**Depthwise separable convolution**

**[출처]** [depthwise convolution / pointwise convolution](https://blog.naver.com/worb1605/221386398035)|**작성자** [worb1605](https://blog.naver.com/worb1605)

사람의 몸을 detecting하는 Open Pose 라는 논문을 제작하면서, VGG-19 network의 연산량이 매우 많아 실시간으로 동작하지 못하는 한계점을 느끼고 이를 개선하고자 여러 논문을 찾아보았다. 그러던 와중 찾은 모델이 “MobileNet”이다.

​

MobileNet은 depthwise convolution을 택하여 연산량을 기하급수로 줄여 real time으로 동작 가능하도록 하게 해준 network이다.

​

그렇다면 depthwise convolution은 어떻게 연산량을 기하급수로 줄일 수 있었을까?

​

이에 대해 알아보다 보니, 연산량을 줄이는 convolution들에 대해 공부하게 되었고, depthwise convolution뿐만 아니라 Pointwise convolution 등 여러 convolution에 대해 알아보게 되었다.

​

우선 depthwise convolution에 대해 알아보도록 하자.

**1. Depthwise convolution**

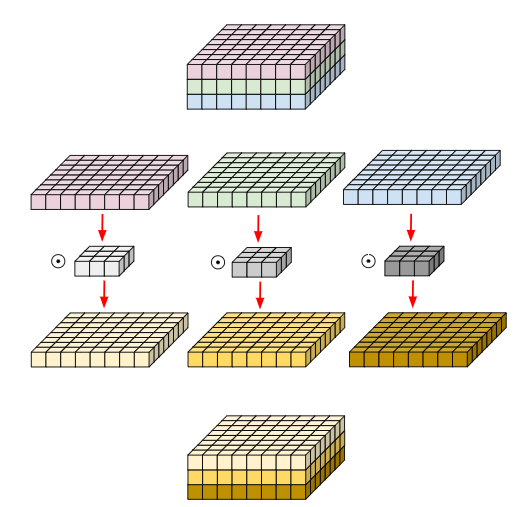
depthwise convolution을 일반 convolution과 비교한 그림이 하나 있다. 이 그림을 보면 단번에 이해가 된다.

위 이미지를 보면

8x8x3의 matrix을 depthwise convolution 하기 위해서 3x3x3 의 커널을 사용하였다.

이때 3x3x3의 커널은 각 채널별로 분리되어 3x3의 2차원 커널을 8x8 크기의 각 분리된 matrix에 붙어 각각 convolution을 진행하고 다시 합쳐지는 구조로 되어있다.

쉽게 말하면 채널 방향의 convolution은 하지 않고, 공간 방향의 convolution만을 진행한다는 것이다.

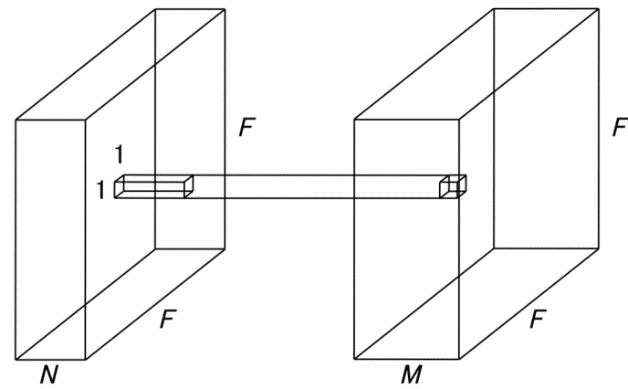
[](https://blog.naver.com/PostView.nhn?blogId=worb1605&logNo=221386398035&from=search&redirect=Log&widgetTypeCall=true&topReferer=https%3A%2F%2Fsearch.naver.com%2Fsearch.naver%3Fsm%3Dtab_hty.top%26where%3Dnexearch%26query%3Ddepthwise%2Bconvolution%26oquery%3DMobileNetV2%26tqi%3DUqHwkdprvmZssRSEa2Cssssstx8-002737&directAccess=false)

**2. Pointwise convolution**

앞서 말했던 것들은 depthwise convolution, depthwise separable convolution들이고, 이들은 공간 방향의 convoluton을 진행한 후, channel 방향의 convolution을 진행하는가/진행하지 않는 가로 나뉘었던 결과물이었다. 이와 비슷하지만 다른 Pointwise convolution은 공간 방향의 convolution은 진행하지 않고, 채널 방향의 convolution을 진행하는 것이다.

