

인공지능

컨볼루션 신경망

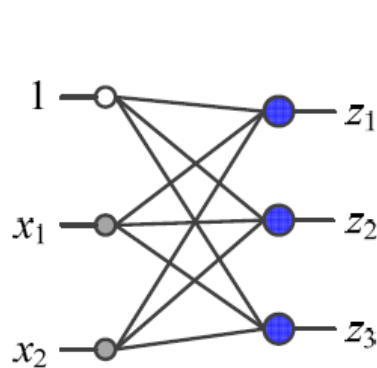
소프트웨어융합대학
소프트웨어학부

본 자료는 해당 수업의 교육 목적으로만 활용될 수 있음.
일부 내용은 다른 교재와 논문으로부터 인용되었으며, 모든 저작권은 원 교재와 논문에 있음.

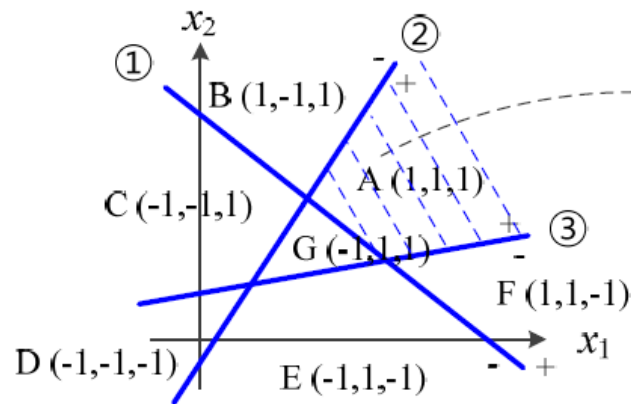
지난 시간...

■ 다층 퍼셉트론의 (은닉)층 의미

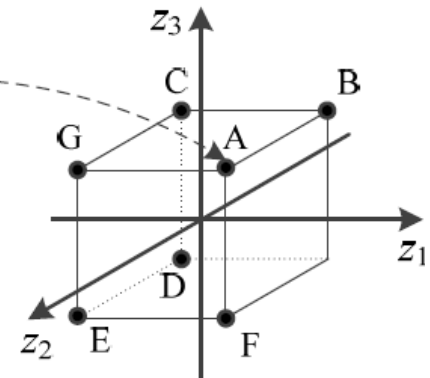
- [그림]처럼 3개 퍼셉트론을 결합하면,
2차원 공간을 7개 영역으로 나누고 각 영역을 3차원 점으로 변환
- ← 활성화함수 τ 로 계단함수를 사용하므로 영역을 점으로 변환
(시그모이드 함수의 경우, 연속적인 새로운 공간으로 변환)



(a) 퍼셉트론 3개를 결합



(b) 7개 부분공간으로 나눔



(c) 3차원 공간의 점으로 매핑

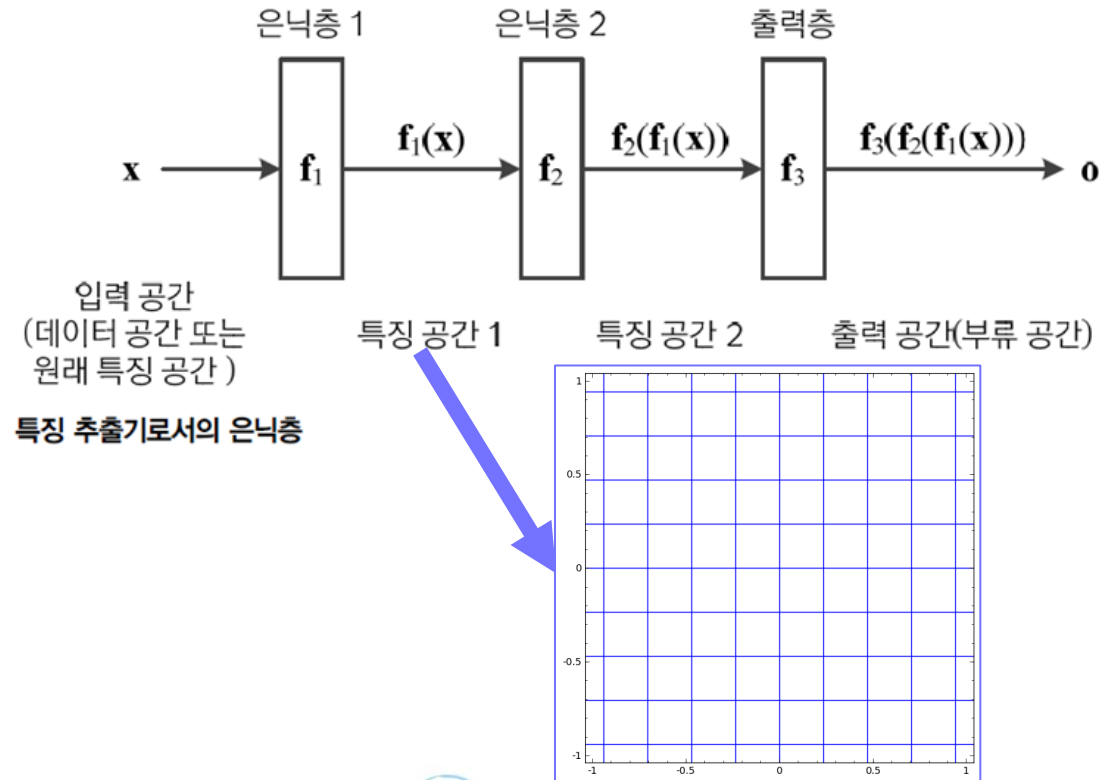
- 일반화하여, p 개 퍼셉트론을 결합하면 p 차원 공간으로 변환

지난 시간...

■ 다층 퍼셉트론의 은닉층은 특징 자동 추출기

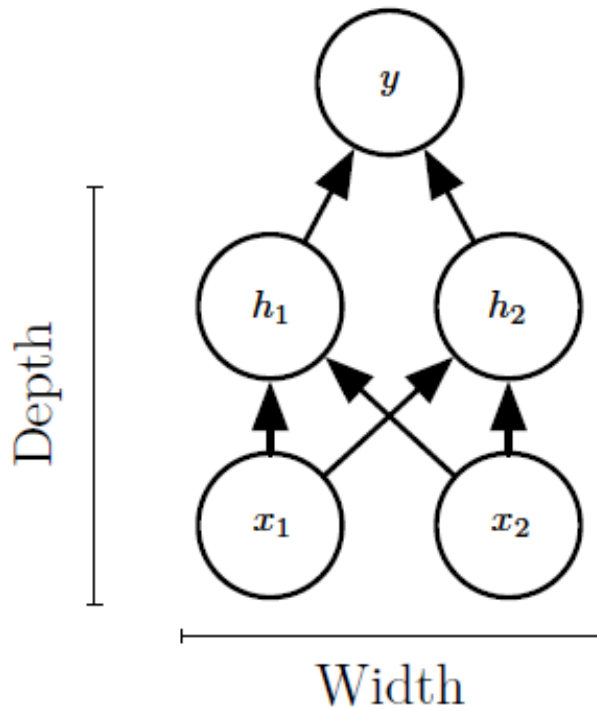
- 은닉층은 특징 벡터를 분류에 더 유리한 새로운 특징 공간으로 변환
- 표현 학습 (representation learning) 을 통해 데이터로부터 자동으로 특징 (data-driven feature) 추출함

(깊은 인공신경망은 더 많은 층을 거쳐 계층적인 표현학습을 함)



지난 시간...

■ 다층 퍼셉트론 기본 구조

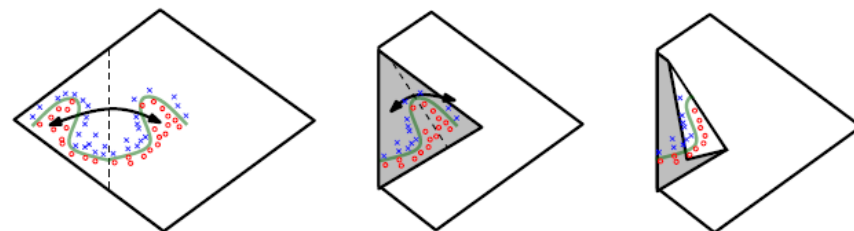
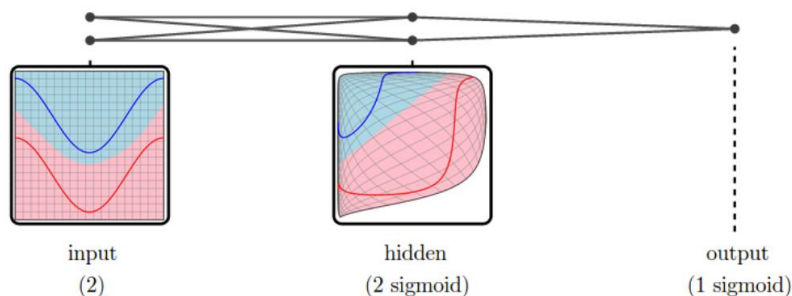


- 범용적 근사 이론 universal approximation theorem
 - 하나의 은닉층은 함수의 근사를 표현
- 다층 퍼셉트론도 공간을 변환하는 함수로 근사됨
- 얇은 은닉층의 구조
 - 더 과잉적합 되기 쉬움
- 일반적으로 깊은 은닉층의 구조가 좋은 성능을 가짐

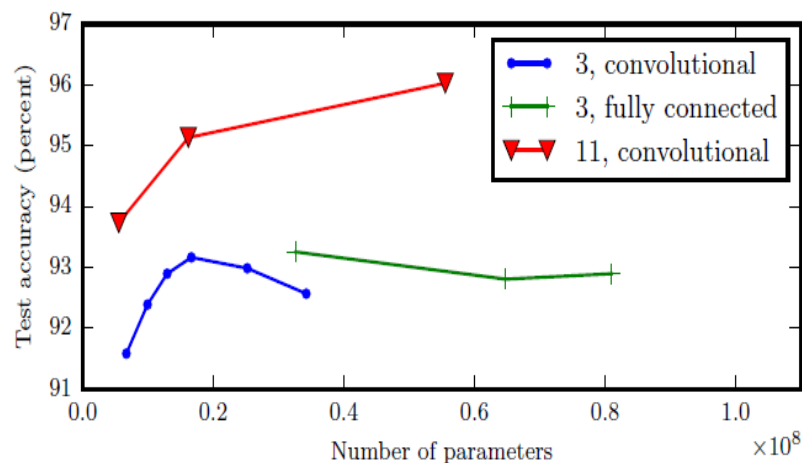
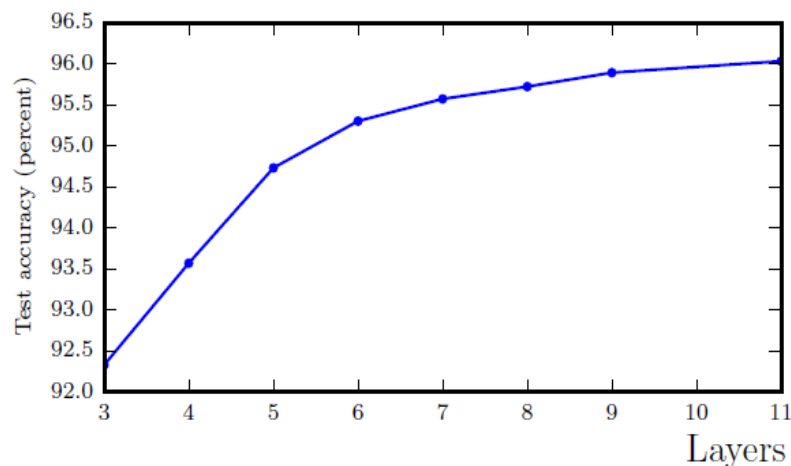
지난 시간...

■ 은닉층의 깊이에 따른 이점

- 공간 변환과 지수의 표현 exponential representation



- 은닉층은 특징 공간 변환 → 지수적으로 많은 선형적인 영역 조각들로 표현
- 여러 은닉층을 통해 계층적으로 특징을 추출, 특징은 층이 깊어짐에 따라 점차 추상화됨
- 좋은 성능



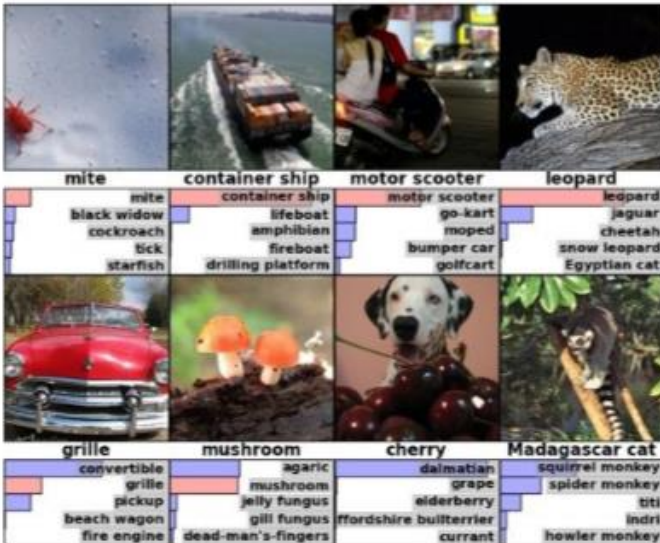
이 장에서는...

