

딥러닝 3단계 1주차

- Convolution -

이승목

2020.08.18.

범위

- Convolution이란 (W1L01~06)

합성곱 신경망 (Convolutional Neural Networks)	
	컴퓨터비전 업데이트 : 2020.07.11 ♥ 9
	모서리 감지 예시 업데이트 : 2020.07.11 ♥ 11
	더 많은 모서리 감지 예시 업데이트 : 2020.07.12 ♥ 8
	패딩 (Padding) 업데이트 : 2020.07.12 ♥ 10
	스트라이드 (Stride) 업데이트 : 2020.07.12 ♥ 8
	입체형 이미지에서의 합성곱 업데이트 : 2020.07.12 ♥ 9

컴퓨터 비전의 세 가지 문제

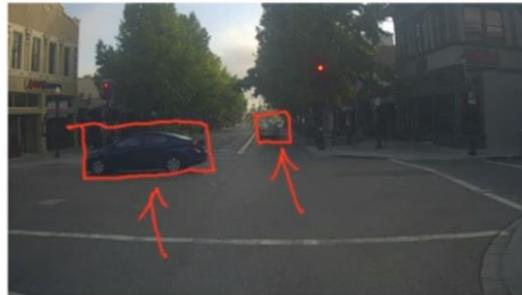
1. 이미지 분류



64x64

→ Cat? (0/1)

2. 사물 인식



3. 스타일 변환



Convolution

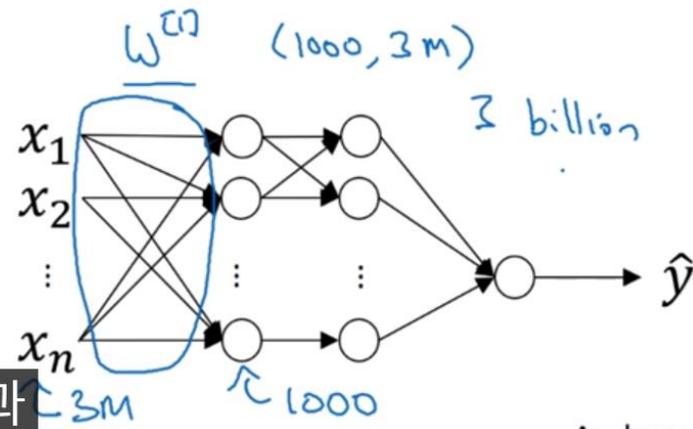
Convolution(합성곱)의 필요성

- 고해상도 인풋에 대해 일반 NN은 너무 무겁다.