위 이미지는 pointwise convolution을 보여주는 예 인데, 1x1xc 크기의 커널을 사용하여 input의 특징을 1개의 채널로 압축시키는 효과가 있다.

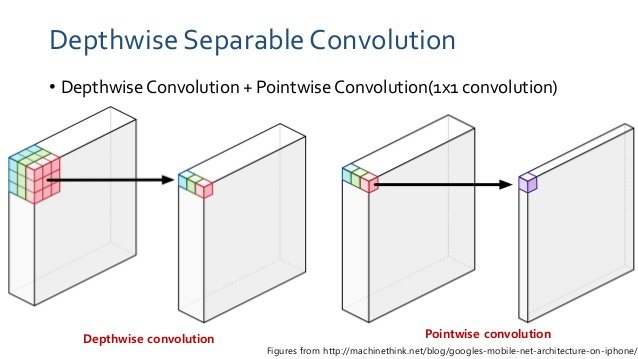
이렇다 보니 input의 채널을 압축시키는 효과가 있어 연산속도를 향상시킬 수 있다는 장점이 있다. 물론 압축되면서 없어지게 되는 데이터도 있지만, 이는 연산 속도와 trade-off 관계에 있는 것이라 그 정도를 어느정도 조정해 나가는 과정이 필요할 것처럼 보인다.

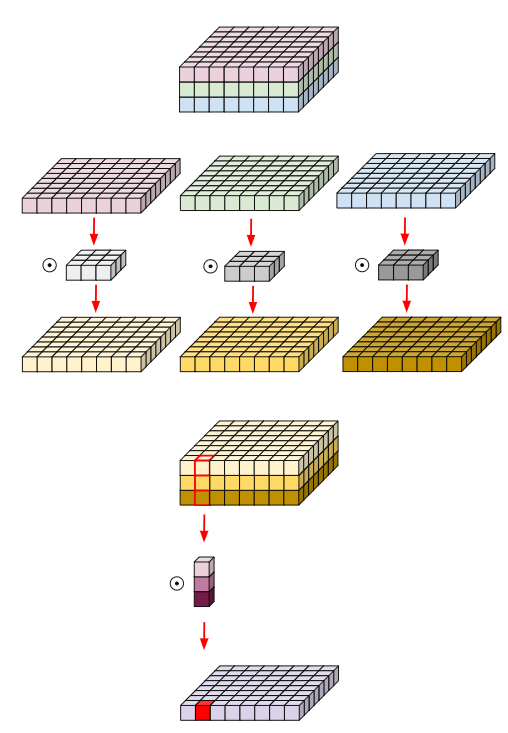
[](https://blog.naver.com/PostView.nhn?blogId=worb1605&logNo=221386398035&from=search&redirect=Log&widgetTypeCall=true&topReferer=https%3A%2F%2Fsearch.naver.com%2Fsearch.naver%3Fsm%3Dtab_hty.top%26where%3Dnexearch%26query%3Ddepthwise%2Bconvolution%26oquery%3DMobileNetV2%26tqi%3DUqHwkdprvmZssRSEa2Cssssstx8-002737&directAccess=false)

**3. Depthwise separable convolution**

이와 반대로 Depthwise separable convolution은 아래와 같은 구조로 되어있으며 각 channel output이 하나로 합쳐지는 특징이 있다.

depthwise convolution을 진행한 결과물에 각 채널을 1개의 채널로 압축할 수 있는 추가 convoluton을 진행하여 결과물이 매우 간소하게 나오도록 하였다.