- 컨볼루션 인공 신경망의 이해
 - 컨볼루션 인공 신경망의 적용 분야인 컴퓨터 비전 소개
 - 컨볼루션 인공신경망 기초를 이해
 - 이를 심화한 사례와 응용을 설명

컴퓨터 비전

■ 오늘날, 영상 분야에서 다양하게 활용됨

분류 classification



검색 retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

검출 detection



Figures copyright Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, 2015. Reproduced with permission.

분할 segmentation



Figures copyright Clement Farabet, 2012. Reproduced with permission.

[Farabet et al., 2012]

컴퓨터 비전

■ 이외에도

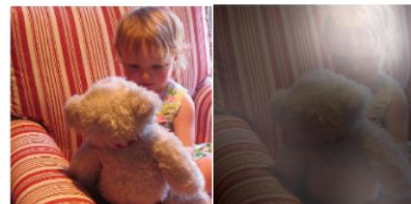
- 동작 인식,
- 영상 캡션 생성



A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A little girl sitting on a bed with a teddy bear.

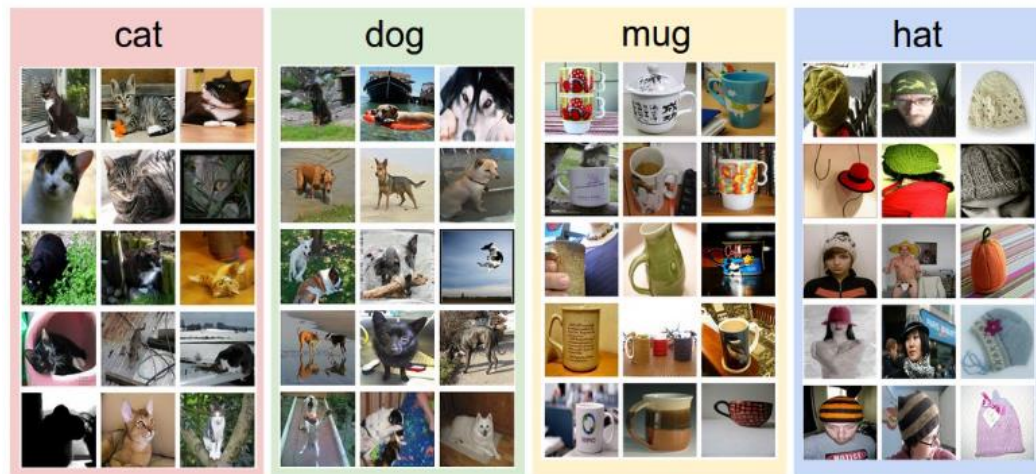


A group of people sitting on a boat in the water.

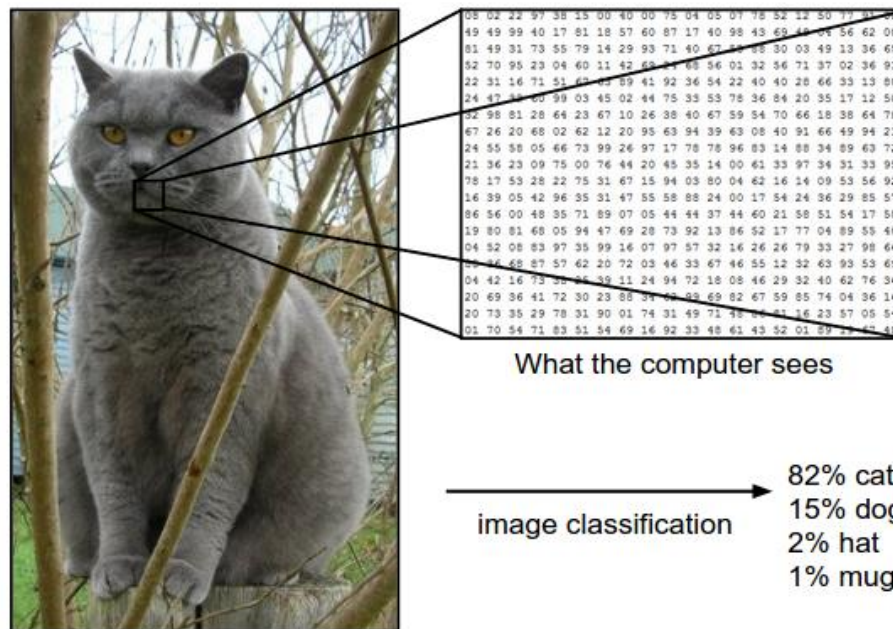
컴퓨터 비전

■ 영상 인식의 예

- 영상마다 분류값이 할당됨
- 수집된 영상으로부터 특징 훈련



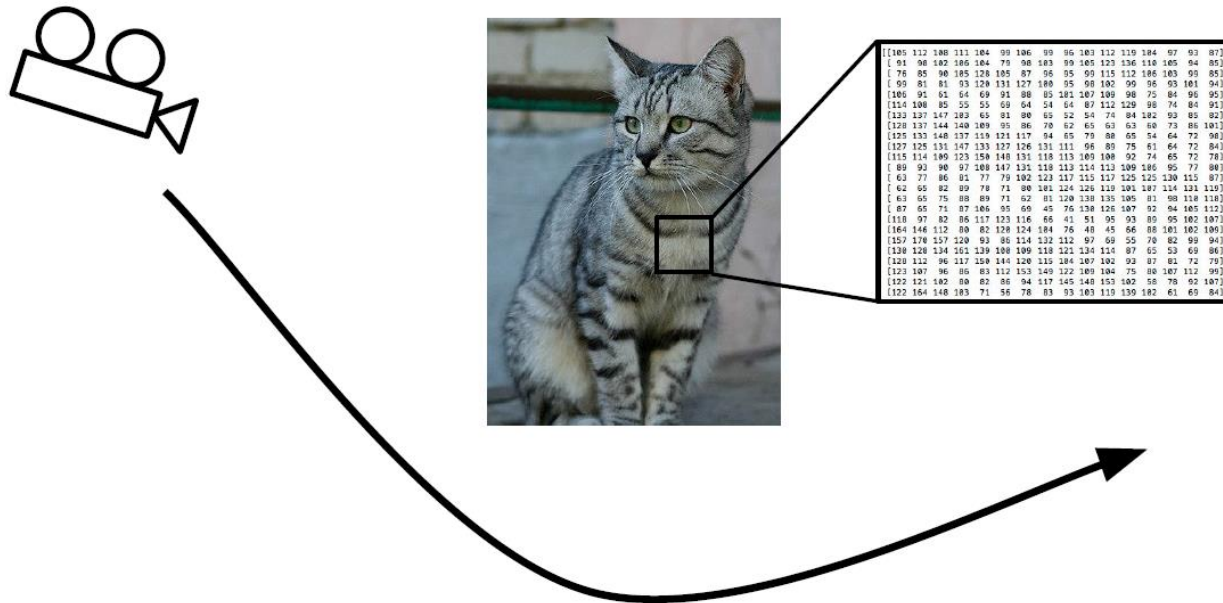
- 새로운 영상의 분류를 추론
 - 영상은 텐서^{tensor} 형태로 표현됨



컴퓨터 비전

■ 컴퓨터 비전의 어려운 점

- 관점의 변화: 동일한 객체라도 영상을 찍는 카메라의 이동에 따라 모든 픽셀값이 변화됨



컴퓨터 비전

■ 컴퓨터 비전의 어려운 점

- 경계색 (보호색)으로 배경과 구분이 어려운 경우



컴퓨터 비전

- 컴퓨터 비전의 어려운 점
 - 조명에 따른 변화



컴퓨터 비전

- 컴퓨터 비전의 어려운 점
 - 기형적인 형태의 영상 존재



컴퓨터 비전

- 컴퓨터 비전의 어려운 점
 - 일부가 가려진 영상 존재



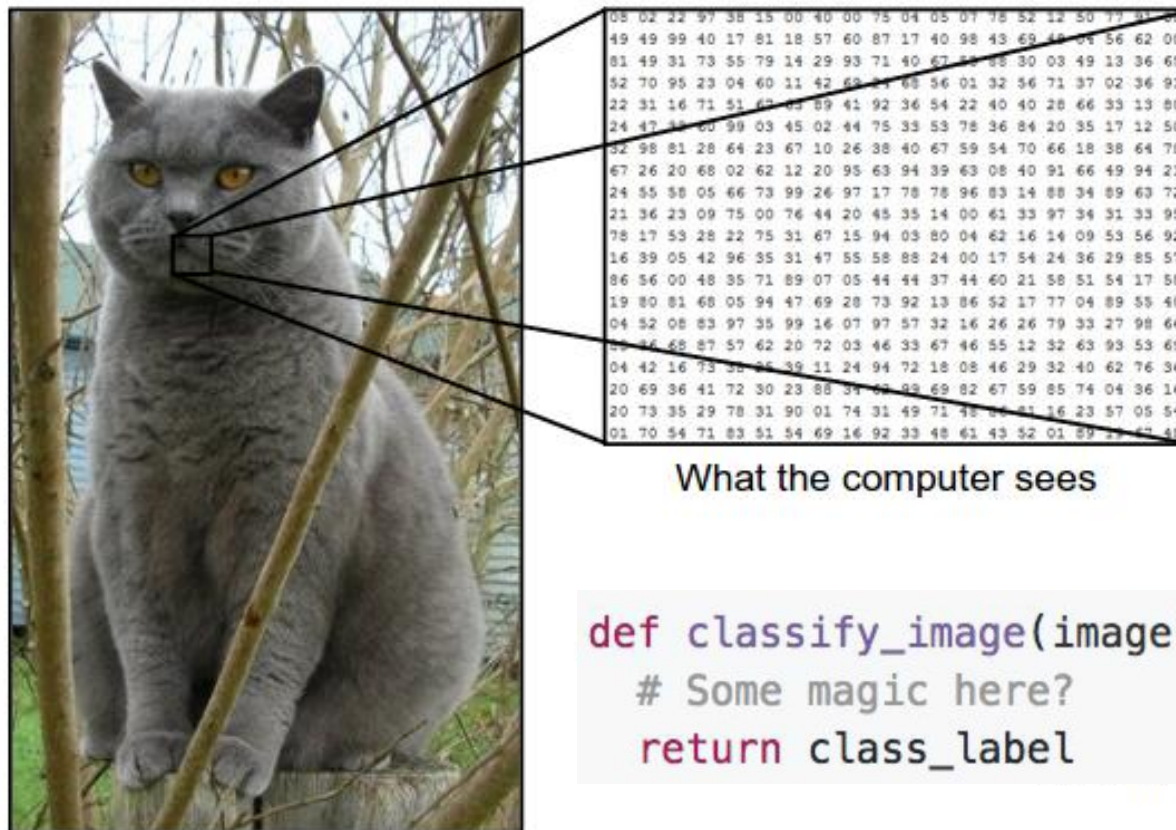
컴퓨터 비전

- 컴퓨터 비전의 어려운 점
 - 같은 종류 간의 변화가 큼



컴퓨터 비전

■ 영상 인식의 예

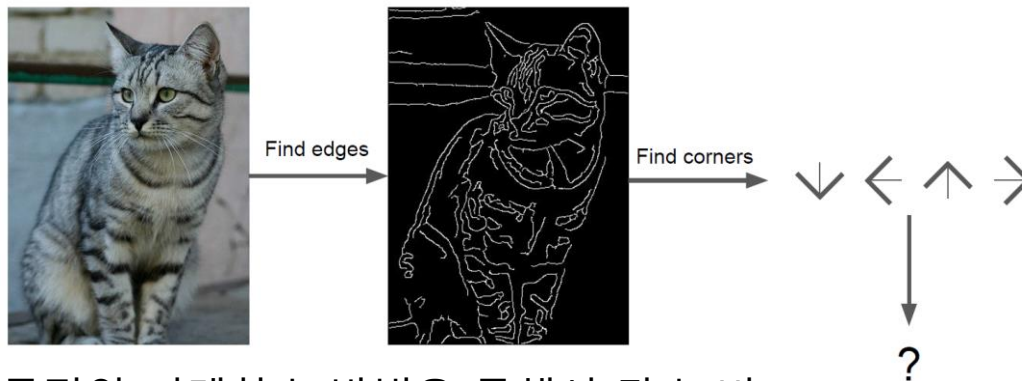


- 정해진 절차에 의해 입력을 처리하고 결과를 만드는 전통적인 프로그래밍 방법으로 해결이 어려움!

