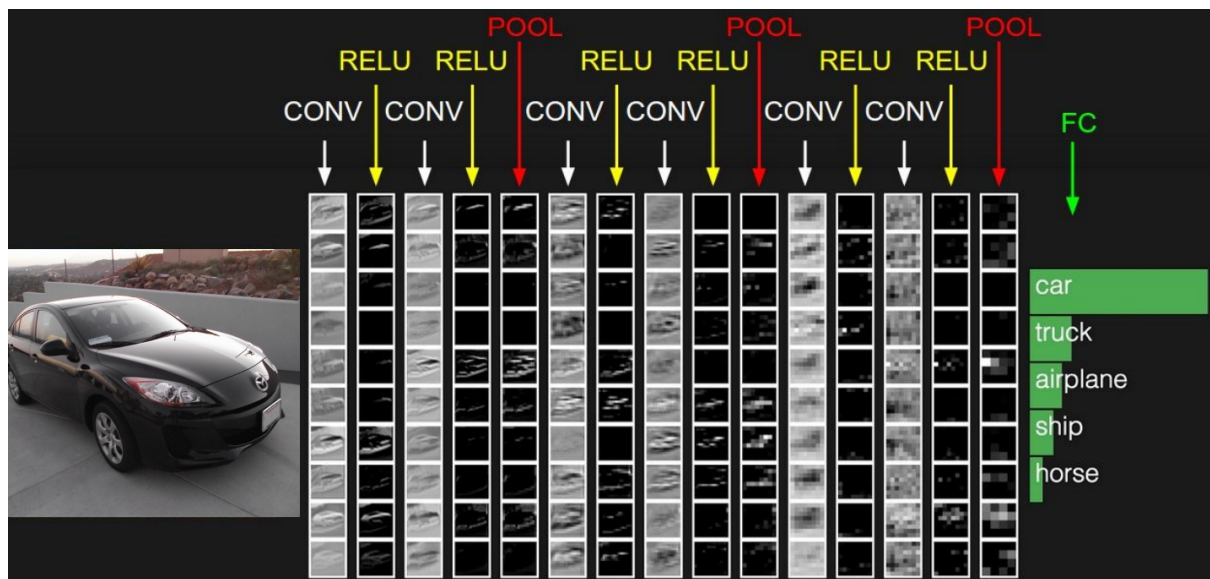


# Lecture 5-2

Title	Convolutional Neural Networks
slide	<a href="http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture5.pdf">http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture5.pdf</a>

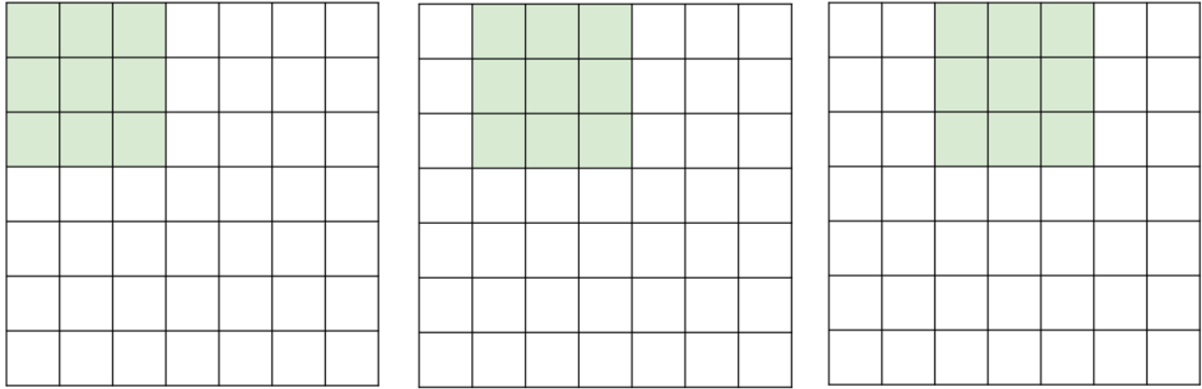
## 정리



정리해보면, CNN의 입력 이미지는 여러 레이어를 통과한다. 가령 첫 번째 Conv Layer 후에 non-linear layer인 ReLU를 통과하여, Conv, ReLU, Conv, ReLU를 거친다. 그 후에 activation maps의 사이즈를 줄이는 역할을 하는 pooling layer를 통과하고 마지막으로 FC-Layer를 사용하여 최종 스코어를 계산한다.

