

4.3 컨볼루션 신경망

2021210088 허지혜

4.3 컨볼루션 신경망(Convolutional Neural Network)

DMLP와 CNN 비교

DMLP

완전 연결 구조로 높은 복잡도를 가진다.
학습이 매우 느리고 과잉 적합의 우려가 있다.

CNN

컨볼루션 연산을 이용한 부분연결(또는 희소연결) 구조로
복잡도를 크게 낮춘다.
컨볼루션 연산을 통해 좋은 특징을 추출한다.

CNN 특징

격자 구조(영상, 음성)를 가지는 데이터에 적합하다.
 j 노드가 연결된 왼쪽 층의 영역을 표시한 파란색 부분을
수용장이라고 한다.
이때 수용장은 인간 시각과 유사하다고 한다.
가변 크기의 입력 처리가 가능하다.

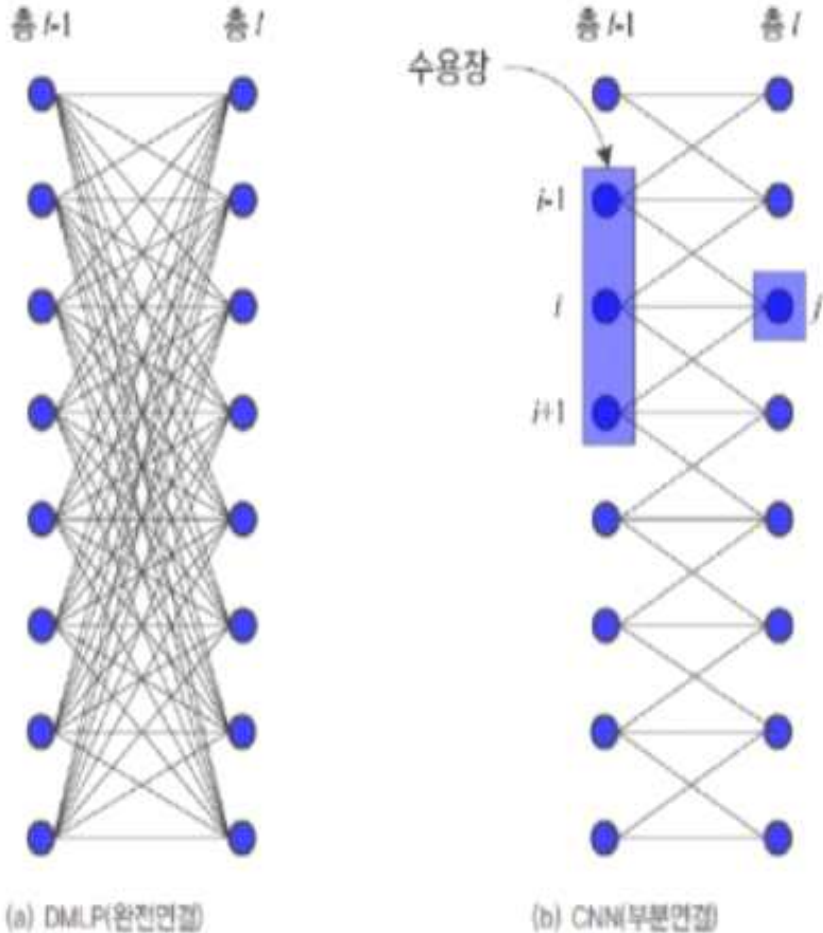


그림 4-5 CNN의 부분연결과 수용장

4.3.1 컨볼루션층

컨볼루션 연산

해당하는 요소끼리 곱하고 결과를 모두 더하는 선형 연산이다.

u : kernel, z : input, s : output(feature map)

$$s(i) = z \odot u = \sum_{x=-(h-1)/2}^{(h-1)/2} z(i+x)u(x) \quad (4.10)$$

1차원 입력

$$s(j,i) = z \odot u = \sum_{y=-(h-1)/2}^{(h-1)/2} \sum_{x=-(h-1)/2}^{(h-1)/2} z(j+y,i+x)u(y,x) \quad (4.11)$$