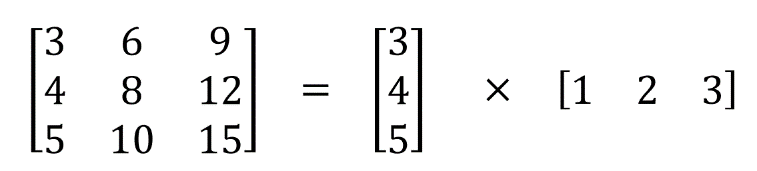
[](https://blog.naver.com/PostView.nhn?blogId=worb1605&logNo=221386398035&from=search&redirect=Log&widgetTypeCall=true&topReferer=https%3A%2F%2Fsearch.naver.com%2Fsearch.naver%3Fsm%3Dtab_hty.top%26where%3Dnexearch%26query%3Ddepthwise%2Bconvolution%26oquery%3DMobileNetV2%26tqi%3DUqHwkdprvmZssRSEa2Cssssstx8-002737&directAccess=false)

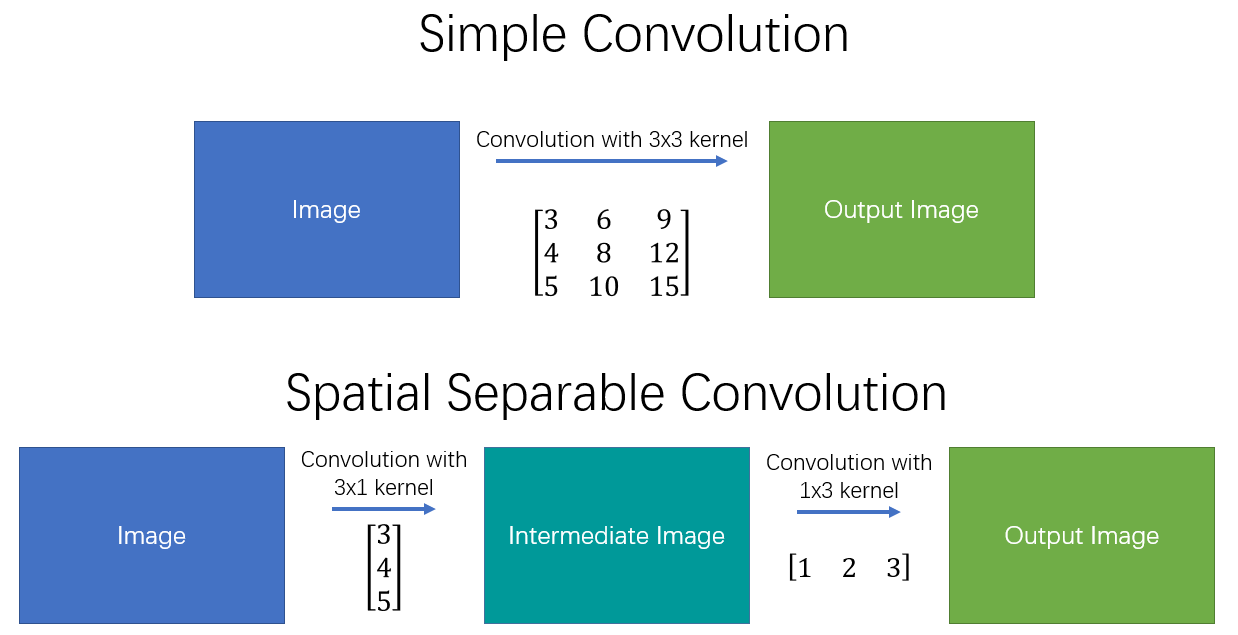
**4. Separable Convolutions (분리 가능형 컨볼루션)**

분리 가능형 컨볼루션에는 공간 분리형 컨볼루션(**Spatial Separable Convolutions**)과 깊이 분리형 컨볼루션(**Depthwise Separable Convolutions**)의 2가지 유형이 있습니다.