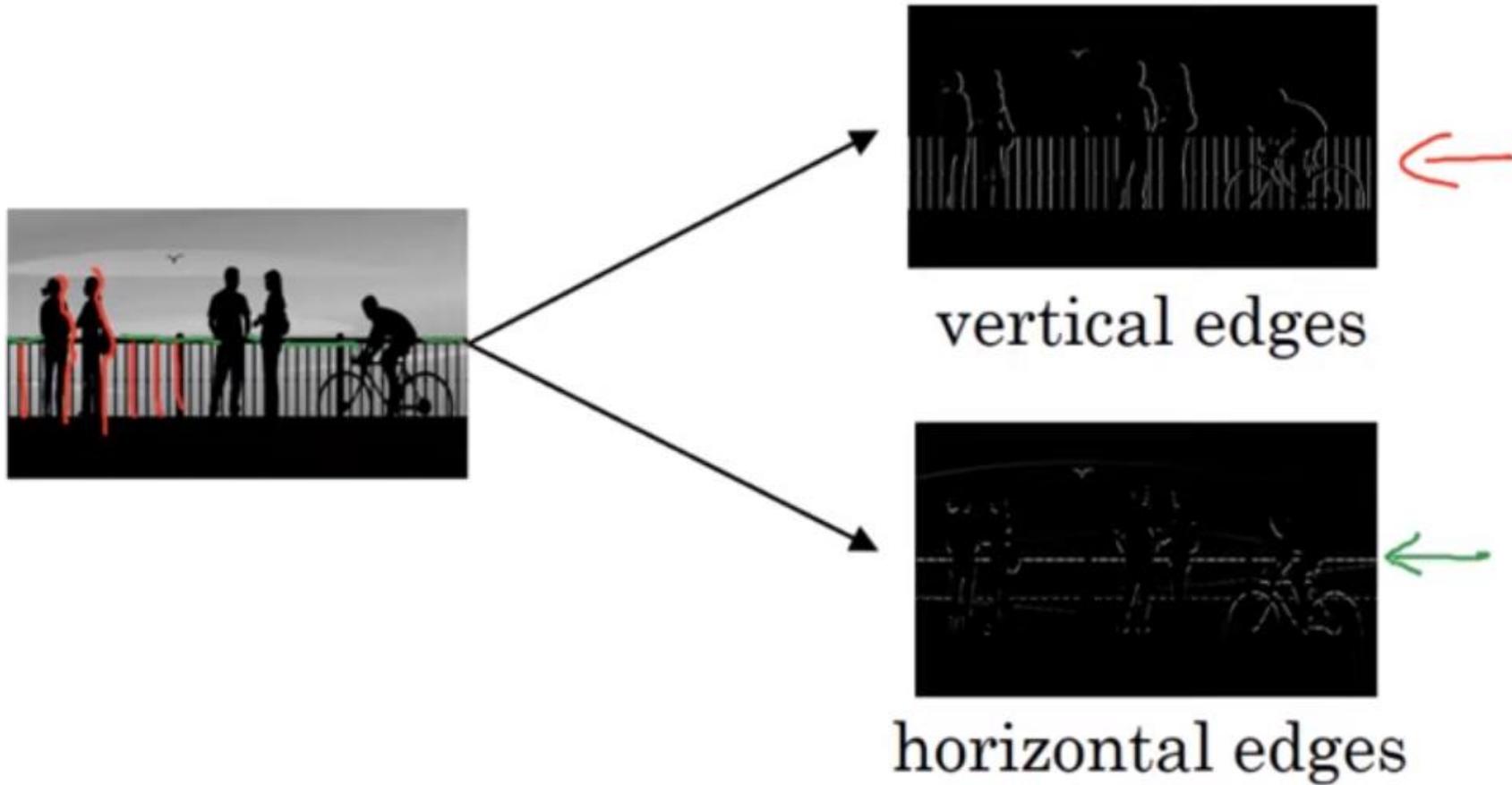


$1000 \times 1000 \times 3$
 $= 3 \text{ million}$

그리고 계산과



Convolution과 모서리 인식

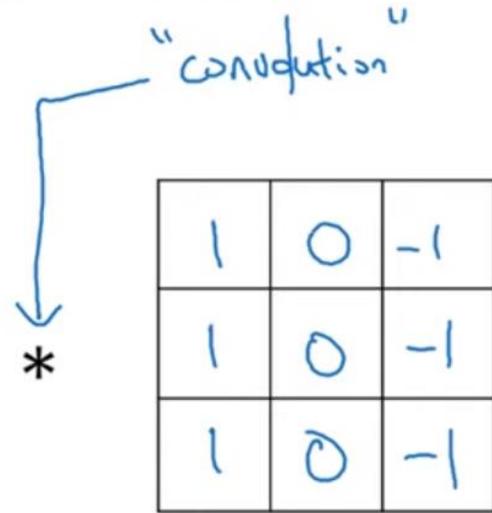


Convolution으로 수직 모서리 인식하기

$$3 \times 1 + 1 \times 1 + 2 \times 1 + 0 \times 0 + 5 \times 0 + 7 \times 0 + 1 \times -1 + 8 \times -1 + 2 \times -1 = -5$$