컴퓨터 비전

■ 지금까지의 시도들... 앞의 컴퓨터 비전의 어려움들을 해결하기엔 부족

- 알려진 영상 특징들을 추출하여 분류 활용



- 군집화 같은 전통적인 기계학습 방법을 통해서 단순 비교



컴퓨터 비전

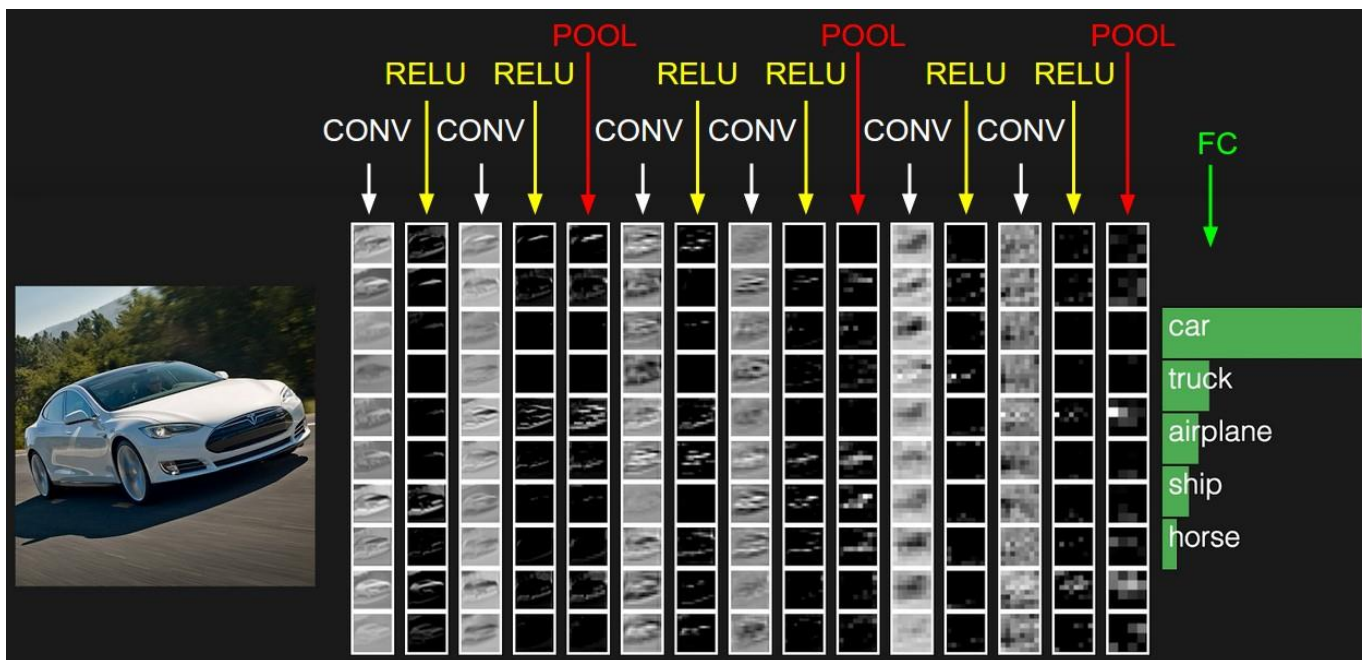
- 컴퓨터 비전 문제에서 컨볼루션 인공신경망은 중요한 해결 방법을 제공
- 실제 컴퓨터 비전 기술은 인공지능 기술의 핵심 요소
 - 로봇, 자율주행, 가상현실 (VR), 영상 제작, 영상의학, 얼굴인식 등등
 - 특히, 지능형 자동차의 핵심 요소 기술!



컨볼루션 신경망 CNN(convolutional neural network)

■ 인공신경망의 하나, 하지만 영상 입력에 특화된 구조를 가짐

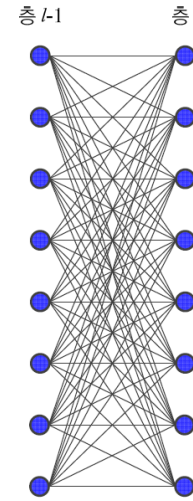
- **컨볼루션층** convolutional layer (CONV)
 - 선형함수인 컨볼루션과 비선형 함수인 활성화함수의 조합
- **풀링층** pooling layer (POOL)
 - 컨볼루션의 얻어진 특징을 통계적 값의 대표화 (압축, 요약)
- **전체 구조**
 - CONV-POOL-...-FC



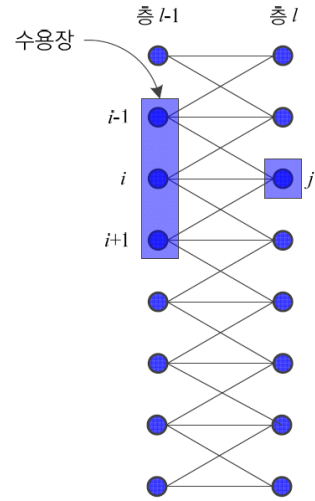
컨볼루션 신경망

■ MLP와 CNN의 비교

- MLP
 - 완전 연결 fully connection 구조로 높은 복잡도
 - 학습이 매우 느리고 과잉적합 우려
- CNN
 - 컨볼루션 연산을 이용한 부분연결 (희소 연결) 구조로 복잡도 크게 낮춤
 - 컨볼루션 연산은 영상의 좋은 특징 추출



(a) DMLP(완전연결)



(b) CNN(부분연결)

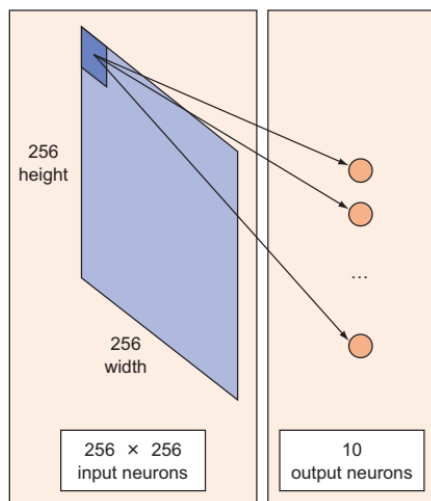
■ 영상에 특화된 CNN

- 격자 grid 구조 (영상, 음성 등)를 갖는 데이터에 적합
- 수용장 receptive field은 인간시각과 유사
- 가변 크기의 입력 처리 가능
- 희소 연결을 통해 효과적인 학습 가능

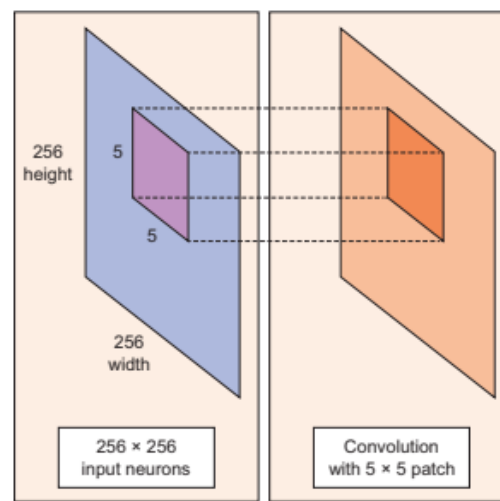
컨볼루션 신경망 CNN(convolutional neural network)

■ CNN의 MLP; 완전 연결 신경망 fully connected neural networks 과 차별

- 각 층의 **입출력의 특징 형상 유지**
- 영상의 **공간 정보**를 유지하면서 **공간적으로 인접한 정보의 특징을 효과적으로 인식**
- **복수의 커널 (혹은 필터)에 대응되는 특징들을 추출**하는 층을 가짐
- 각 커널 (필터)은 **파라미터를 공유**함으로써 MLP 대비 **학습 파라미터가 매우 적음**



256*256 크기 입력과 10개 출력
 $256 \times 256 \times 10 = 655360$ 가중치



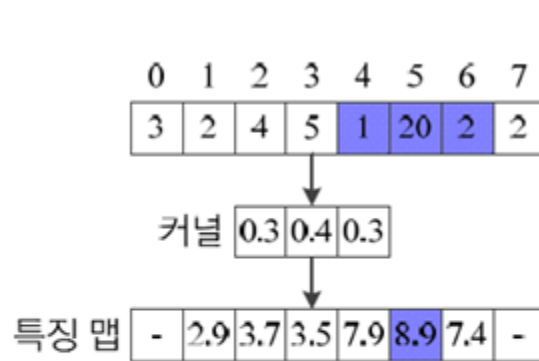
하나의 5*5 컨볼루션 필터
 $5 \times 5 = 25$ 가중치
(가중치 공유)

컨볼루션 신경망 - 컨볼루션층

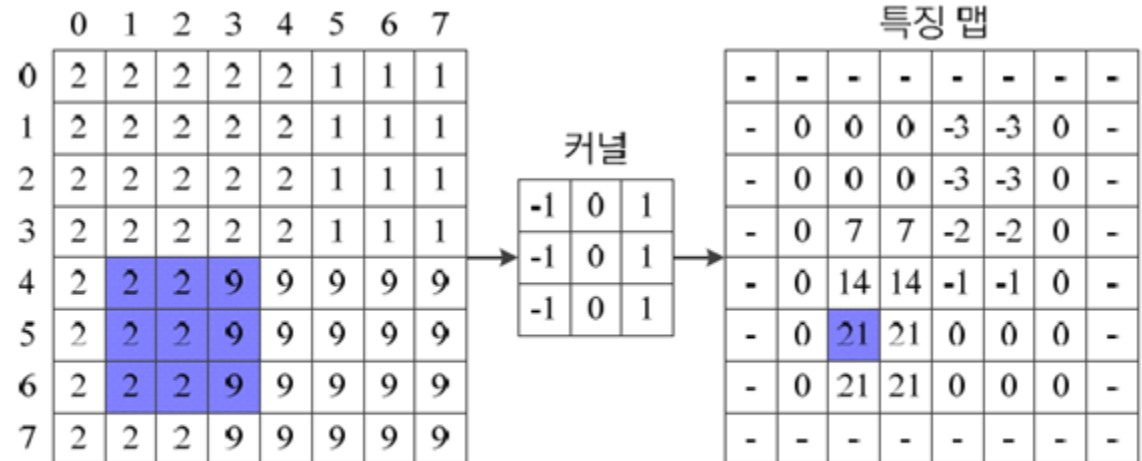
■ 컨볼루션convolution 연산

- 컨볼루션: 해당하는 요소끼리 곱하고 결과를 모두 더하는 **선형 연산 (합성곱)**
- z **입력** * (합성곱) u **커널** kernel (혹은 **필터** filter) = s 는 **출력** (**특징 맵** feature map)
 - 영상에서 **특징을 추출하기 위한 용도**로 사용됨 (= **공간 필터** spatial filtering)

