

밑바닥부터 시작하는 딥러닝

CHAPTER 8

더 깊은 신경망

손글씨 숫자를 인식하는 심층 CNN



- 3*3의 작은 필터를 사용한 합성곱 계층(Conv)
- 활성화 함수(ReLU)
- 풀링 계층 추가해 중간 데이터의 공간 크기를 줄여나감
- 완전연결 계층 뒤에 드롭 아웃 계층 사용
 - 완전 연결 신경망이란, 인접하는 계층의 모든 뉴런과 결합되어 있는 신경망을 말한다. 입력 데이터와 가중치가 1대 1로 대응하는 것을 완전 연결(Fully-Connected)이라고 하며, 행렬의 내적(Affine)으로 표현된다.
 - 드롭 아웃은 오버피팅(over-fit)을 막기 위한 방법으로 뉴럴 네트워크가 학습중일 때, 랜덤하게 뉴런을 꺼서 학습을 방해함으로써, 학습이 학습용 데이터에 치우치는 현상을 막아준다. 일반적으로 CNN에서는 이 드롭아웃 레이어를 Fully connected network 뒤에 놓지만, 상황에 따라서는 max pooling 계층 뒤에 놓기도 한다.
- 가중치 매개변수 갱신(Adam을 사용해 최적화)
- 가중치 초기값: 'He' 초기값
(ReLU계의 활성화 함수 사용 시 He 초기화 방법이 효율적이다.)

데이터 확장(Data augmentation)

데이터 확장은 손쉬운 방법이면서도 정확도 개선에 아주 효과적이다.
입력 이미지(훈련 이미지)를 알고리즘을 동원해 '인위적'으로 확장한다.

(이미지의 개수를 늘린다.)

1. 변형

- 회전에 의한 변형(rotate)
- 이동에 의한 변형(translation)

1. 이미지 일부를 잘라낸다(crop)

2. 좌우를 뒤집는다(flip)

- 이미지의 대칭성을 고려하지 않아도 되는 경우에만 쓸 수 있다.

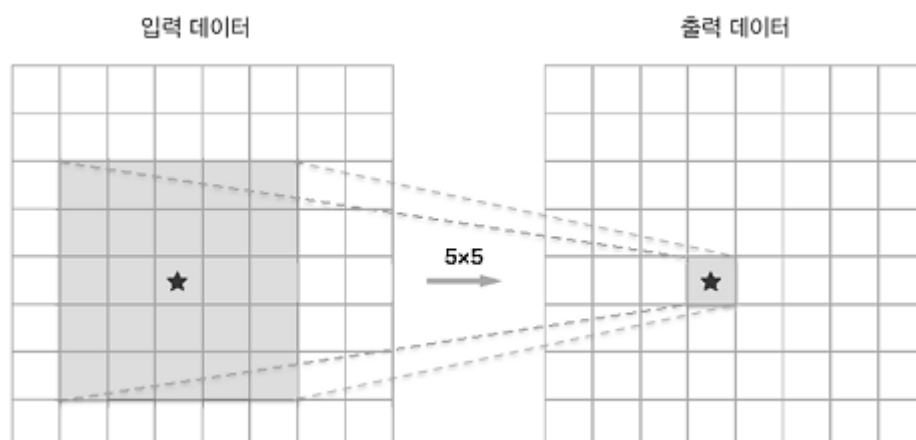
1. 크기수정(rescale)

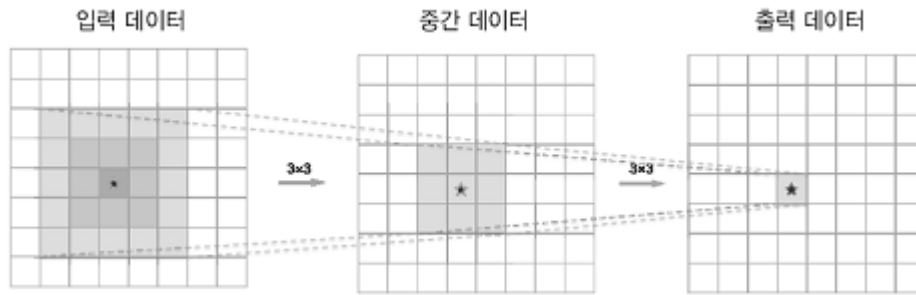
2. 밝기변화(lightning condition) 등

층을 깊게하는 이유

- ILSVRC(ImageNet Large Scale Visual Recognition Challenge) 대규모 이미지 인식 대회에서 최근 상위를 차지한 기법 대부분은 딥러닝 기반이며 그 경향은 신경망을 더 깊게 만드는 방향으로 가고 있다.(층의 깊이에 비례해 정확도가 좋아진다.)
- 층을 깊게 한 신경망은 깊지 않은 경우보다 적은 매개변수로 같은(혹은 그 이상) 수준의 표현력을 달성할 수 있다.