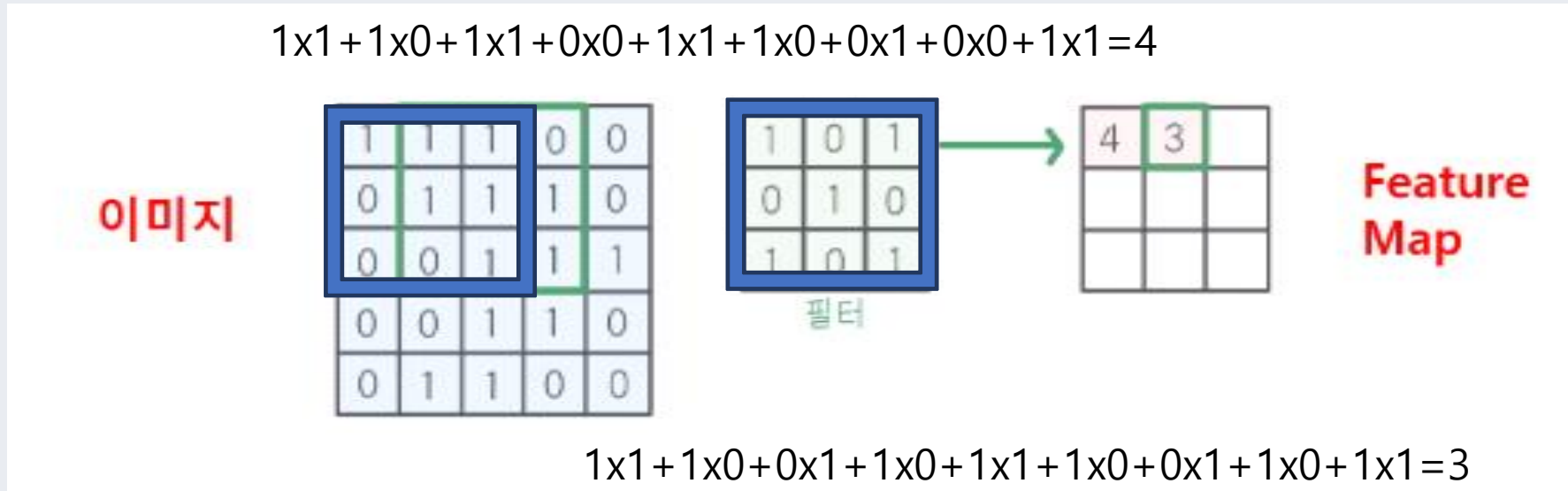
2차원 입력



그림 4-6 컨볼루션 연산

4.3.1 컨볼루션층

컨볼루션 연산



4.3.1 컨볼루션층

덧대기(Padding)

가장 자리에서 영상의 크기가 줄어드는 효과를 방지하기 위해 padding을 한다(feature map 형상 유지)

종류) 0 덧대기, 복사 덧대기



그림 4-7 덧대기(회색 노드가 덧댄 노드)

바이어스 추가

커널에 편향을 추가한다.

