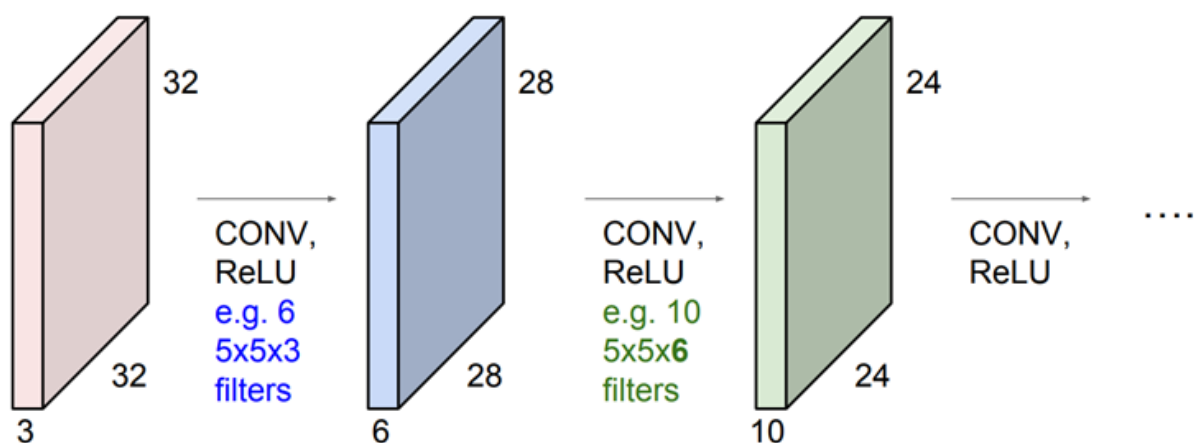


Lecture 5-3

Title	Convolutional Neural Networks
slide	http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture5.pdf

어떤 stride와 filter를 쓸 것인지에 대해 일반적으로 정하는 방법이 있다. 일반적으로 filter는 3x3, 5x5, 7x7을 사용하고, 3x3 filter에는 stride를 1로, 5x5 filter에는 stride를 2로, 7x5 filter에는 stride를 3으로 사용한다. 다시 한 번 말하지만 zero-padding을 하는 목적은 입력 사이즈와 출력 사이즈를 같도록 하기 위해서이다. 레이어가 여러 겹 쌓이고 zero-padding을 하지 않는다면, 출력 사이즈는 아주 빠르게 줄어들게된다. 따라서 만약 엄청 깊은 네트워크가 있다고 생각해보면 각 코너에 있는 값들을 계산하지 못하기 때문에 Activation Map은 점점 줄어들어 엄청 작아지게 될 것이다. 이렇게 되면 일부 정보를 잃게 되고, 원본 이미지를 표현하기에 너무 작은 값을 사용하게 된다.



입력 이미지 32 x 32 x 3이 있고 10개의 5 x 5 필터가 있을 때, stride=1, padding=2로 했을 때 출력 사이즈는 얼마일까? 이전에 배웠던 공식에 적용해보면, 입력 사이즈 F=32인데, padding으로 이 값을 2씩 증가시켜서 32+4가 된다. 따라서 $((32+4)-5)/1+1=32$ 가 되어 각 필터는 32x32가 되고, 10개의 필터의 개수가 있으므로 전체 출력의 크기는 32x32x10 이 된다.

그렇다면 이 레이어의 파라미터는 총 몇 개일까? 10개의 5x5 필터가 있고, 입력의 Depth만큼도 통과해야 하며, bias term 또한 존재한다. 따라서 5 x 5 x 3 가중치에 하나의 bias term이 들어있으므로 필터당 76개의 파라미터($5*5*3+1$)가 있고, 총 10개의 필터가 있으므로 760개의 파라미터가 존재한다.