예) 5*5의 합성곱 연산 1회는 3*3의 합성곱 연산을 2회 수행하여 대체할 수 있다.





전자의 매개변수가 25개(5×5)인 반면 후자는 총 18개($2 \times 3 \times 3$)이며 매개변수의 수는 층을 반복할수록 적어진다.

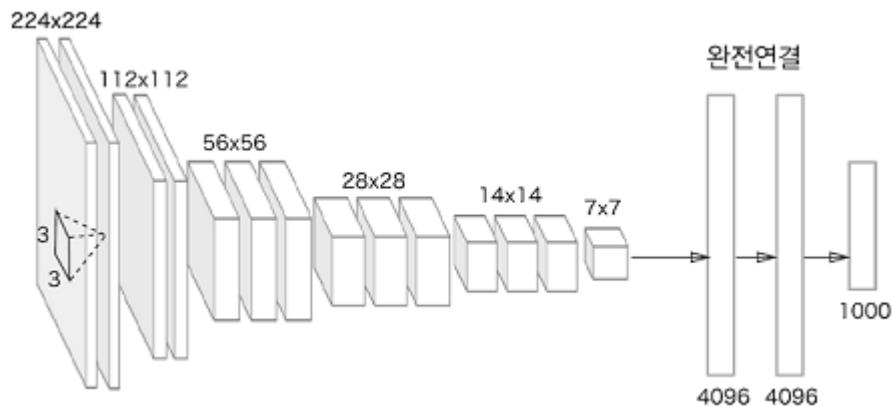
- 층을 깊게 함으로써 학습 데이터의 양을 줄여 학습을 고속으로 수행할 수 있다.(신경망을 깊게 하면 학습해야 할 문제를 계층적으로 분해할 수 있다. 즉, 각 층이 학습해야 할 문제를 더 단순한 문제로 대체할 수 있다.)

대표적인 신경망

VGG

합성곱 계층과 풀링 계층으로 구성되는 기본적인 CNN

비중 있는 층(합성곱 계층, 완전연결 계층)을 모두 16층(혹은 19층)으로 심화함.



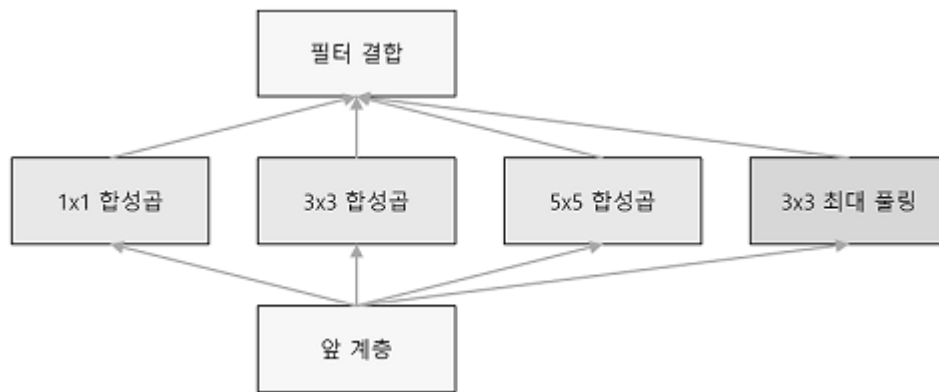
3*3의 작은 필터를 사용한 합성곱 계층을 연속으로 거친다.

합성곱 계층을 2~4회 연속으로 풀링 계층을 두어 크기를 절반으로 줄이는 처리를 반복하고 마지막에는 완전연결 계층을 통과시켜 결과를 출력한다.

GoogLeNet

기본적으로는 기존의 CNN과 다르지 않지만 세로 방향 깊이 뿐 아니라 가로 방향도 깊다는 점이 특징이다.

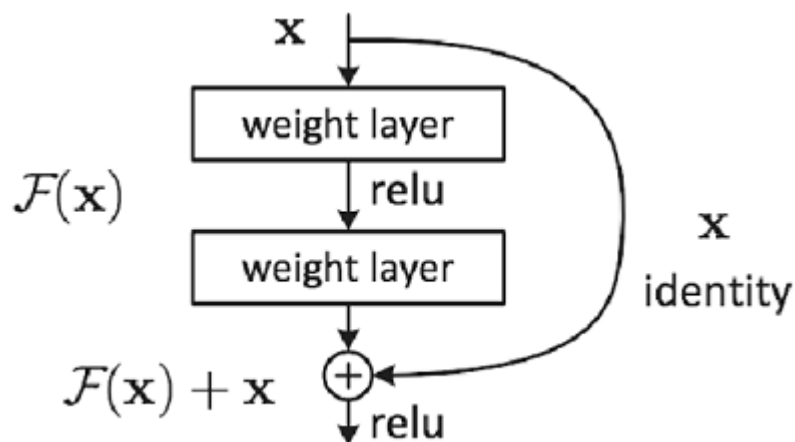
GoogLeNet에는 가로 방향에 폭이 있고 이를 **인셉션 구조**라 한다.