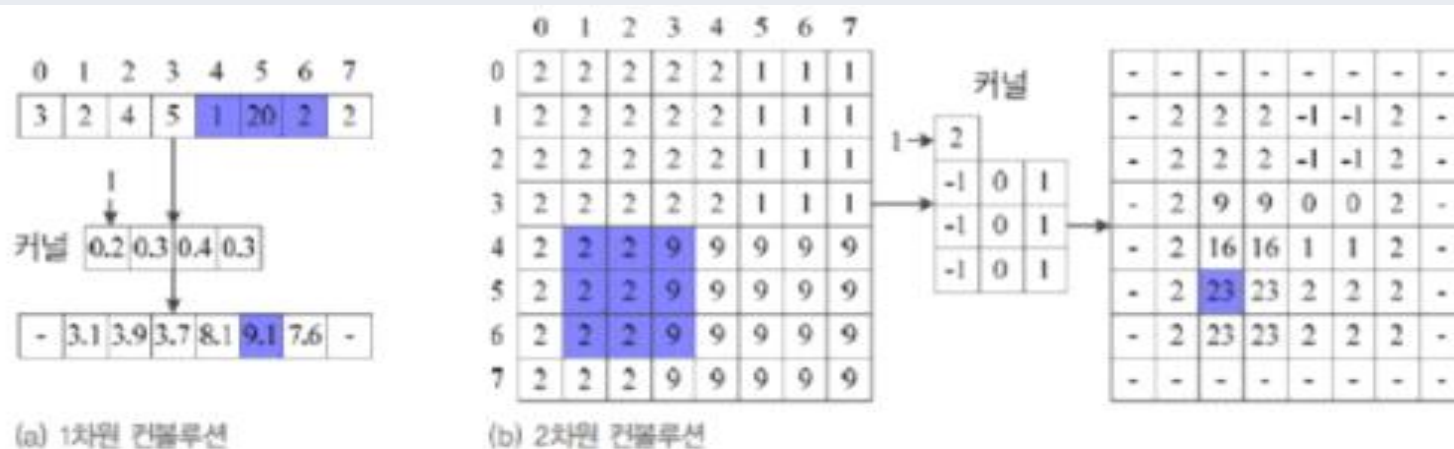


그림 4-8 바이어스

4.3.1 컨볼루션층

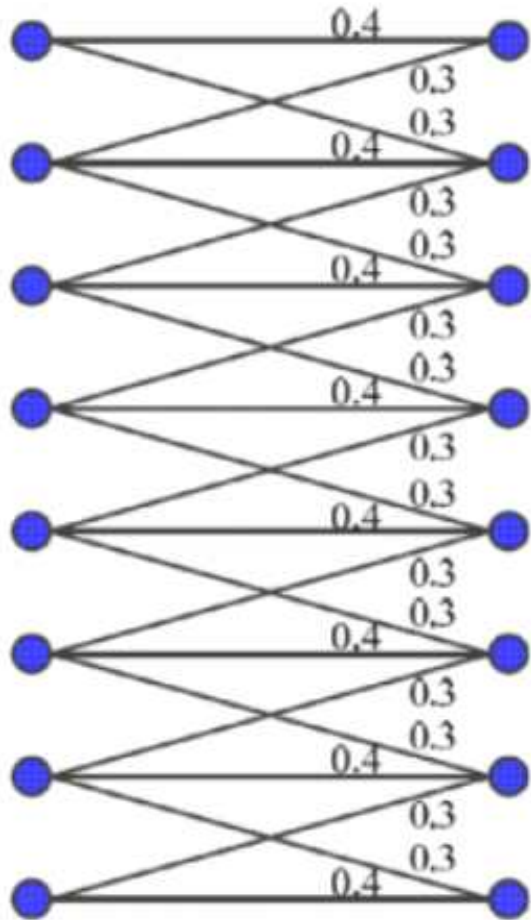


그림 4-9 CNN의 가중치 공유

가중치 공유(묶인 가중치)

모든 노드가 동일한 커널을 사용(즉 가중치를 공유)
하므로 매개변수는 3개에 불과하다.
모델의 복잡도가 크게 낮아진다.

$$\text{CNN 가중치 개수} : 3 * 8 - 2 = 22$$

4.3.1 컨볼루션층

다중 특징 맵 추출

커널의 값에 따라 커널이 추출하는 특징이 달라진다.

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 에지 추출

하나의 커널만 사용하면 너무 빈약한 특징이 추출되었다.

아래 사진은 3개 커널을 사용해 3개 특징 맵을 추출하는 상황이다. 실제로는 아래보다 더 많은 커널을 사용해 특징 맵을 추출한다.

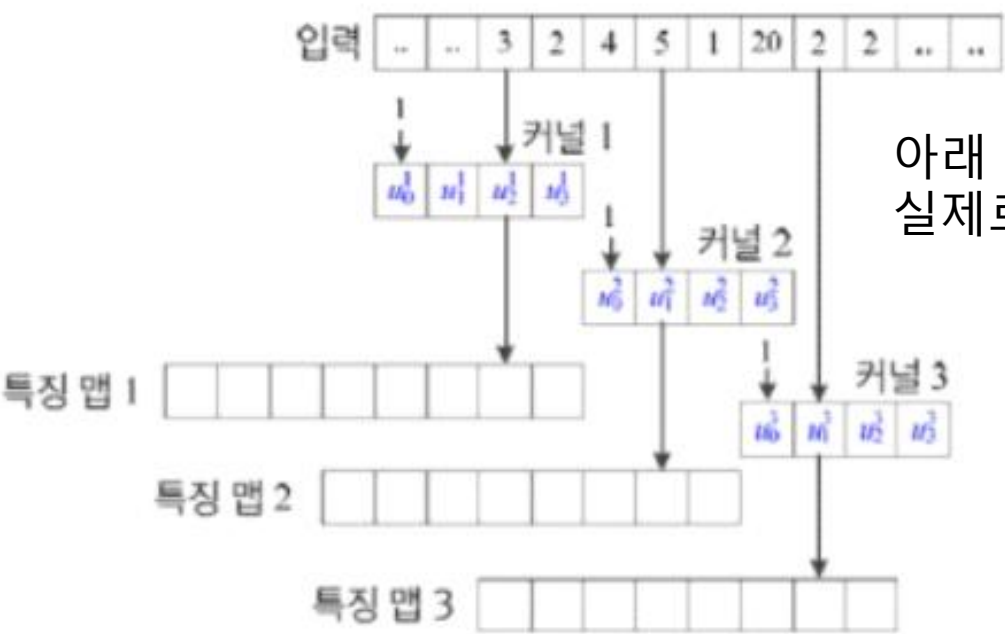


그림 4-10 다중 특징 맵 추출

4.3.1 컨볼루션층

특징 학습

커널을 사람이 설계하던 기존의 수작업 특징방식을 벗어나, 기계학습으로 훈련집합에 최적인 커널을 자동으로 찾아준다.

w_{ki} 는 k번째 커널의 i번째 매개변수

예) 2차원 영상이 7*7 커널을 64개 사용한다면,
학습은 $(7*7+1)*64 = 3200$ 개의 매개변수를 찾아내야한다.

커널은 어떤 특징 맵을 추출할지 규정하므로, 이러한 학습 과정을 특징 학습 또는 표현 학습이라고 부른다. DMLP와 마찬가지로 오류 역전파로 커널을 학습한다.