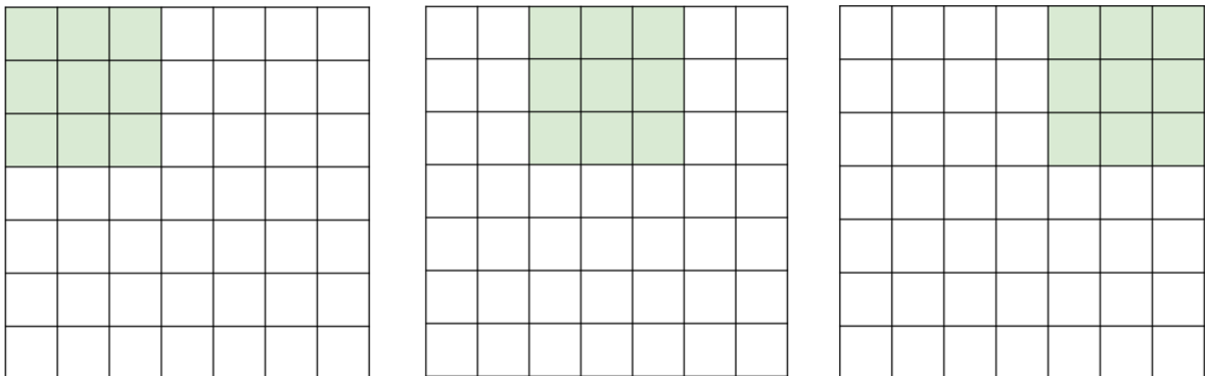
## Spatial dimension

**stride = 1**



간단한 예시로 7x7 입력에 3x3 필터가 있다고 해보자. 필터를 이미지의 왼쪽 상단부터 씌우고 해당 값들의 내적을 수행하면, 이 값들은 activation map의 왼쪽 상단에 위치하게 된다. 다음 단계로 필터를 오른쪽으로 한 칸 움직여서 같은 작업을 수행한다. 이렇게 계속 반복하게 되면 5x5의 출력을 얻을 수 있다. 이 슬라이드 필터는 좌우 방향으로 5번, 상하 방향으로 5번을 수행할 수 있다. 이 예제에서는 슬라이딩을 한 칸씩 진행하는데, 이 때 움직이는 칸을 "stride" 라고 한다. 즉, stride=1을 사용했다.

## stride = 2



stride=2라면, 왼쪽 위부터 시작해서 1칸은 건너뛰고 다음 칸으로 이동해서 계산하게 된다. 따라서 stride=2이면 결국 출력은 3x3이 되게 된다.

마찬가지로 stride=3이면 출력의 사이즈를 생각해보면, stride가 3인 경우에는 이미지를 슬라이딩해도 필터가 모든 이미지를 커버 할 수 없게 된다. 이렇게 되면 불균형한 결과를 야기할 수 있기 때문에 사용하면 안된다.

Output size:  
 **$(N - F) / \text{stride} + 1$**

e.g.  $N = 7, F = 3$ :

stride 1  $\Rightarrow (7 - 3) / 1 + 1 = 5$

stride 2  $\Rightarrow (7 - 3) / 2 + 1 = 3$

stride 3  $\Rightarrow (7 - 3) / 3 + 1 = 2.33$

출력의 사이즈를 계산하는 유용한 수식이 있다. 입력의 차원을  $N$ , 필터 사이즈를  $F$ 라 했을 때, stride가 주어지게 되면, 출력의 크기 **Output size =  $(N - F) / \text{stride} + 1$** 로 나타낼 수 있다. 이를 이용하여 어떤 필터 크기를 사용해야 하는 지를 알 수 있으며, 어떤 stride를 사용했을 때 이미지에 꼭 맞는 지, 그리고 몇 개의 출력 값을 낼 수 있는지도 알 수 있다. 위의 예제로 생각해보면,  $N=7, F=3, \text{stride}=1$ 이고 이것을 수식에 적용해보면, 5x5 출력이 나올 것이라는 것을 알 수 있다. 마찬가지로  $\text{stride}=3$ 이라면,  $\text{output size}=2.33$ 이기 때문에 잘 동작하지 않게 될 것을 알 수 있다.

**zero-pad**

0	0	0	0	0	0			
0								
0								
0								
0								

출력의 사이즈 의도대로 만들어 주기 위해 가장 흔히 사용하는 기법은 zero-pad이다. 이것은 코너를 어떻게 처리 하는지와도 연결된다. 여러 레이어를 거치면 출력의 사이즈는 입력의 사이즈보다 작아지기 때문에, zero-padding을 함으로써 레이어를 거쳐가며 작아지는 출력 사이즈를 유지 시켜주고, 이미지의 가장자리에 0을 채워 넣으면 필터의 중앙이 닿지 않는 곳에도 연산을 할 수 있게 해준다.

그렇다면 7x7 입력에 3x3 필터 연산을 수행할 때, zero-padding을 하면 출력이 어떻게 될까? 앞선 수식을 통해,  $N=9$ ,  $F=3$ ,  $\text{stride}=1$ 로 대입하면  $(9 - 3) / 1 + 1 = 6$ 으로 6x6의 출력을 얻게 된다. activation map의 depth는 우리가 가진 필터의 개수이기 때문에, 정확하게 말해서는 6x6x"필터의 개수"를 출력하게 된다.