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

6x6



3x3
filter

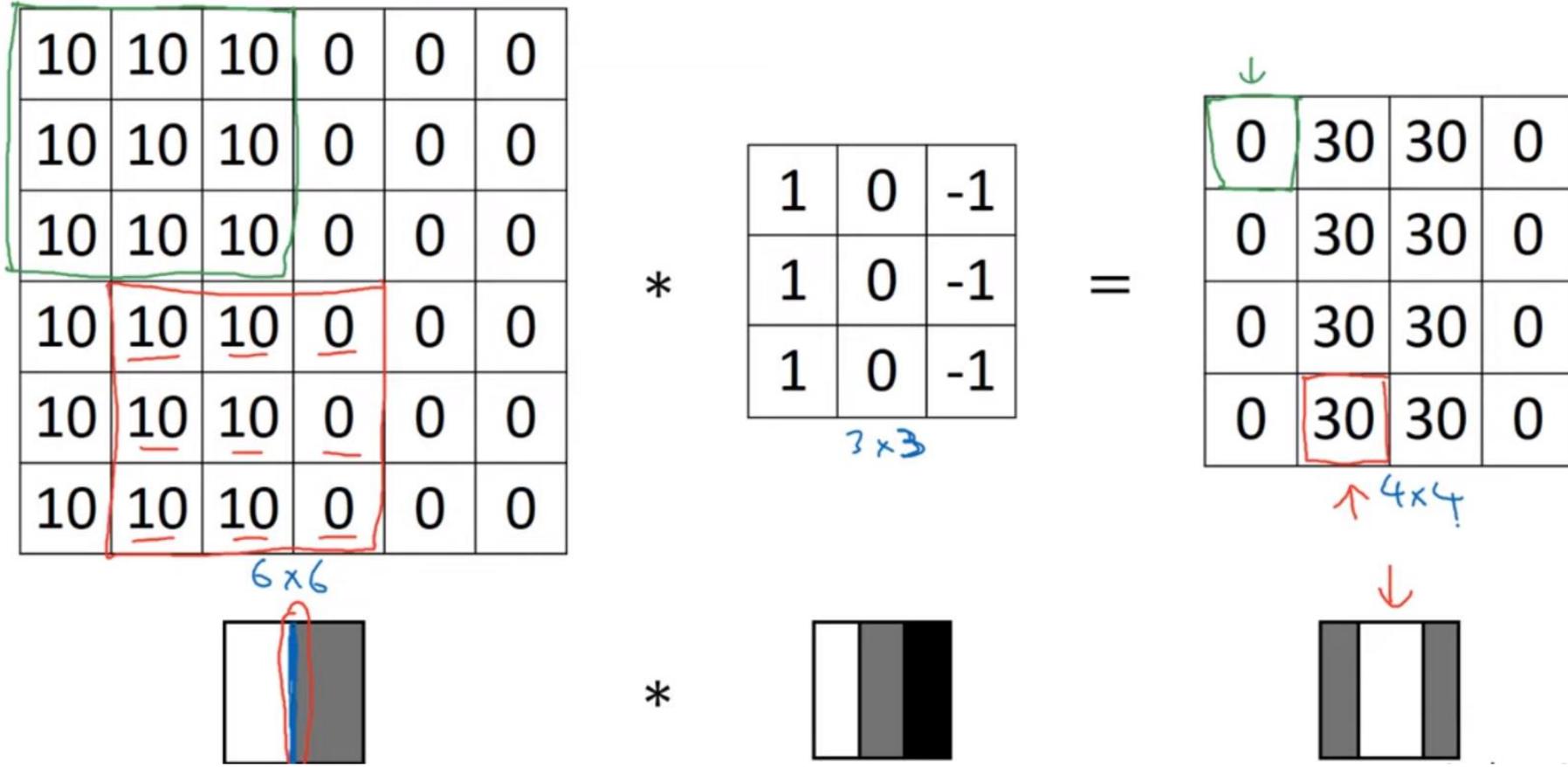
= kernel

=

-5			

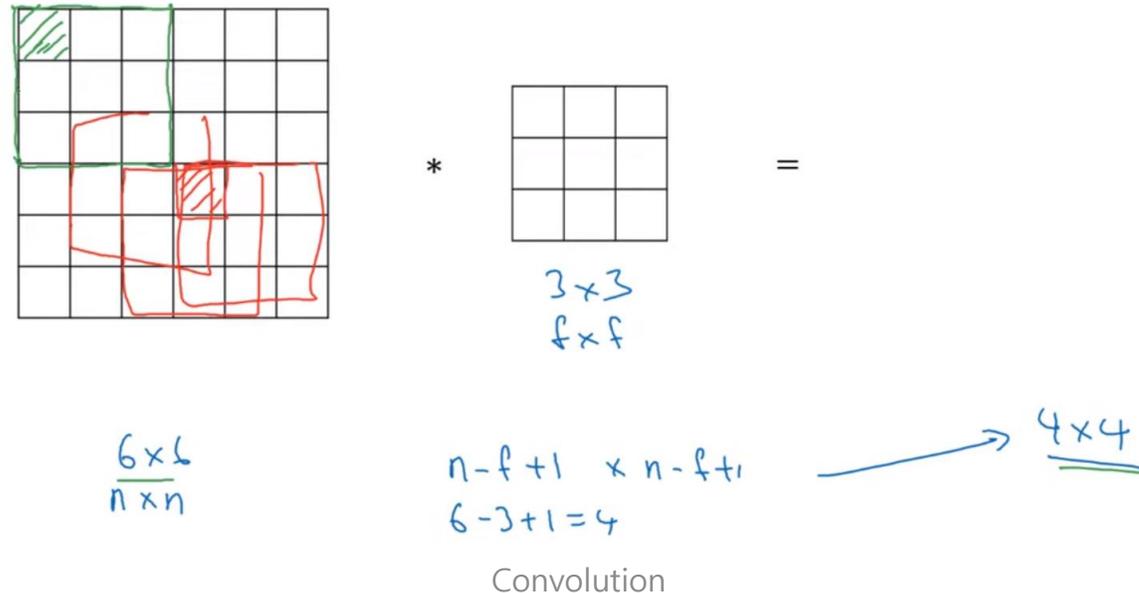
4x4

Convolution으로 수직 모서리 인식하기



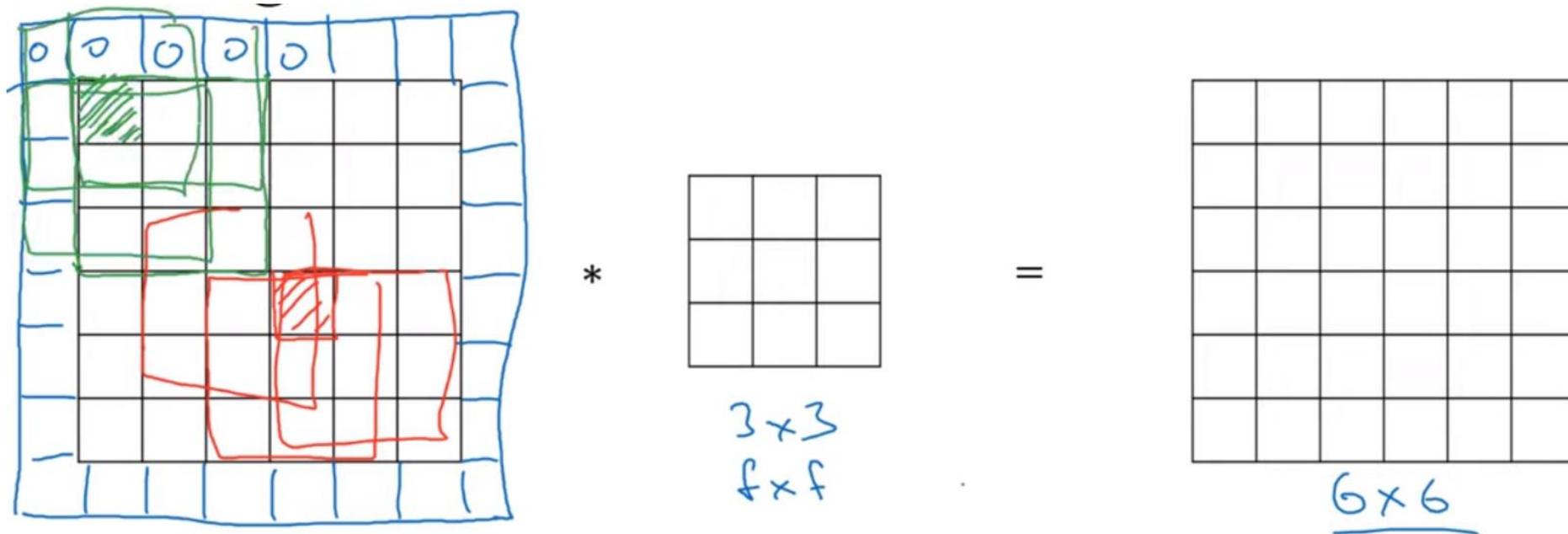
Padding

- 그냥 Convolution을 하면 이미지가 점점 작아진다.
 - $N \times N$ 에 $f \times f$ 필터를 적용하면?
 - $(N-f+1) \times (N-f+1)$ 출력!
- 그냥 Convolution을 하면 가장자리 픽셀은 더 적게 반영된다.



Padding

- Padding을 하면 된다!
 - $N \times N$ 에 p 만큼 padding하고서 $f \times f$ 필터를 적용하면?
 - $(N+2p-f+1) \times (N+2p-f+1)$ 출력!



Valid Convolution과 Same Convolution

- Valid: padding 없이 그냥 하는 것.
- Same: 출력 이미지의 크기가 입력 이미지의 크기와 같아지도록 padding하는 convolution.
 - $N+2p-f+1 = N$
 - $p=(f-1)/2$
 - 의례적으로 f 는 홀수. 그래야 padding이 좌우대칭이 되고, 필터의 중심 픽셀이라는 개념을 사용할 수 있다.

Stride

2	3	7	4	6	2	9
6	6	9	8	7	4	3
3 ³	4 ⁴	8 ⁴	3	8	9	7
7 ¹	8 ⁰	3 ²	6	6	3	4
4 ⁻¹	2 ⁰	1 ³	8	3	4	6
3	2	4	1	9	8	3
0	1	3	9	2	1	4