※필터: 무언가를 걸러내는 도구



(a) 1차원 컨볼루션

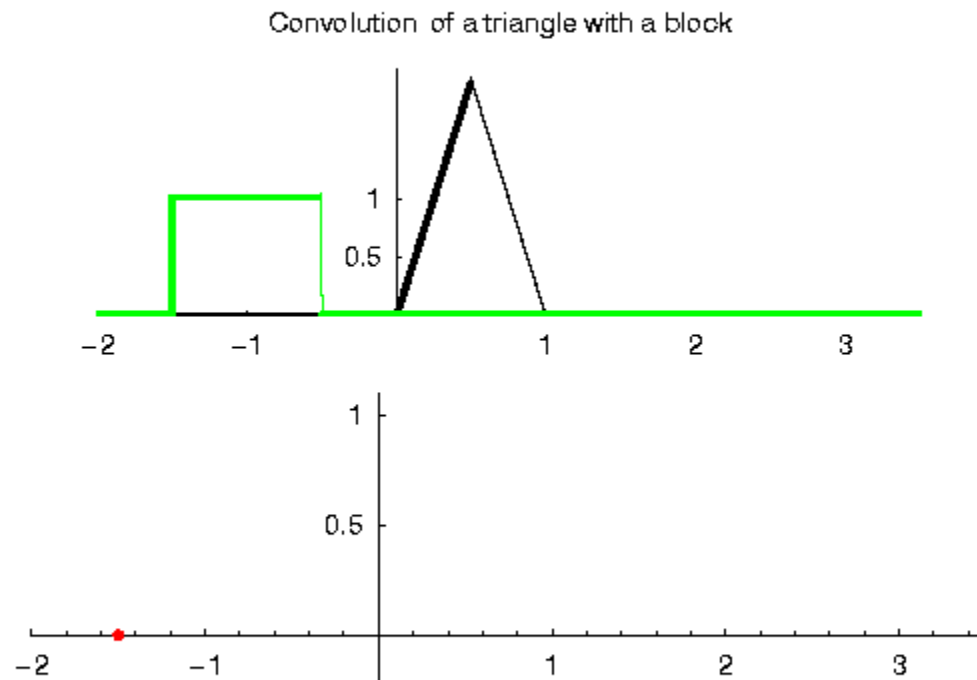


(b) 2차원 컨볼루션

컨볼루션 연산

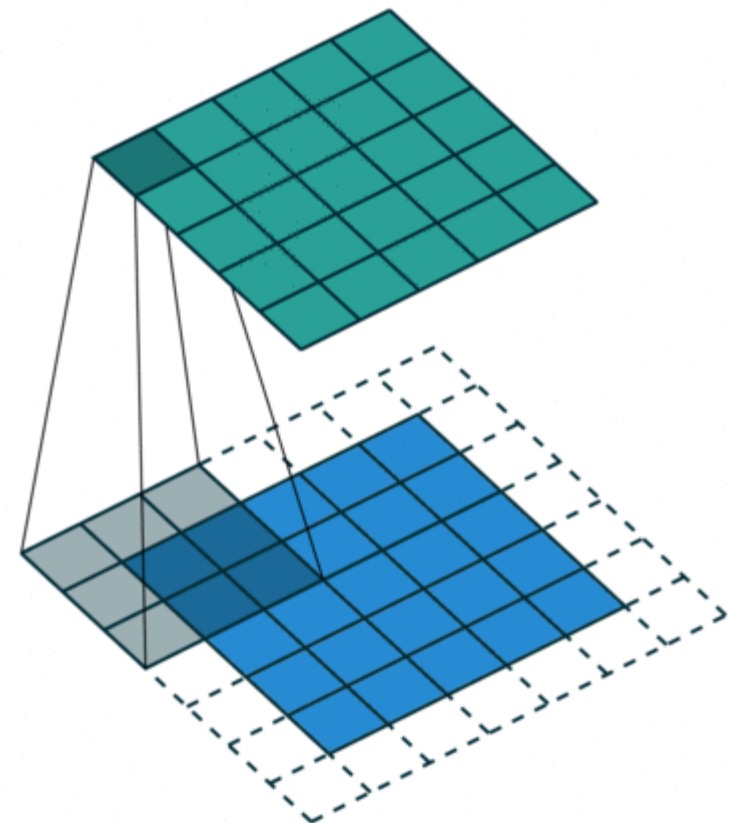
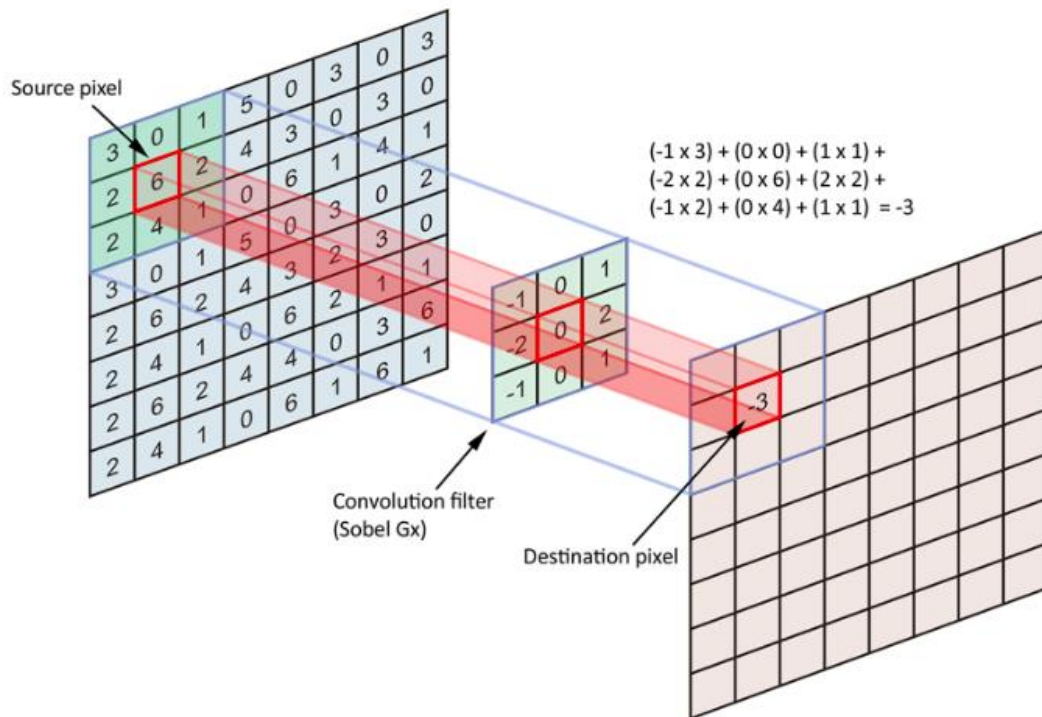
컨볼루션 신경망 - 컨볼루션층

■ 1차원 컨볼루션 연산의 예



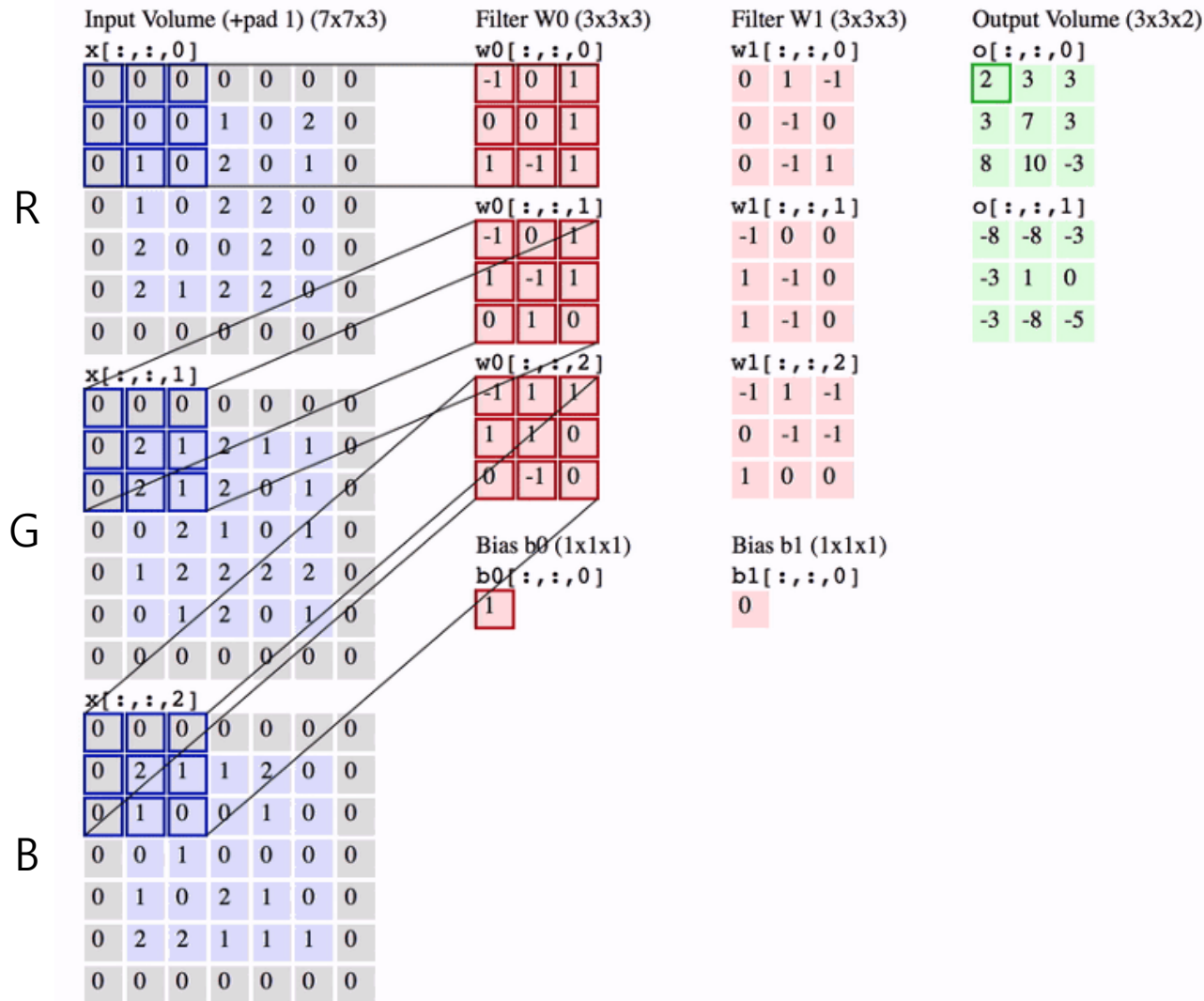
컨볼루션 신경망 - 컨볼루션층

■ 2차원 컨볼루션 연산의 예



컨볼루션 신경망 - 컨볼루션층

■ 3차원 (혹은 RGB 채널 channel) 컨볼루션 연산의 예



컨볼루션 신경망 - 컨볼루션층

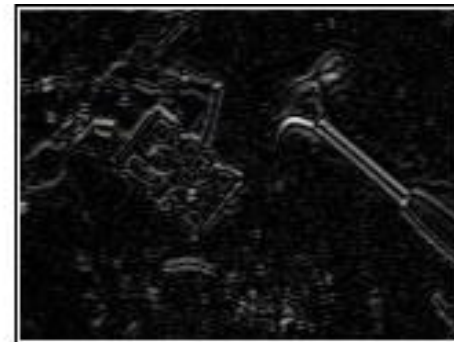
- 영상에서의 일반적인 컨볼루션(convolution) 연산 수행의 예

$$\begin{Bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{Bmatrix} \times$$

Horizontal



=

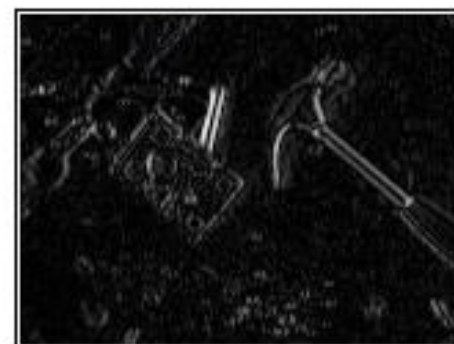


$$\begin{Bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{Bmatrix} \times$$

Vertical



=



컨볼루션 신경망 - 컨볼루션층

- 영상에서의 인공신경망에서 컨볼루션(convolution) 연산 수행의 예

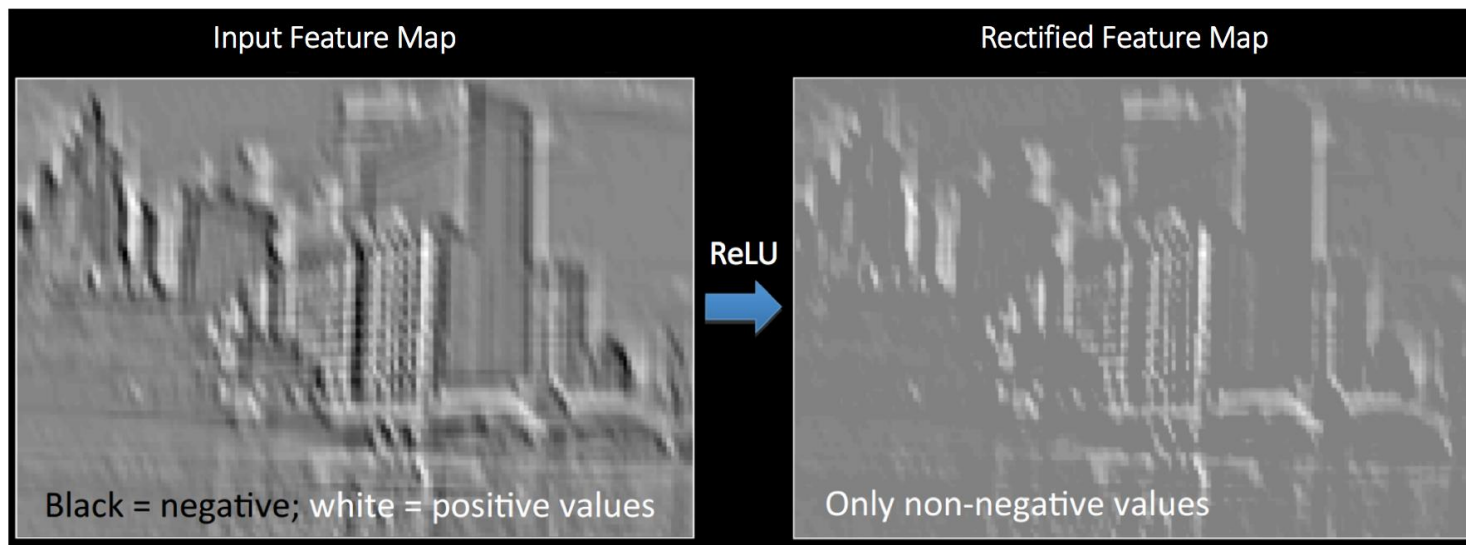
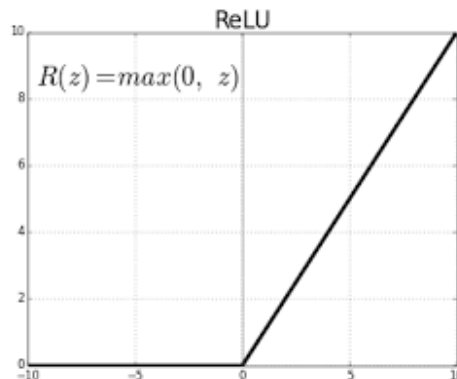


Input

커널(필터)의 값에 따라 추출하는 특징(결과)이 달라짐

컨볼루션 신경망 - 컨볼루션층

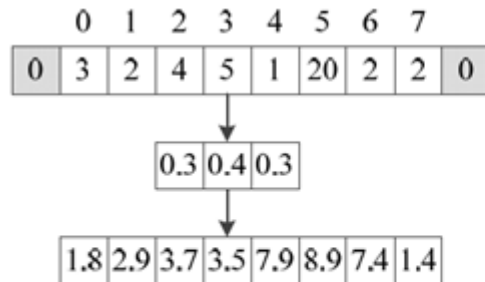
- 선형 연산: **컨볼루션** 연산
- 비선형 연산: 영상에서의 **ReLU (활성함수)** 연산과 수행 예