4.3.1 컨볼루션층

컨볼루션 연산에 따른 CNN의 특성

CNN은 기본 연산이 컨볼루션이기 때문에 컨볼루션 성질에 직접적인 영향을 받는다.

1) 컨볼루션은 이동에 동변이다.

즉, 신호가 이동하면 이동 정보가 그대로 특징 맵에 반영된다.

따라서 영상 인식에서 물체의 이동이나 음성 인식에서 발음 지연에 효과적으로 대처된다.

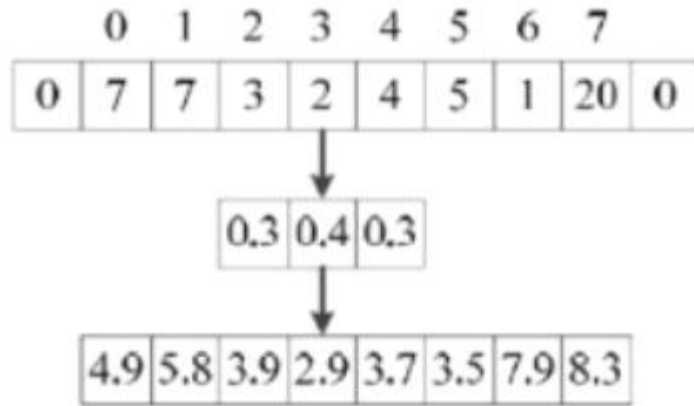


그림 4-11 CNN의 이동에 동변한 특성

2) 병렬분산 구조이다.

각 노드는 독립적으로 계산 가능하므로 병렬 구조이다.

노드는 깊은 층을 거치면서 전체에 영향을 미치므로 분산 구조이다.

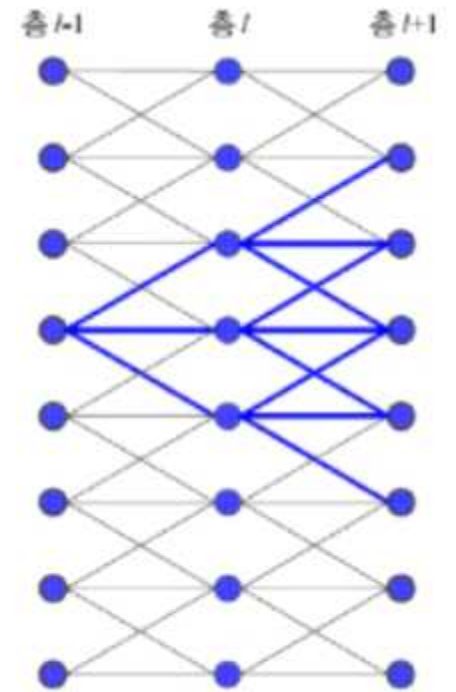
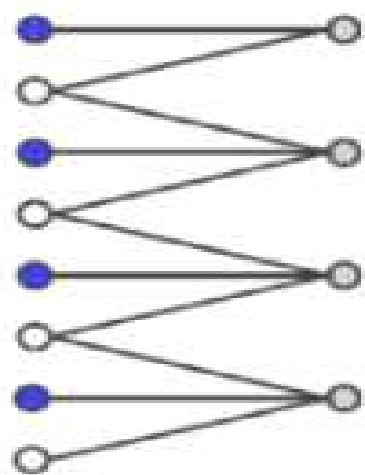


그림 4-12 CNN의 병렬 분산 구조

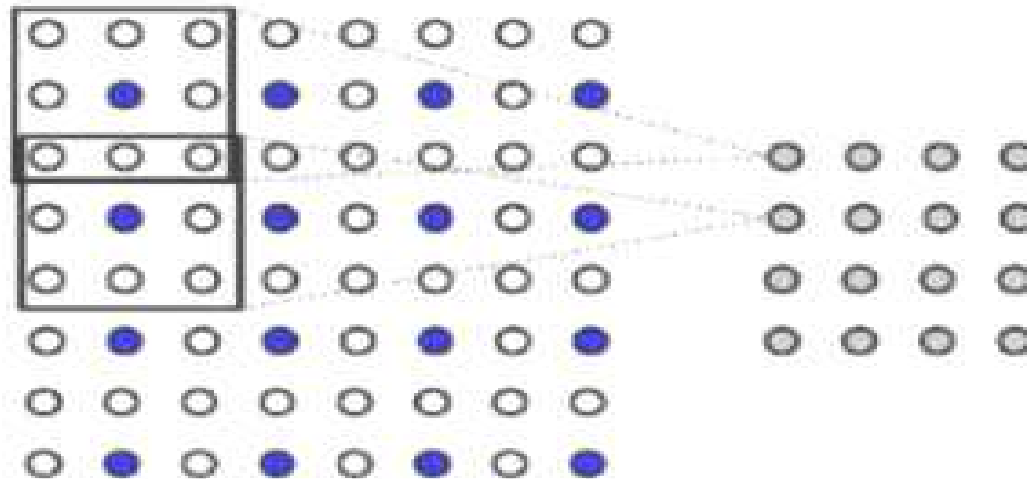
4.3.1 컨볼루션층

큰 보폭에 의한 다운샘플링(Stride)