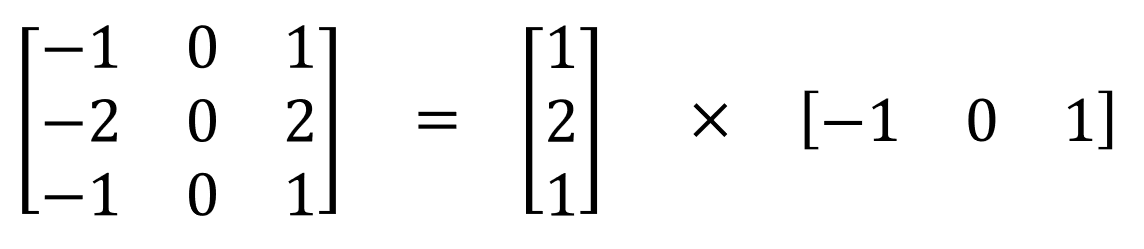
**Spatial Separable Convolutions**

공간 분리형은 이미지와 커널의 공간 차원 , 즉 너비와 높이를 주로 다루기 때문에 그렇게 명명 됩니다. 공간 분리형은 단순히 커널을 두 개의 작은 커널로 나눕니다. 가장 일반적인 경우는 3x3 커널을 3x1 및 1x3 커널로 나누는 것입니다.





9 곱셈으로 한 번의 컨볼루션을 수행하는 대신 동일한 효과를 얻기 위해 각각 3 번의 곱셈 (총 6 번)으로 두 번의 컨볼루션을 수행합니다. 곱셈이 적을수록 계산 복잡성이 줄어들고 네트워크를 더 빠르게 실행할 수 있습니다.



[ Sobel kernel ]

공간적으로 분리 될 수 있는 가장 유명한 회선 중 하나는 모서리를 감지하는 데 사용되는 Sobel 커널입니다. 공간 분리형의 주요 문제는 모든 커널을 두 개의 작은 커널로 분리할 수 없다는 것입니다. 네트워크가 채택 할 수 있는 모든 가능한 커널이 있기 때문에 훈련 중에는 특히 귀찮게 됩니다. 두 개의 작은 커널로 분리 할 수 있는 작은 부분 중 하나만 사용할 수 있습니다.

**Depthwise Separable Convolutions**

공간 분리형 컨벌루션과 달리 깊이 분리형 컨벌루션은 두 개의 작은 커널로“factored=인수”할 수 없는 커널에서 작동합니다. 따라서 더 일반적으로 사용됩니다. 이것은 keras.layers.SeparableConv2D 또는 tf.layers.separable\_conv2d에서 볼 수 있는 분리 가능한 컨볼루션 유형입니다.

깊이 분리 가능한 컨벌루션은 공간 차원뿐만 아니라 깊이 차원 (채널 수)도 다루기 때문에 그렇게 명명되었습니다. 입력 이미지에는 RGB의 3 가지 채널이 있을 수 있습니다. 몇 번의 컨볼 루션 후 이미지에는 여러 채널이 있을 수 있습니다. 각 채널을 해당 이미지의 특정 해석으로 이미지화 할 수 있습니다. 예를 들어, "적색"채널은 각 픽셀의 "적색"을 해석하고 "청색"채널은 각 픽셀의 "청색"을 해석하고 "녹색"채널은 각 픽셀의 "녹색"을 해석합니다. 64 개의 채널이 있는 이미지는 해당 이미지에 대해 64 개의 다른 해석을 갖습니다.

공간 분리형 컨벌루션과 유사하게, 깊이 분리형 컨벌루션은 커널을 depthwise convolution과 pointwise convolution이라는 두 개의 컨볼루션을 수행하는 2개의 개별 커널로 분할합니다.