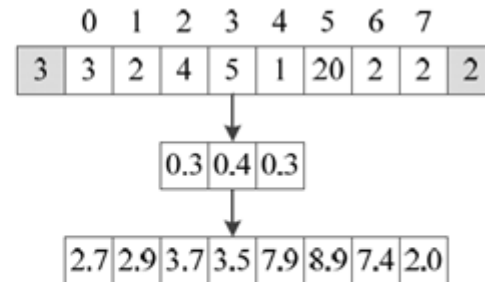
컨볼루션 신경망 - 컨볼루션층

■ 덧대기padding

- 가장자리에서 영상의 크기가 줄어드는 효과 방지 (각 층의 입출력의 특징 **형상 유지**)



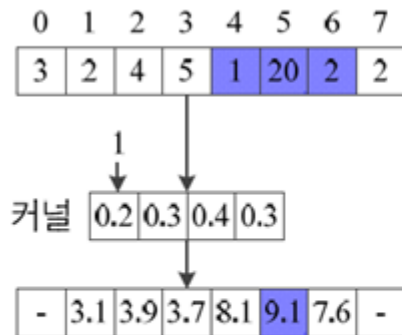
(a) 0 덧대기



(b) 복사 덧대기

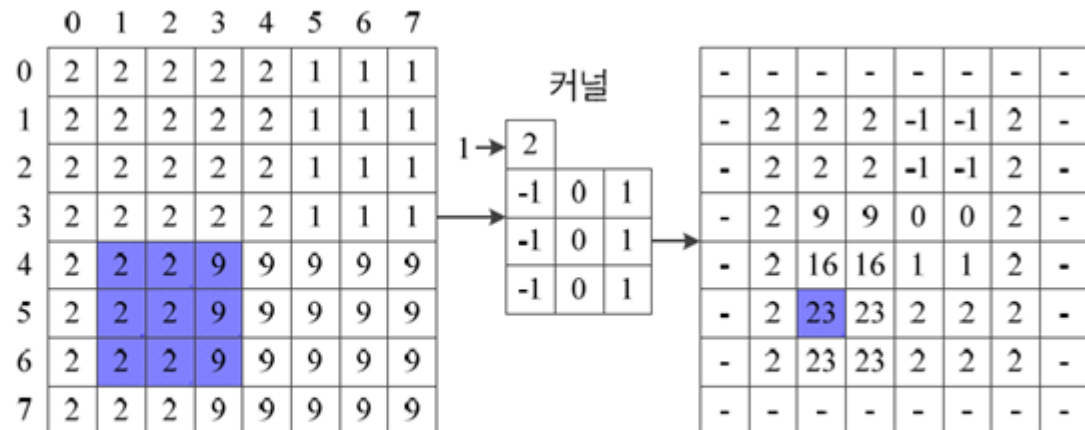
덧대기(회색 노드가 덧댄 노드)

■ 편향bias 추가



(a) 1차원 컨볼루션

바이어스

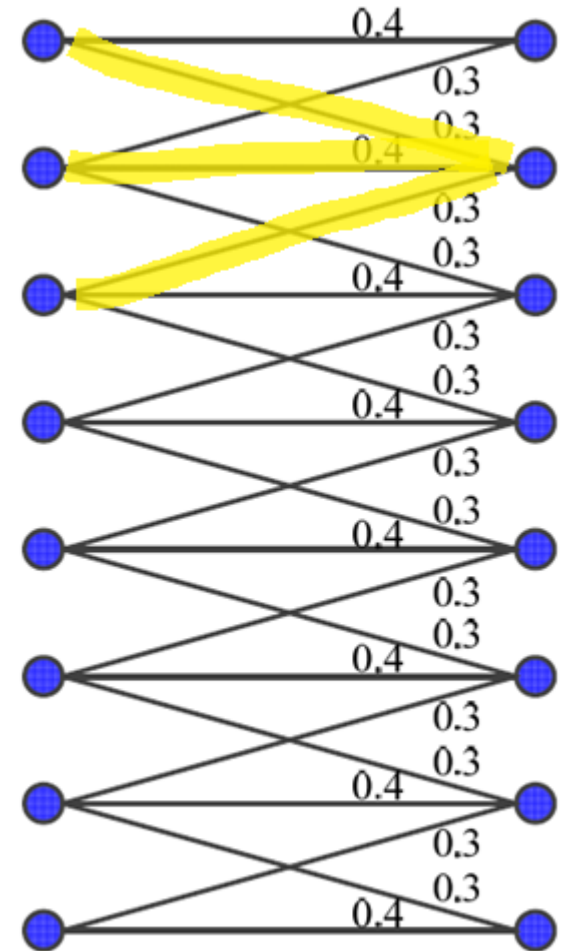


(b) 2차원 컨볼루션

컨볼루션 신경망 - 컨볼루션층

■ 가중치 공유 weight sharing 혹은 parameter sharing (묶인 가중치)

- 모든 노드가 동일한 커널(kernel)(필터)을 사용
 - 즉 가중치를 공유하므로 매개변수는 희소화
- 모델의 복잡도가 크게 낮아짐

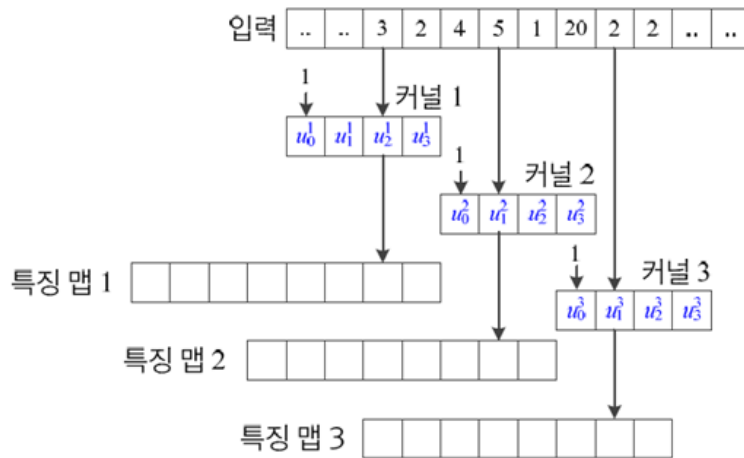


CNN의 가중치 공유

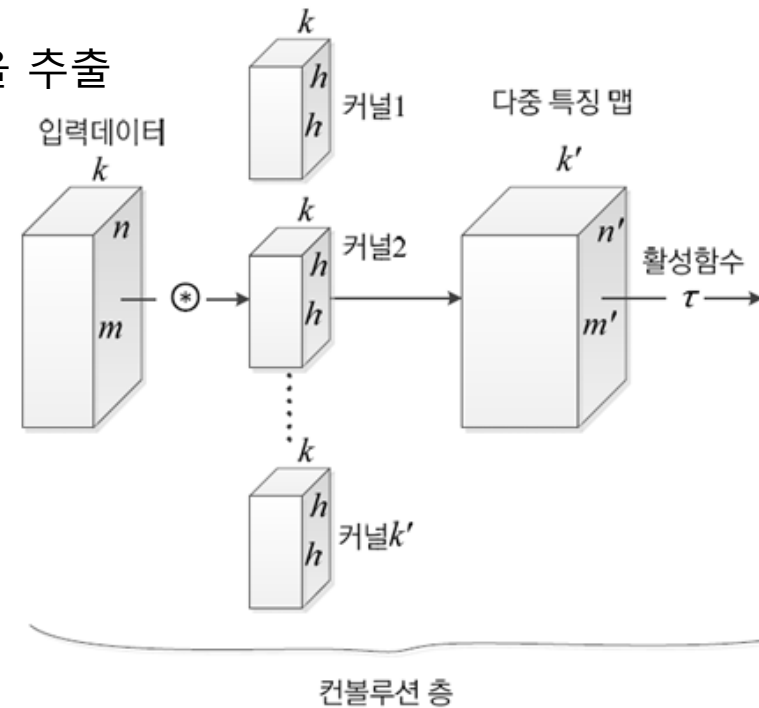
컨볼루션 신경망 - 컨볼루션층

■ 다중 특징맵 feature map 추출

- 커널(필터)의 요소값에 따라 추출되는 특징이 다름
- 예) $\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$: 수직방향, $\begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$: 수평방향 선 혹은 모서리 추출
- 하나의 커널(필터)만 사용하면 빈약한 특징이 추출됨
- 아래 그림은 3개 커널(필터)을 사용하여 3개 특징맵을 추출



다중 특징 맵 추출



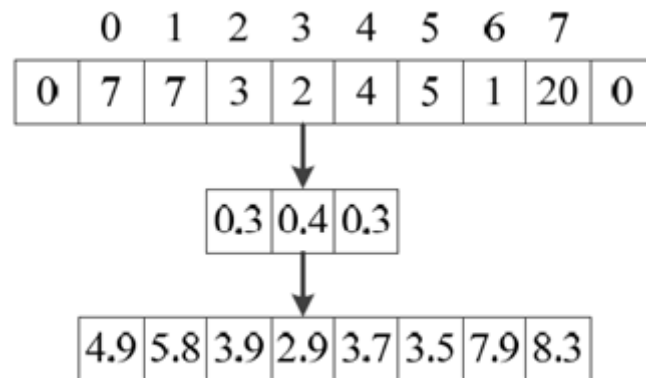
CNN의 빌딩블록

- 실제로는 수십~수백 개의 커널들을 사용
- 커널(필터)을 사람이 설계하지 않고, 학습 (오류 역전파)으로 찾음
- ← 표현 학습

컨볼루션 신경망 - 컨볼루션층

■ 컨볼루션 연산에 따른 컨볼루션 인공신경망의 특성

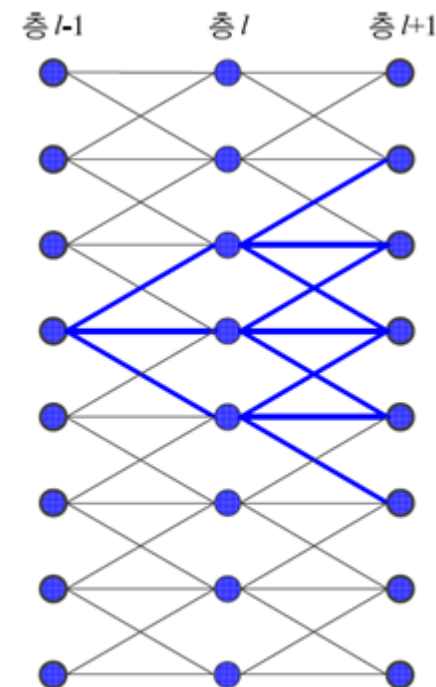
- 커널(필터)이 이동하면 입력의 신호 성분을 형상 유지하면서 특징맵에 반영 (특징 추출)



CNN의 이동에 동변한 특성

■ 병렬분산 구조

- 깊어짐에 따라 계층적인 구조 학습이 가능
- 이전층에서 추출된 정보들의 조합으로 다음층의 특징이 추출됨
→ 계층적 추상화

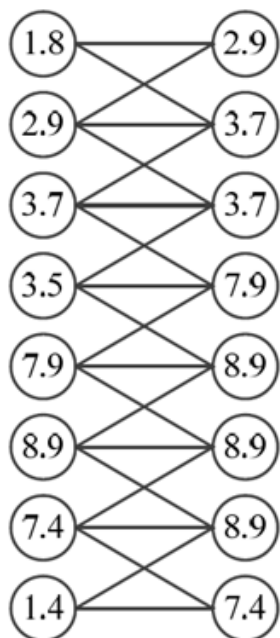


CNN의 병렬 분산 구조

컨볼루션 신경망 - 풀링층

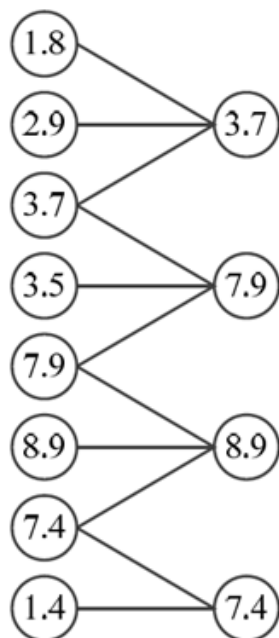
풀링 pooling 연산

- 통계적 대표값으로 특징 요약, 압축
 - 최대 풀링, 평균 풀링, 가중치 평균 풀링 등
- 보폭을 크게 하면 다운샘플링 downsampling 효과
- 최대 풀링 max pooling 예