지금까지는 모든 화소에 커널을 적용할 때 보폭을 1로 설정해서 하였다.
밑의 사진은 보폭이 2인 상황이다.
일반적으로 보폭이 k 이면, k 개 마다 하나씩 샘플링하여 커널을 적용한다.
이때, 2차원 영상의 경우 특징 맵이 $1/k^2$ 로 작아진다.



(a) 1차원 데이터(예: 음성)

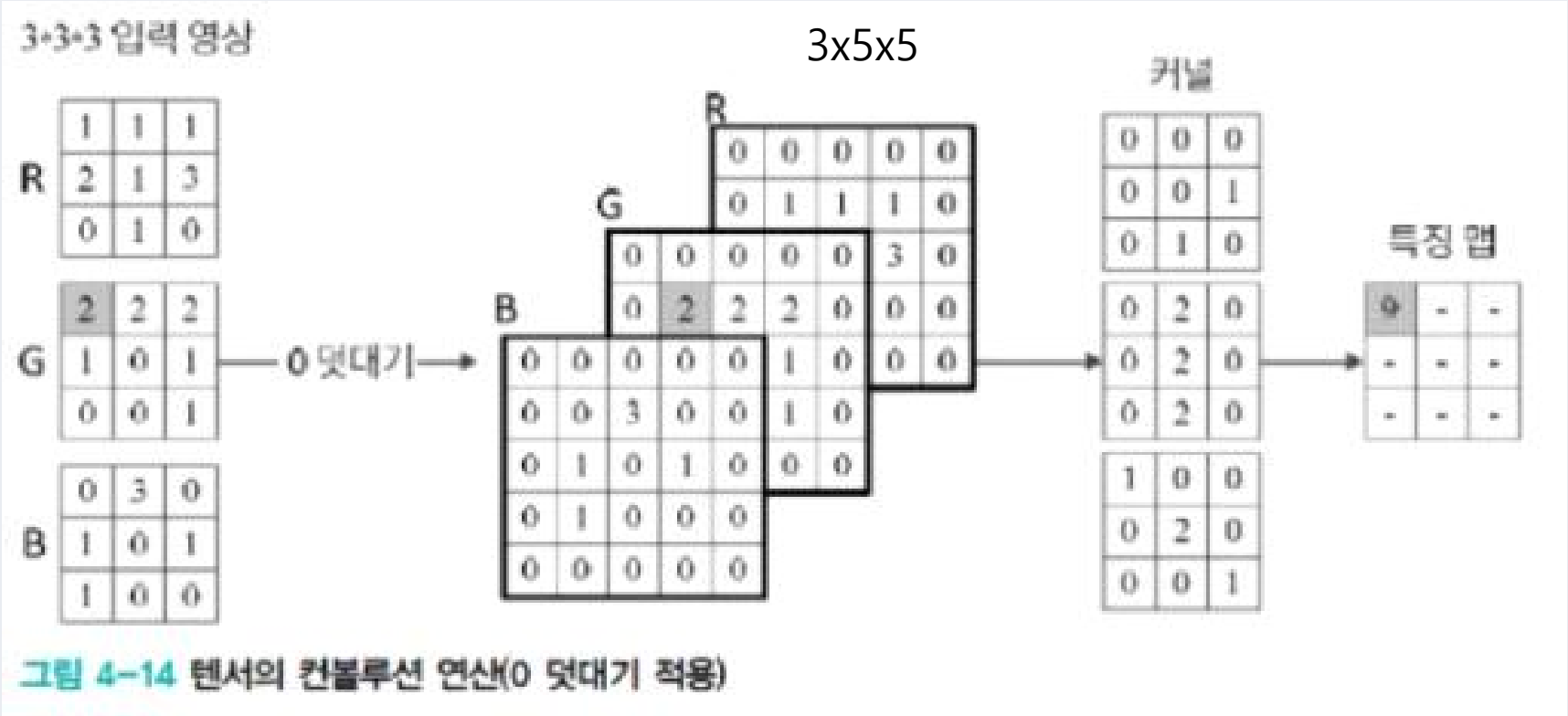


(b) 2차원 데이터(예: 영상)

그림 4-13 보폭이 2인 컨볼루션 연산

4.3.1 컨볼루션층

텐서에 적용 3차원 이상의 구조에도 적용이 가능하다.