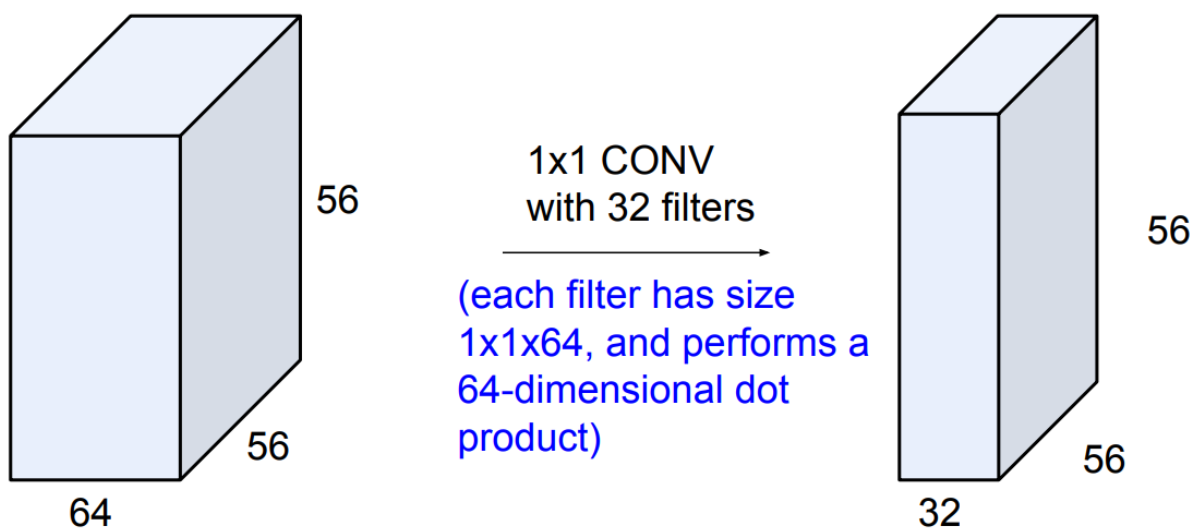
summary

Summary. To summarize, the Conv Layer:

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
 - Number of filters K ,
 - their spatial extent F ,
 - the stride S ,
 - the amount of zero padding P .
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
 - $W_2 = (W_1 - F + 2P)/S + 1$
 - $H_2 = (H_1 - F + 2P)/S + 1$ (i.e. width and height are computed equally by symmetry)
 - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.
- In the output volume, the d -th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the d -th filter over the input volume with a stride of S , and then offset by d -th bias.

일반적으로 필터 사이즈는 3x3, 5x5를 사용하고, Stride는 보통 1이나 2가 가장 흔하며, padding은 설정에 따라 조금씩 다르다. 그리고 보통 필터의 개수는 32, 64, 128, 512와 같이 2의 제곱 수로 한다.

stride를 크게 선택할수록 출력은 점점 작아지기 때문에 Pooling처럼 다운샘플링하는 효과를 얻으면서도 더 좋은 성능을 낸다. 또한 activation map의 사이즈를 줄이는 것은 추후 모델의 전체 파라미터의 개수에도 영향을 미친다. 가령 Conv Layer의 마지막에는 FC Layer가 있을 수 있는데, FC Layer는 Conv의 출력 모두와 연결되어 있어서 Conv Layer의 출력이 작을수록 FC Layer에서 필요한 파라미터의 수가 더 작게 된다. 즉, 파라미터의 수, 모델의 사이즈, Overfitting 과 같은 것들은 다양한 trade-off 가 있으므로, 이를 잘 고려해서 stride를 선택해야 한다.



1x1 Convolution도 마찬가지로 슬라이딩 하면서 값을 구하지만, 5x5에서처럼 공간적인 정보를 이용하지는 않는다. 하지만 이 필터는 여전히 Depth만큼 연산을 수행하기 때문에 1x1 Conv는 입력의 전체 Depth에 대한 내적을 수행하는 것과 같다. 입력은 56x56x64이고 여기에 32개의 1x1 Conv 수행하면, 56x56x32의 출력 값이 나오게 된다.