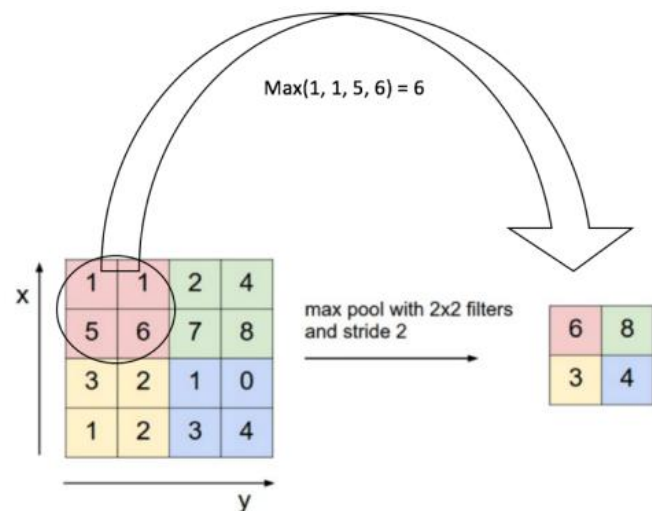


(a) 보폭 1

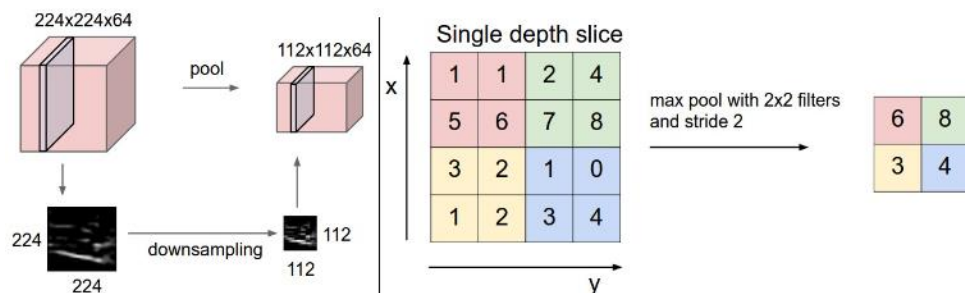
최대 풀링



(b) 보폭 2



Rectified Feature Map

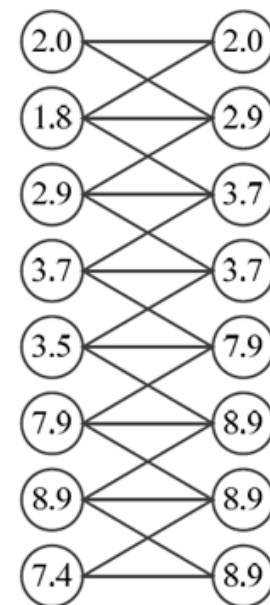


컨볼루션 신경망 - 풀링층

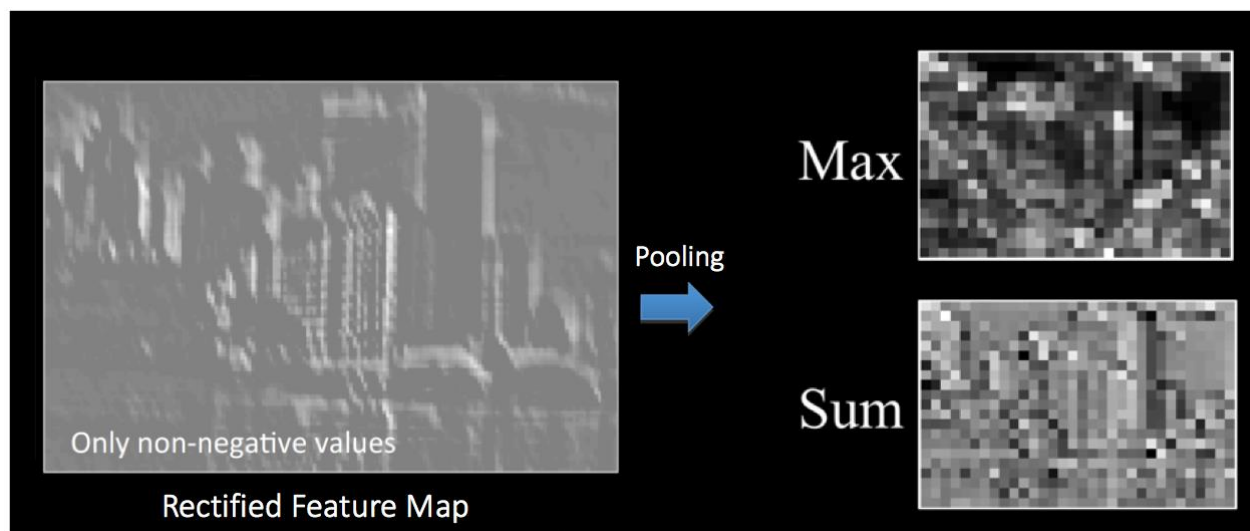
■ 풀링 연산의 특성

- 매개변수가 없음
- 특징맵의 수(깊이)를 그대로 유지함 (가로/세로 크기 축소)
- 작은 변화를 덜 민감하게 함

- 영상을 입력으로 가지는 경우, 풀링 연산 수행 예
 - 컨볼루션층의 출력 결과인 특징맵에 적용 함



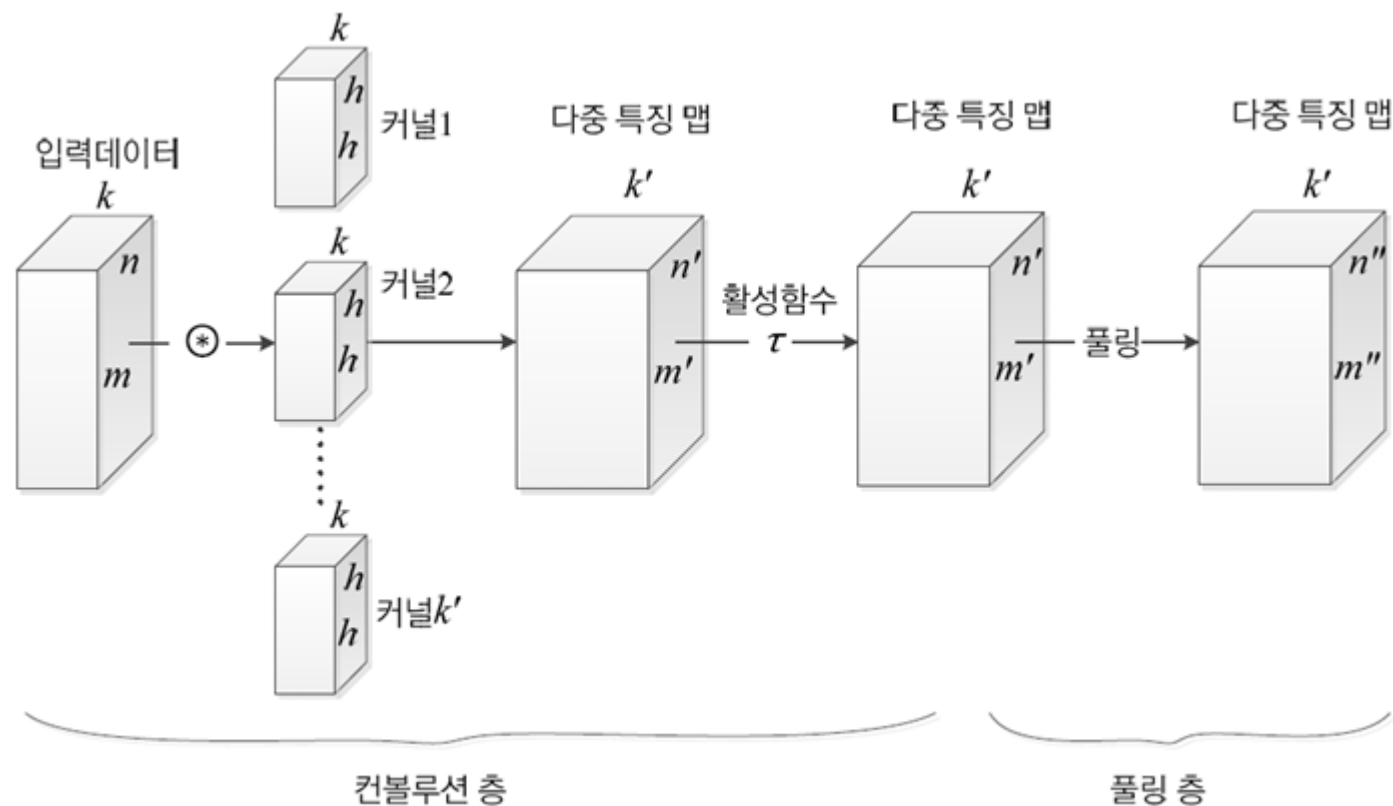
작은 이동에 둔감한 최대 풀링



컨볼루션 신경망 - 전체 구조

■ 빌딩 블록 building block

- 컨볼루션 인공신경망은 빌딩 블록을 이어 붙여 **깊은 구조로 확장**
- 전형적인 빌딩블록: **컨볼루션층**(컨볼루션 연산+활성함수 [주로 ReLU 사용]) → **풀링층**
- 다중 커널(필터)을 사용하여 다중 특징맵을 추출함

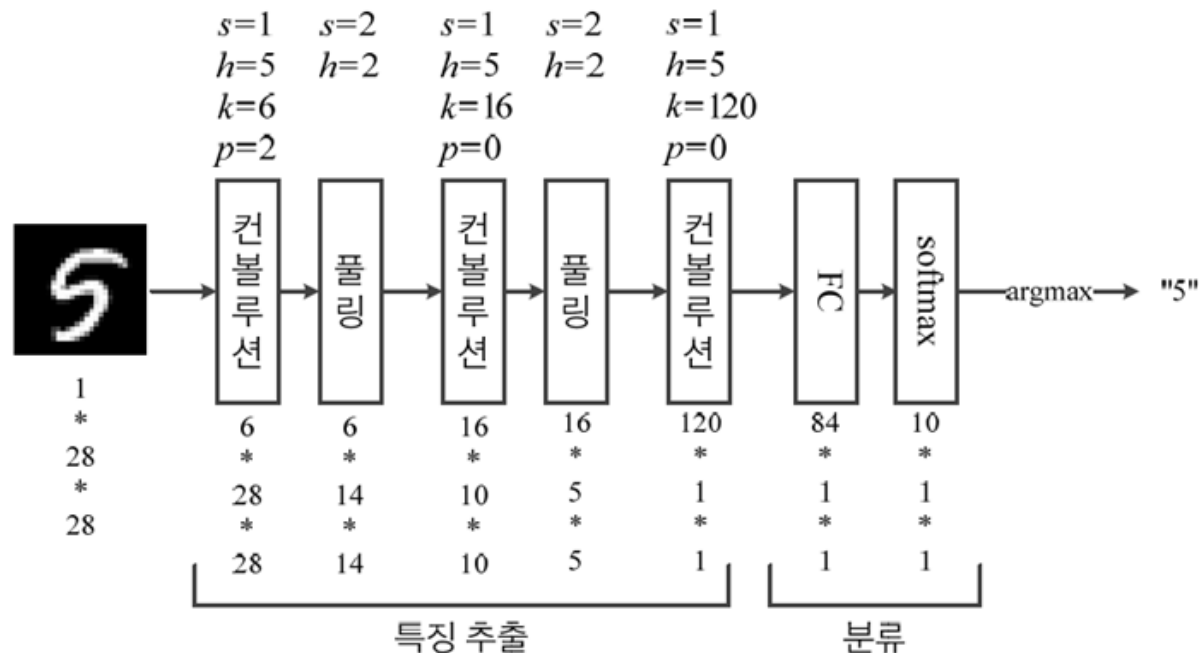


CNN의 빌딩블록

컨볼루션 신경망의 응용 - LeNet

■ 초창기 CNN 사례로서 LeNet-5

- **특징 추출**: CONV-POOL-CONV-POOL-CONV의 다섯 층을 통해 28*28 명암 영상을 120차원의 특징 벡터로 변환
 - 평균 풀링
- **분류**: 은닉층이 하나인 MLP
- CNN의 첫 번째 성공사례: 필기 숫자 인식기 만들어 수표 인식 자동화 시스템 구현



LeNet-5 구조

컨볼루션 신경망의 응용

■ 영상 분류 image classification: 도전적 문제

■ ImageNet 데이터베이스

- 2만 2천여 부류에 대해 부류별로 수백~수만장의 영상을 인터넷에서 수집하여 1500만여 장의 영상을 구축하고 공개

■ ILSVRC ImageNet Large Scale Visual Recognition Competition 영상 분류 대회 (CVPR 학술대회에서 개최)