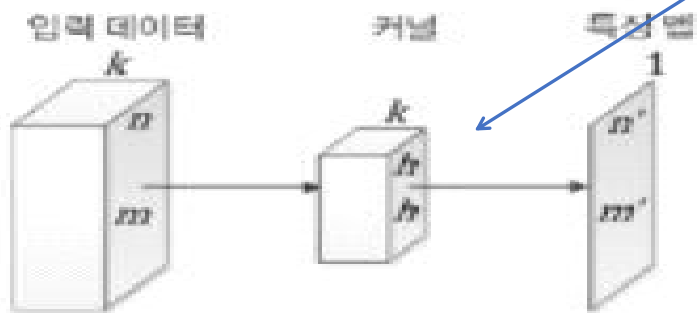
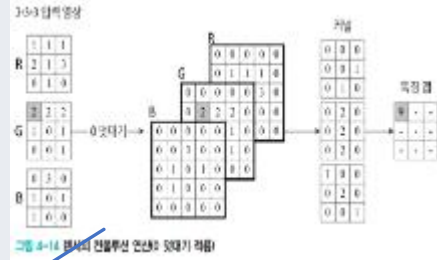


$$\underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix}}_R \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 1 & 0 \end{pmatrix}}_G \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 1 & 0 \end{pmatrix}}_B \ominus \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{c_1} \underbrace{\begin{pmatrix} 0 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 0 \end{pmatrix}}_{c_2} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{c_3} = 9$$

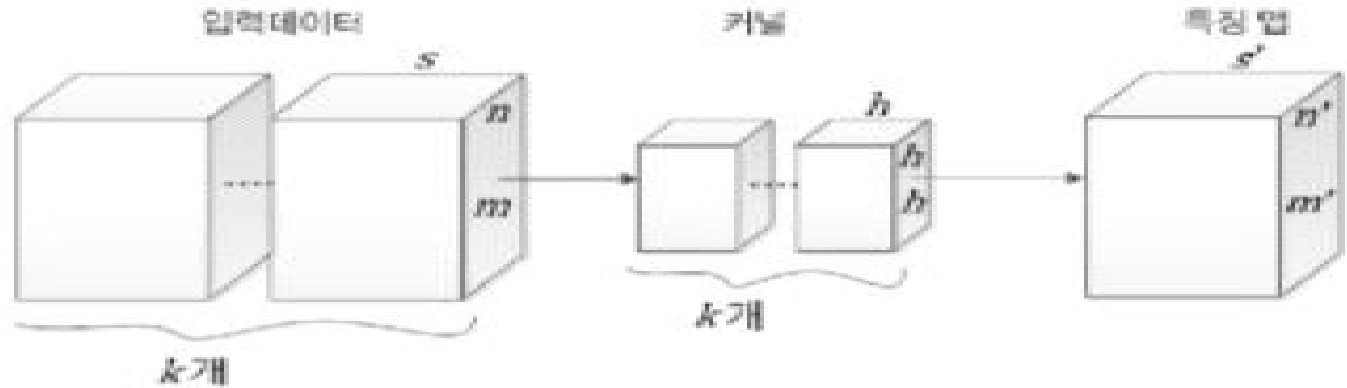
4.3.1 컨볼루션층

3차원 구조의 데이터에 적용

채널이 k 개인 3차원 격자 구조



(a) 다중채널 데이터(예: RGB 컬러 영상)



(b) 3차원 데이터(예: 동영상, MRI 뇌 영상)

그림 4-15 텐서의 컨볼루션 연산(직육면체로 표현하기)

4차원 텐서로 표현하는 데이터

예) 컬러 동영상($3 \times s \times m \times n$), MRI 뇌영상($1 \times s \times m \times n$)

4.3.2 풀링층

풀링 연산

컨볼루션 연산 + 활성화 함수 => 풀링 연산
최대 풀링, 평균 풀링, 가중치 평균 풀링, L2놈 풀링 등
보폭을 크게 하면 다운샘플링 효과가 있다.

예) 보폭 2 -> 특징 맵 크기 1/2

풀링 연산 특징

풀링은 상세 내용에서 요약통계를 추출한다.
매개변수가 없다.
특징 맵의 수를 그대로 유지한다.
작은 이동에 둔감해 물체 인식이나 영상 검색 등에 효과적이다.