7x7

*

3	4	4
1	0	2
-1	0	3

3x3

Stride = 2

=

91	100	83
69		

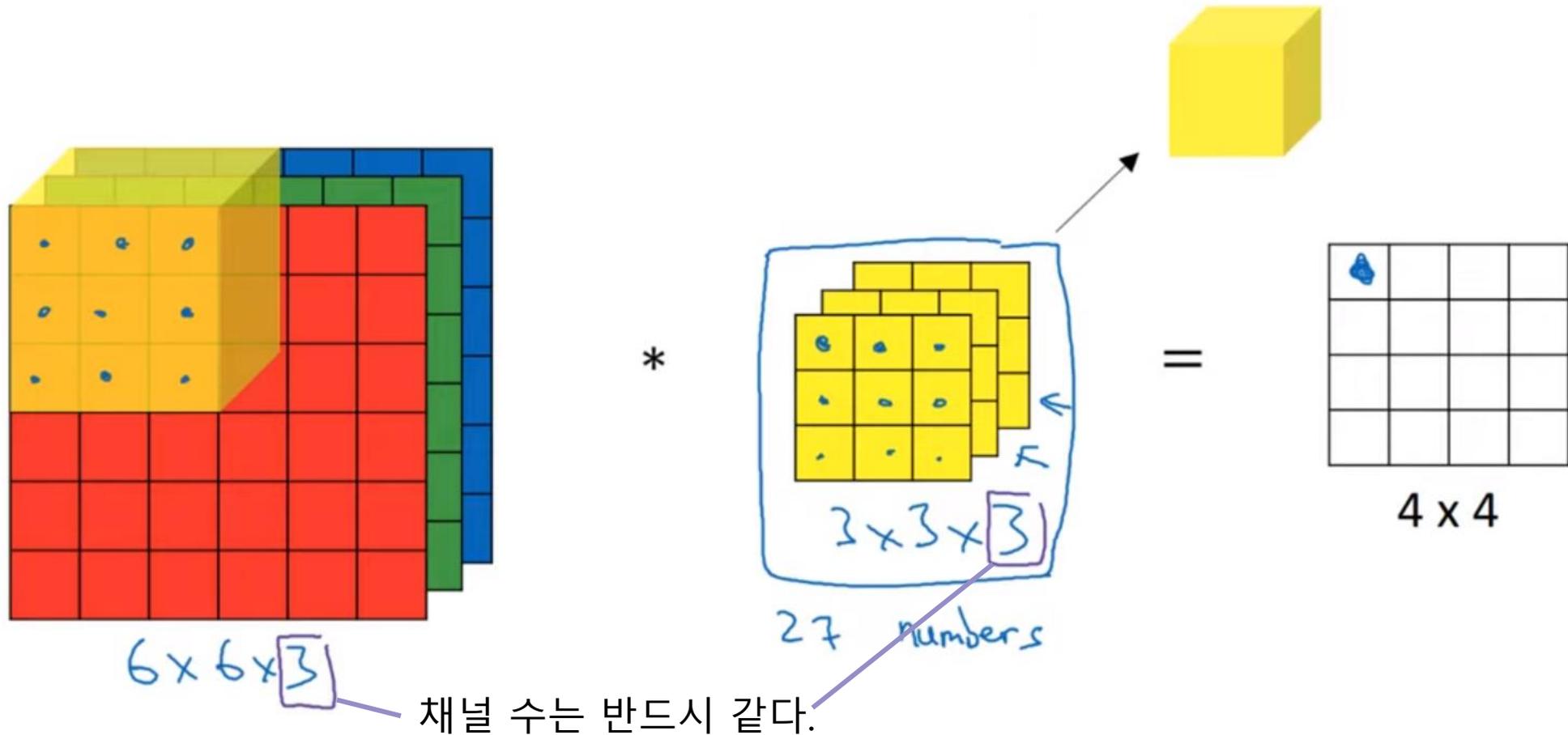
Stride

$n \times n$ image $f \times f$ filter

padding p stride s

$$\left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor$$

채널이 있는 경우의 Convolution

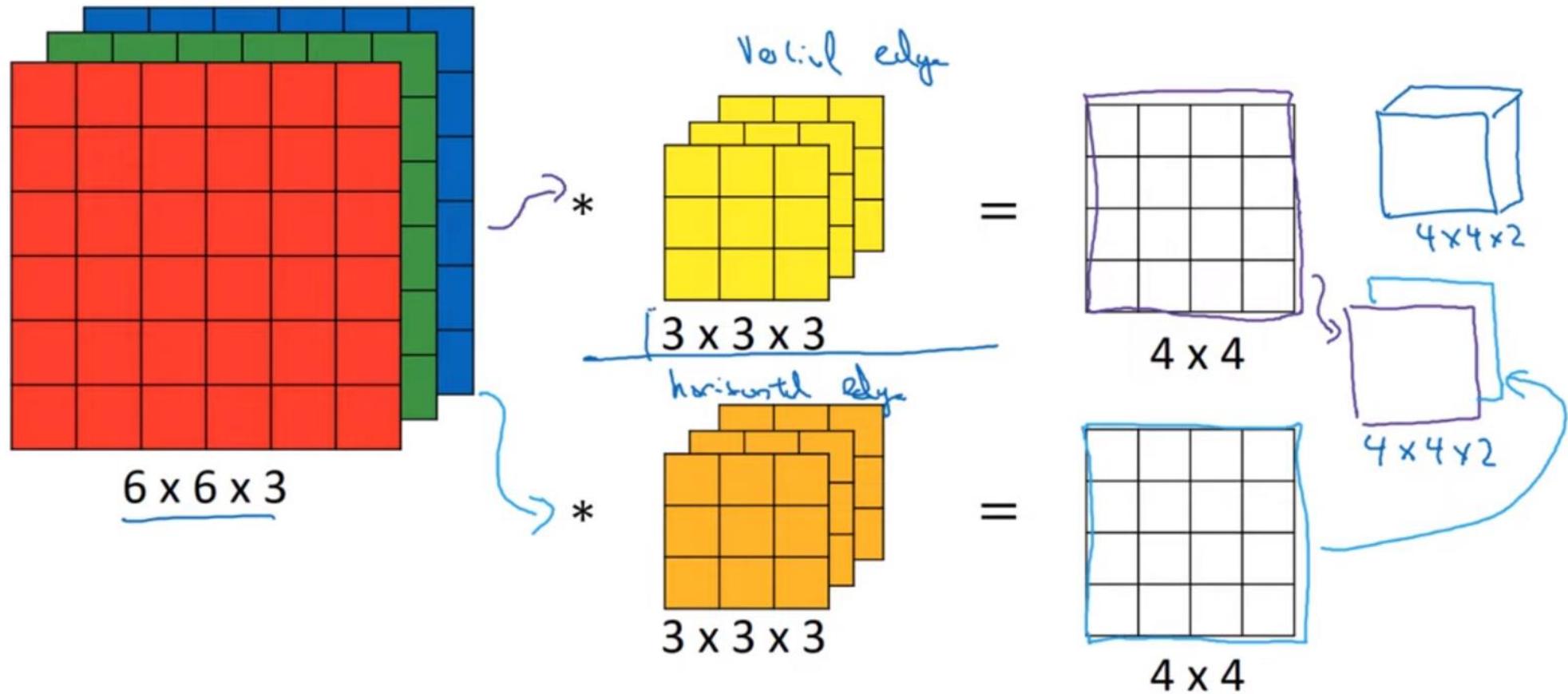


채널이 있는 경우의 Convolution

- 붉은 세로 선을 검출하는 필터



여러 개의 필터



입출력 특성의 크기

- 입력 ($n \times n \times n_c$)
 - (height X width X channels)
 - Channels 대신 depth를 쓰기도 하지만 신경망의 층 수를 의미하는 depth와 헷갈릴 여지가 있어 channels를 사용하길 권장
- 필터 ($f \times f \times n_c$)인 필터를 n_c' 개 적용
 - Parameter 수: $f \times f \times n_c \times n_c'$
- Stride s , Padding p