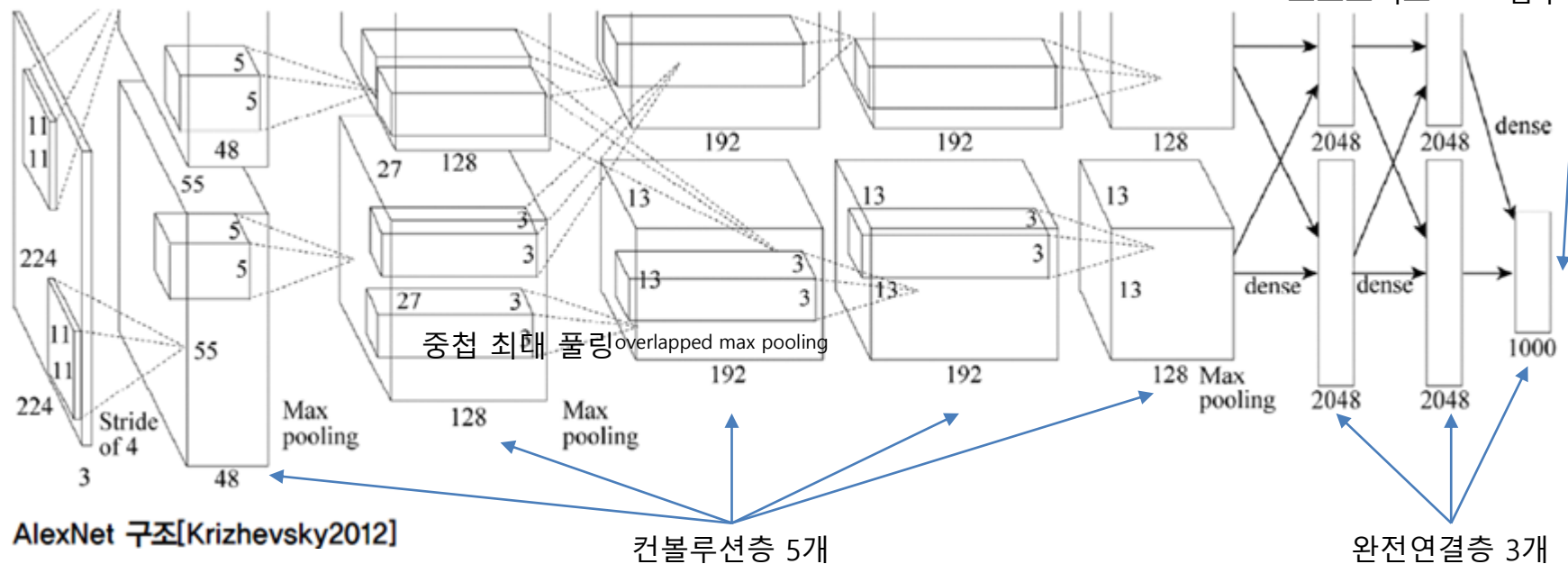
- 1000부류에 대해 분류, 검출, 위치 지정 문제: 1순위와 5순위 정확도 측정
 - 120만 장의 훈련집합, 5만 장의 검증집합, 15만 장의 테스트집합
- 우승한 CNN은 프로그램과 가중치를 공개함으로써 널리 사용되는 표준 신경망이 됨



컨볼루션 신경망의 응용 - AlexNet

■ 구조

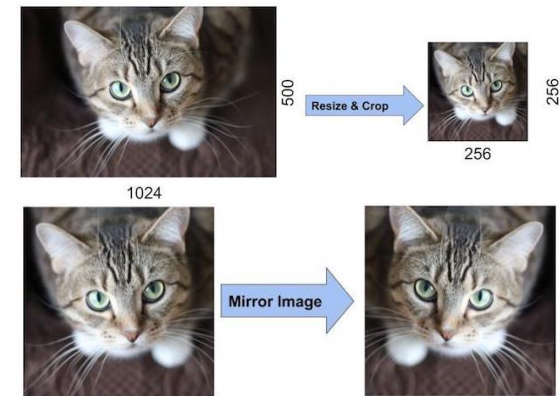
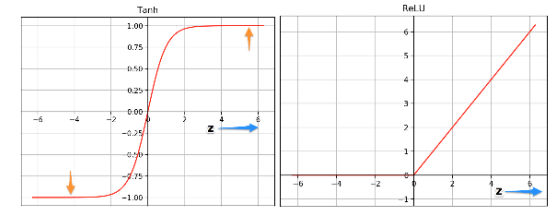
- 컨볼루션층 5개와 완전연결 fully connected (FC) 층 3개
 - 8개 층에 290400-186624-64896-43264-4096-4096-1000개의 노드 배치
 - 컨볼루션층은 200만개, FC층은 6500만개 가량의 매개변수
 - FC층에 30배 많은 매개변수 → 향후 CNN은 FC층의 매개변수를 줄이는 방향으로 발전
- 1000개의 분류를 위해 소프트맥스 softmax 함수 사용



컨볼루션 신경망의 응용 - AlexNet

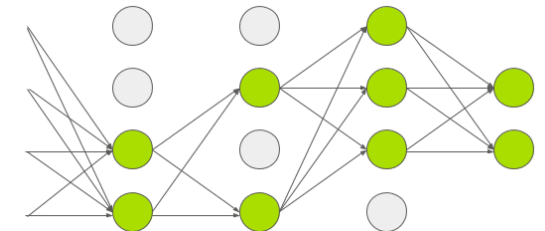
■ AlexNet이 학습에 성공한 요인

- 외적 요인
 - ImageNet이라는 대용량 데이터베이스
 - GPU를 사용한 병렬처리 (GPU#1, GPU#2으로 분할)
- 내적 요인
 - 활성화함수로 ReLU 사용
 - 지역 반응 정규화^{local response normalization} 기법 적용
 - 1번째, 3번째 최대 풀링 전 적용
 - 과잉적합 방지하는 여러 규제 기법 적용
 - 데이터 확대 (잘라내기^{cropping}와 반전^{mirroring}으로 2048배로 확대)
 - 드롭아웃^{dropout} 등
 - » 완전연결층에서 사용됨



■ 테스트 단계에서 앙상블 적용

- 입력된 영상을 잘라내기와 반전을 통해 증가하고, 증가된 영상들의 예측 평균으로 최종 인식
- 2~3%만큼 오류율 감소 효과

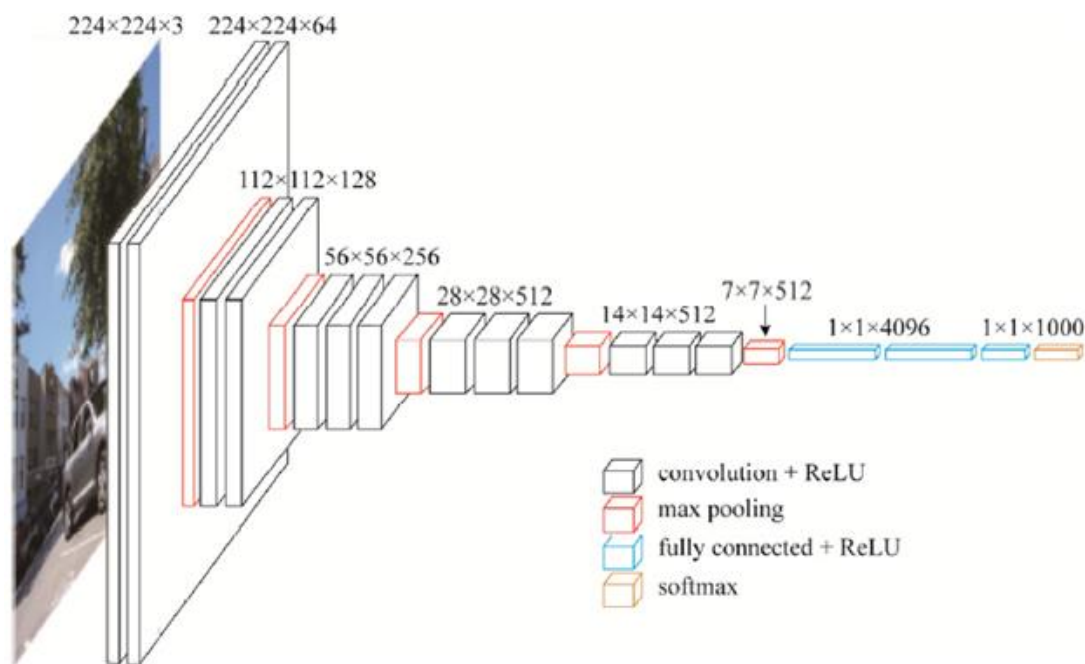


컨볼루션 신경망의 응용 - VGGNet

■ VGGNet의 핵심 아이디어

- 3*3의 작은 커널을 사용
- 신경망을 더욱 깊게 만듦 (신경망의 깊이가 어떤 영향을 주는지 확인)
- 컨볼루션층 8~16개를 두어 AlexNet의 5개에 비해 2~3배 깊어짐

■ 16층짜리 VGG-16 (CONV 13층+FC 3층)

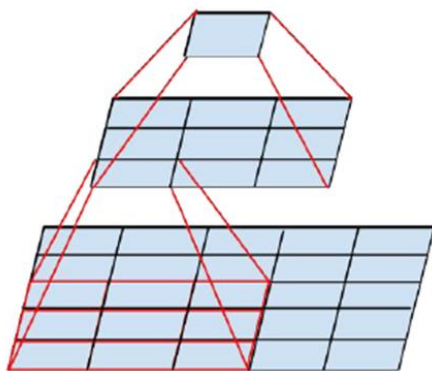


VGGNet 구조[Simonyan2015]

컨볼루션 신경망의 응용 - VGGNet

■ 작은 커널의 이점

- 큰 크기의 커널은 여러 개의 작은 크기 커널로 분해될 수 있음
→ 매개변수의 수는 줄어들면서 신경망은 깊어지는 효과
- 예) 5×5 커널을 2층의 3×3 커널로 분해하여 구현



매개변수

5×5 커널인 경우, 25

3×3 커널인 경우, $9 + 9 = 18$

컨볼루션 신경망의 응용 - VGGNet

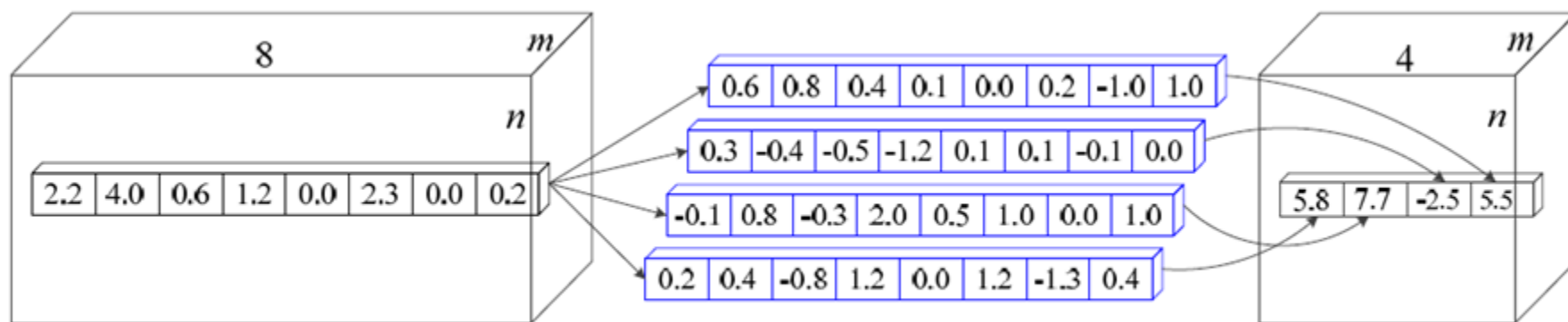
■ 1*1 커널

■ 차원 축소 효과

- $c_2 > c_3$: 차원 축소 (연산량 감소), 깊은 신경망

■ 예

- $m*n$ 의 특징 맵 8개에 1*1 커널을 4개 적용 \rightarrow $m*n$ 의 특징 맵 4개가 됨
- 다시 말하면, $8*m*n$ 텐서에 $8*1*1$ 커널을 4개 적용하여 $4*m*n$ 텐서를 출력하는 셈



1*1 컨볼루션 예제

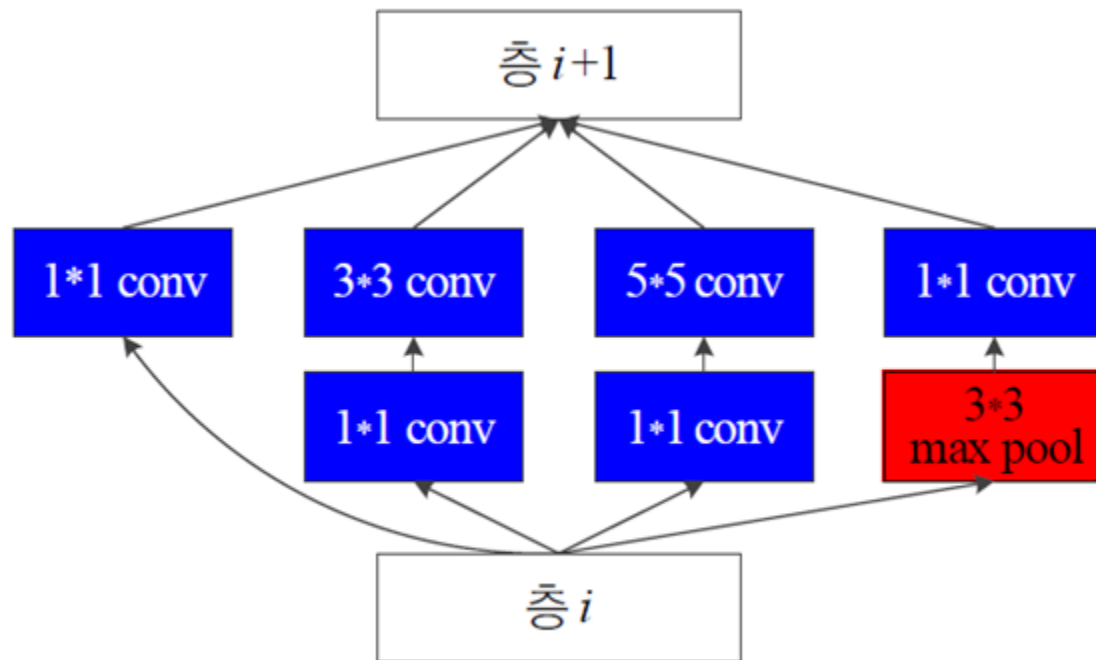
- VGGNet은 적용 실험을 하였지만 최종 선택하지는 않음 (GoogLeNet에서 사용됨)