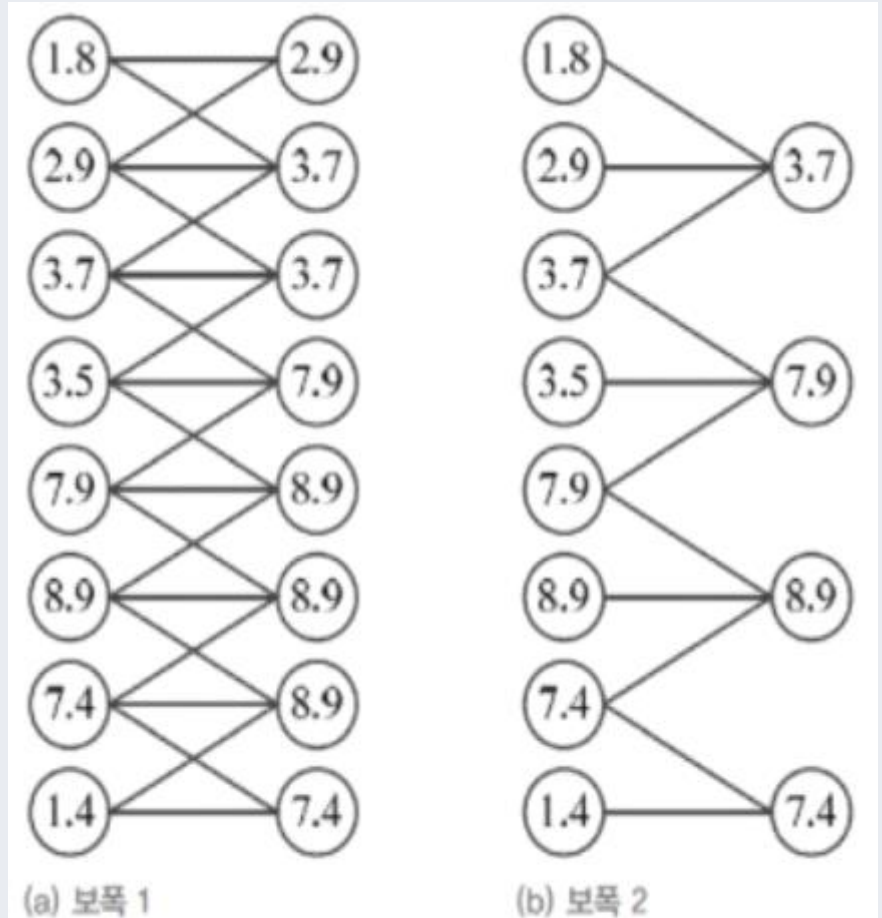


그림 4-16 최대 풀링

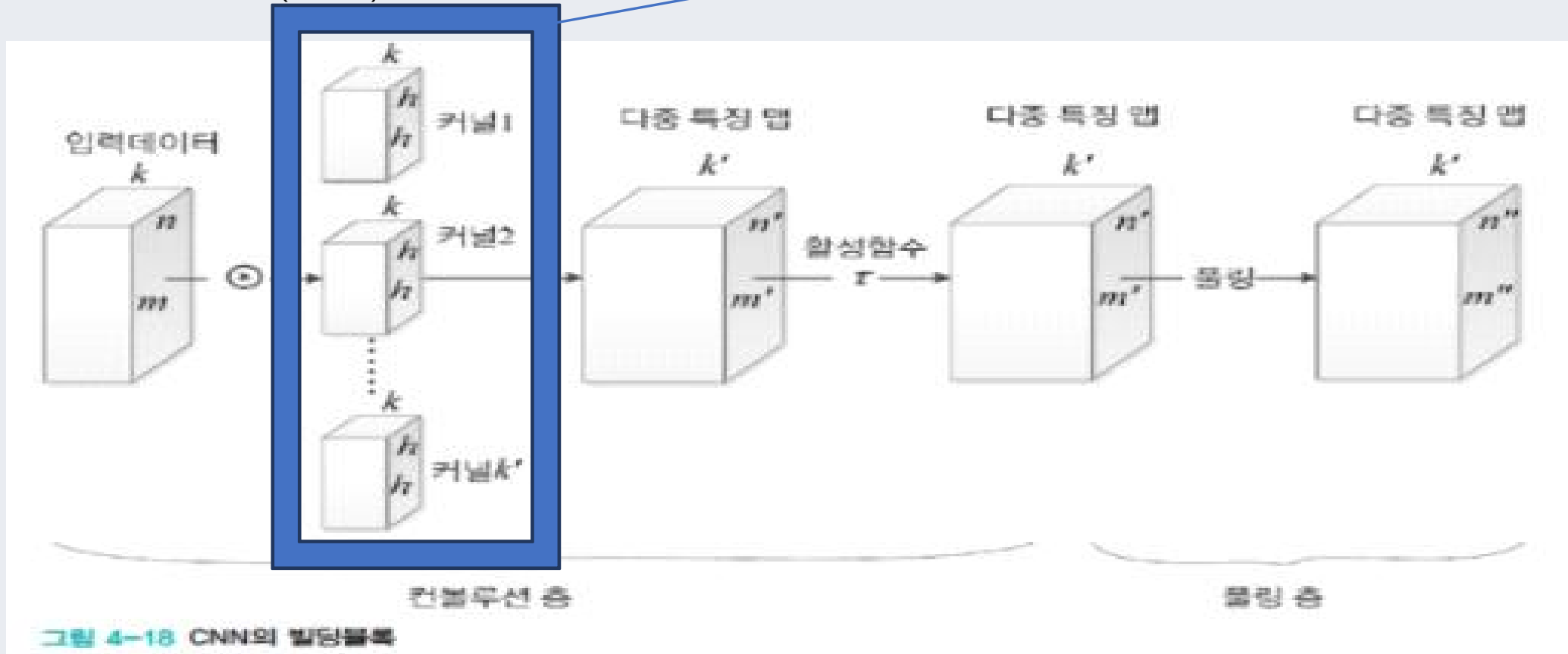
4.3.3 전체 구조

빌딩 블록

CNN은 빌딩블록을 이어 붙여 깊이 구조를 만든다.
아래의 사진은 전형적인 빌딩블록을 의미한다.

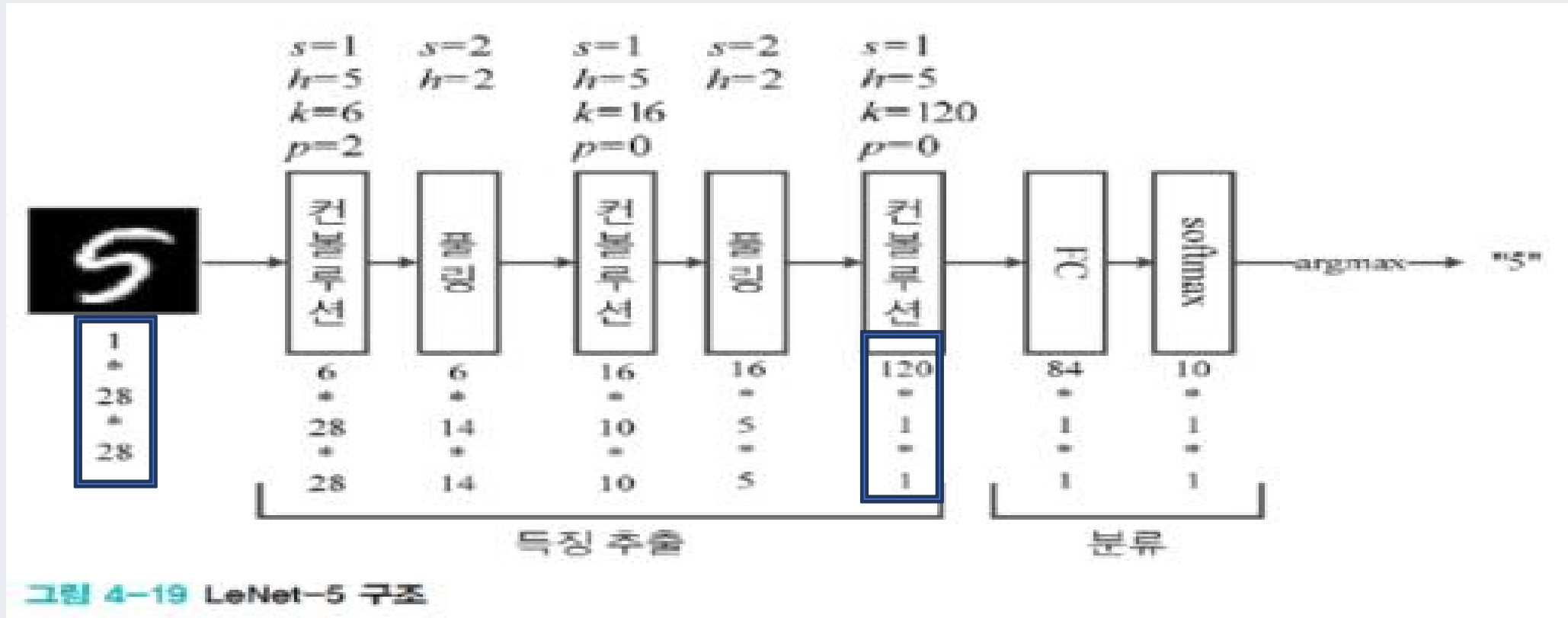
컨볼루션층 -> 활성화함수(ReLU) -> 풀링층

다중 커널을 사용하여 다중 특징 맵을 추출한다.



4.3.3 전체 구조

초창기 CNN 사례인 LeNet-5



CNN 첫 번째 성공 사례 : 필기 숫자 인식기로 수표의 금액을 읽는 자동화 시스템

4.3.3 전체 구조

가변 크기의 데이터 다루기

DMLP는 특징 벡터의 크기가 달라지면 연산이 불가능하다.

CNN은 가변 크기를 다룰 수 있는 강점이 있다.

컨볼루션층에서 보폭을 조정한다거나, 풀링층에서 커널이나 보폭을 조정하여 특징 맵 크기를 조절한다.

끝