- 출력 $\left(\left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor \times n_c' \right)$

Convolution vs Cross-Correlation

- 수학에서 말하는 Convolution은 filter의 상하좌우를 반전시킨 뒤 곱해준다.
 - 그렇게 해야 결합법칙이 성립하고 각종 신호 처리에 이점이 있다.
- 하지만 CNN에는 mirroring이 주는 이점이 없기 때문에 그냥 곱한다.
 - 정확한 수학 용어는 cross correlation이지만 컴퓨터 공학을 하는 사람들은 모두 convolution이라고 부르고 있다.

Convolution in math textbook:

2 ⁷	3 ²	7 ⁵	4	6	2
6 ⁹	6 ⁰	9 ⁴	8	7	4
3 ¹	4 ¹	8 ³	3	8	9
7	8	3	6	6	3
4	2	1	8	3	4
3	2	4	1	9	8

*	3	4	5
	1	0	2
	-1	9	7

7	2	5
9	0	4
-1	1	3

전통적인 컴퓨터 비전 연구자

- 필터를 직접 만들고 시험해서 좋은 필터를 일일이 찾아냈다.

1	0	-1
1	0	-1
1	0	-1



1	0	-1
2	0	-2
1	0	-1

Sobel filter

3	0	-3
10	0	-10
3	0	-3

Scharr filter

딥러닝 컴퓨터 비전 연구자

- 이제는 더 좋은 필터를 학습을 통해 찾아낸다.