컨볼루션 신경망의 응용 - GoogLeNet

■ GoogLeNet은 NIN 개념을 확장한 신경망

■ 인셉션 모듈

- 다양한 크기의 필터들을 컨볼루션 연산 사용 → 다양한 특징 추출
- 1*1 컨볼루션을 사용하여 차원 축소 dimension reduction
 - 매개변수의 수 (=특징 맵의 수)를 줄임 + 깊은 신경망 → 연산량 감소
- 3*3, 5*5 같은 다양한 크기의 컨볼루션들을 통해서 다양한 특징들을 추출

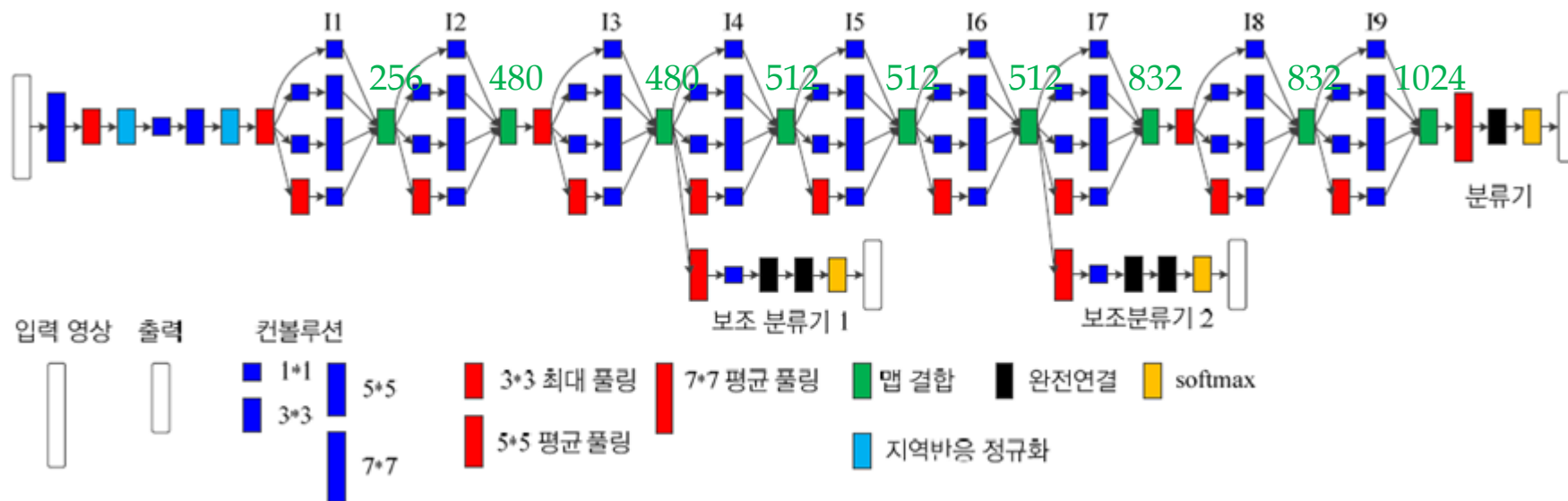


GoogLeNet의 인셉션 모듈

컨볼루션 신경망의 응용 - GoogLeNet

■ 인셉션 모듈(I)을 9개 결합한 GoogLeNet

- 매개변수가 있는 층 22개, 없는 층 (풀링) 5개로 총 27개 층
- 완전 연결층은 1개에 불과
 - 1백만 개의 매개변수를 가지며, VGGNet의 완전 연결층에 비하면 1%에 불과



GoogLeNet의 구조

- 보조 분류기 (auxiliary classifier)
- 원 분류기의 오류 역전파 결과와 보조 분류기의 오류 역전파 결과를 결합함으로써 **그레디언트 소멸 문제 완화**
- 학습시에만 도우미 역할을 하고, 동작시에는 제거됨

컨볼루션 신경망의 응용 - ResNet

■ ResNet

- 잔류 학습(residual learning)이라는 개념을 이용하여 성능 저하를 피하면서 층 수를 대폭 늘림 (최대 1202층까지) ← 깊은 신경망일수록 데이터의 대표적인 특징을 잘 추출할 수 있음

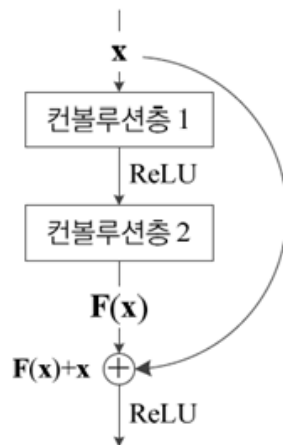
- 원래 컨볼루션 신경망

$$\mathbf{F}(\mathbf{x}) = \tau(\mathbf{x} \circledast \mathbf{w}_1) \circledast \mathbf{w}_2$$
$$\mathbf{y} = \tau(\mathbf{F}(\mathbf{x}))$$

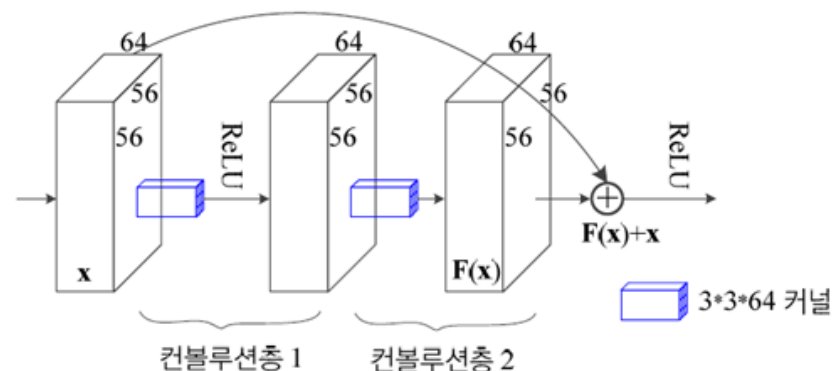
- 잔류 학습은 지름길 연결된 \mathbf{x} 를 더한 $\mathbf{F}(\mathbf{x}) + \mathbf{x}$ 에 τ 를 적용. $\mathbf{F}(\mathbf{x})$ 는 잔류(residual)

$$\mathbf{y} = \tau(\mathbf{F}(\mathbf{x}) + \mathbf{x})$$

→ 그레디언트 소멸 문제 해결 $\frac{\partial \mathcal{E}}{\partial \mathbf{x}_l} = \frac{\partial \mathcal{E}}{\partial \mathbf{x}_L} \frac{\partial \mathbf{x}_L}{\partial \mathbf{x}_l} = \frac{\partial \mathcal{E}}{\partial \mathbf{x}_L} \left(1 + \frac{\partial}{\partial \mathbf{x}_l} \sum_{i=l}^{L-1} \mathbf{F}(\mathbf{x}_i) \right)$



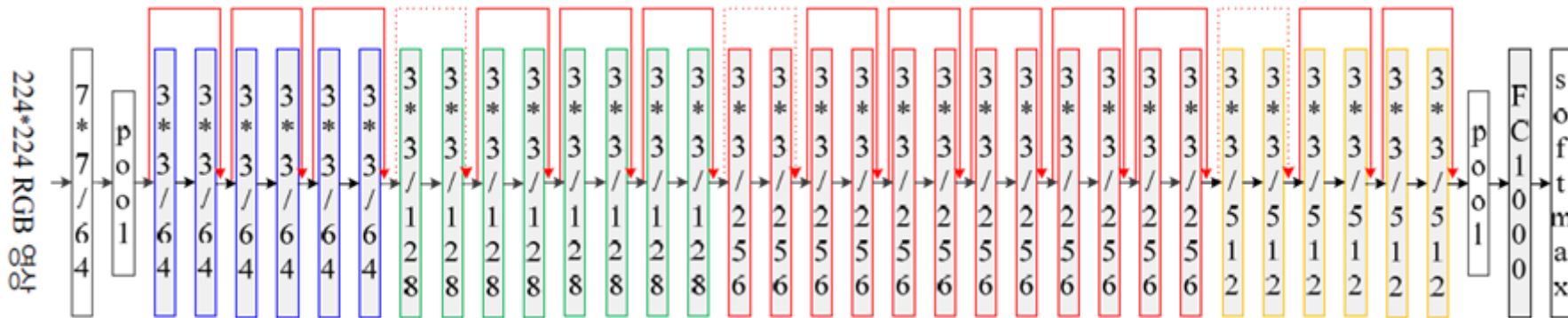
(a) 빌딩블록



(b) 빌딩블록 사례

컨볼루션 신경망의 응용 - ResNet

■ 34층짜리 ResNet 예시



ResNet 예제(34층)

- VGGNet과 같은 점

- 3*3 커널 사용

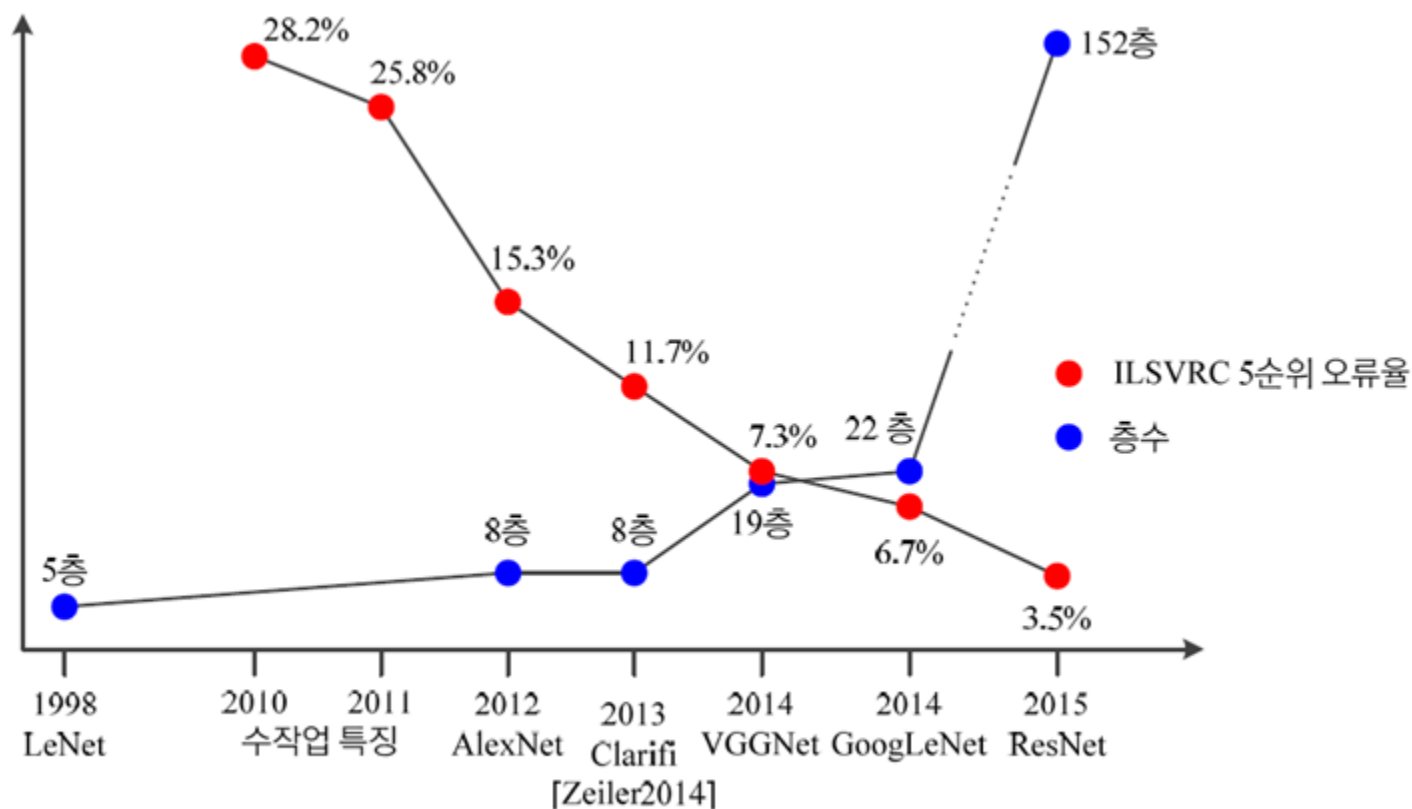
■ VGGNet과 다른 점

- 잔류 학습 사용
- 전역 평균 풀링 사용 (FC 층 제거)
- 배치 정규화^{batch normalization} 적용 (드롭아웃 적용 불필요)

컨볼루션 신경망의 응용

■ ILSVRC 대회 성적

- 2012년 AlexNet의 15.3% 오류율은 당시로서 경이로운 성능
- 2015년에 ResNet은 3.5% 오류율 달성
 - 사람의 경우, 5.1% 오류율을 가짐



컨볼루션 신경망 정리

- 컴퓨터 비전 문제의 효과적인 해결책 제시
 - 사람인 시각인지와 유사한 접근법
- 영상에 특화된 인공신경망 구조 제시
 - CONV-POOL-FC 조합을 쌓아 입력의 공간적 특징들을 자동 추출하여 문제 해결
 - 작은 필터들로 구성된 층이 깊게 만드는 경향
 - POOL/FC 층을 제거하는 경향
- 실제 다양한 응용에서 주목할 만한 결과를 보임
 - 지능형 자동차의 핵심 요소 기술임 (객체 인식)