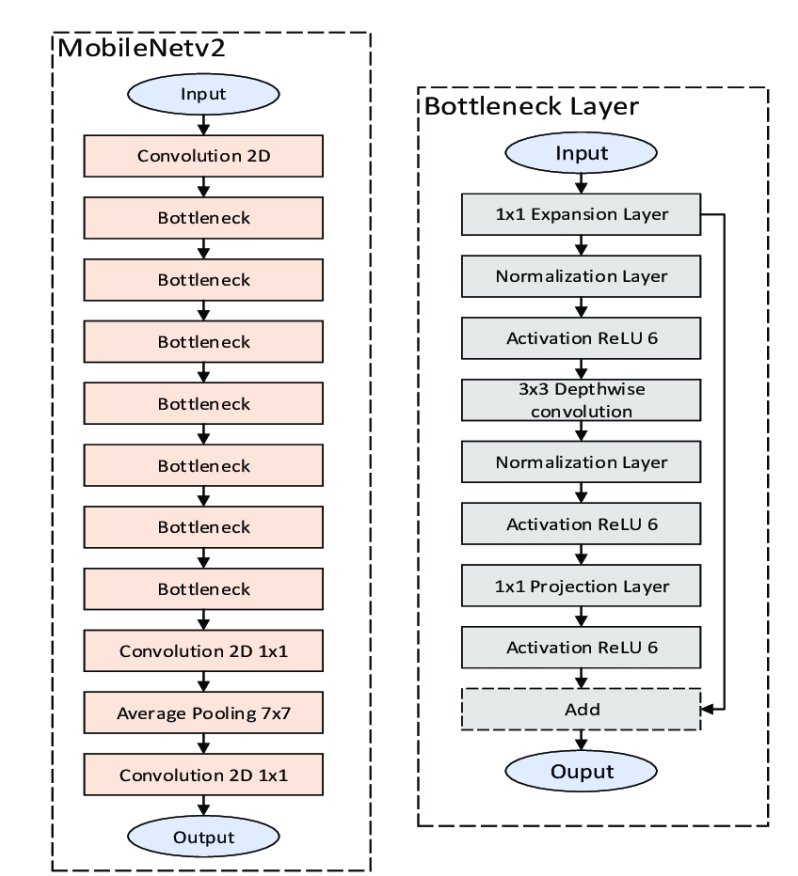
→ produces better results with shortcuts b/w bottlenecks.

- In typical, the **primary network** (width multiplier 1, 224×224), has a computational cost of **300 million multiply-adds** and uses **3.4 million parameters**. (Width multiplier is introduced in MobileNetV1.)
- The performance trade offs are further explored, for **input resolutions from 96 to 224**, and **width multipliers of 0.35 to 1.4.**
- The network computational cost up to 585M MAdds, while the model size vary between 1.7M and 6.9M parameters.
- To train the network, 16 GPU is used with batch size of 96.

Size	MobileNetV1	MobileNetV2	ShuffleNet (2x,g=3)
112x112	64/1600	16/400	32/800
56x56	128/800	32/200	48/300
28x28	256/400	64/100	400/600K
14x14	512/200	160/62	800/310
7x7	1024/199	320/32	1600/156
1x1	1024/2	1280/2	1600/3
max	1600K	**400K**	600K

Number of Maximum Channels/Memory in Kb) at Each Spatial Resolution for Different Architecture with 16-bit floats for activation

Overall Structure:



Classification results:

Network	Top 1	Params	MAdds	CPU
MobileNetV1	70.6	4.2M	575M	113ms
ShuffleNet (1.5)	71.5	**3.4M**	292M	-
ShuffleNet (x2)	73.7	5.4M	524M	-
NasNet-A	74.0	5.3M	564M	183ms
MobileNetV2	**72.0**	**3.4M**	**300M**	**75ms**
MobileNetV2 (1.4)	**74.7**	6.9M	585M	**143ms**