인셉션 구조는 크기가 다른 필터(와 풀링)을 여러 개 적용하여 그 결과를 결합하는데 GoogLeNet은 이 인셉션 구조를 하나의 빌딩 블록(구성요소)으로 사용한다.

또한 11크기의 필터를 사용한 합성곱 계층을 많은 곳에서 사용한다. (11의 합성곱 연산은 채널 쪽으로 크기를 줄이는 것으로 매개변수 제거와 고속 처리에 기여한다.)

ResNet(Residual Network)



- **스킵연결(skip connection)**: 층의 깊이에 비례해 성능을 향상시킨다. 입력 데이터를 합성곱 계층(weight layer)을 건너뛰어 출력에 바로 더하는 구조역전파 때 스킵 연결이 신호 감쇠를 막아주기 때문에 층이 깊어져도 학습을 효율적으로 할 수 있다.

딥러닝 가속화

1. 계산 능력

- 딥러닝 프레임워크 대부분은 GPU(Graphics Processing Unit)를 활용해 대량의 연산을 고속으로 처리한다.
- GPU는 병렬 수치 연산을 고속으로 처리할 수 있다.

- 딥러닝 계산을 더욱 고속화하고자 다수의 GPU와 기기로 계산을 분산한다.(딥러닝 학습의 수평 확장(scale out) 즉, 분산학습)
- 다수의 GPU와 컴퓨터를 이용한 분산 학습을 지원한 딥러닝 프레임워크의 예 구글-텐서플로 마이크로소프트-CNTK(Computational Network Toolkit)

1. 메모리 용량

대량의 가중치 매개변수와 중간 데이터를 메모리에 저장해야 한다.

2. 버스 대역폭

GPU(혹은 CPU)의 버스를 흐르는 데이터가 많아져 한계를 넘어서면 병목이 된다.

=> 2와 3의 문제점을 해결하기 위해서는 네트워크로 주고받는 데이터의 비트 수를 최소로 만드는 것이 바람직하다.

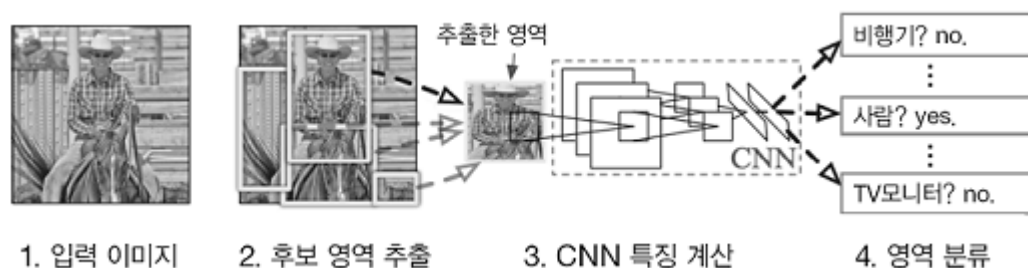
- 신경망의 견고성: 딥러닝은 높은 수치 정밀도(수치를 몇 비트로 표현하느냐)를 요구하지 않는다.

딥러닝의 활용

사물 검출

이미지 속에 담긴 사물의 위치와 종류(클래스)를 알아내는 기술

- R-CNN(Regions with Convolutional Neural Network)



R-CNN 에서 주목할 곳은 2. 후보 영역 추출(사물처럼 보이는 물체를 찾아 처리) 과 3. CNN 특징 계산이다. 먼저 사물이 위치한 영역을 (어떤 방법으로) 찾아내고 추출한 각 영역에 CNN을 적용하여 클래스를 분류한다.

- R-CNN논문에서는 Selective Search 기법을 사용했다.

분할(segmentation)

이미지를 픽셀 수준에서 분류한다.

픽셀 단위로 객체마다 채색된 지도(supervised) 데이터를 사용해 학습한다.



- FCN(Fully Convolutional Network)합성곱 계층만으로 구성된 네트워크.단 한 번의 forward 처리로 모든 픽셀의 클래스를 분류해준다.공간 볼륨을 유지한 채 마지막 출력 까지 처리할 수 있다.마지막에 공간 크기를 확대하는 처리를 도입했다.(이를 통해 줄어 든 중간 데이터를 입력 이미지와 같은 크기까지 단번에 확대할 수 있다. **이중 선형 보간 (bilinear interpolation)**에 의한 선형 확대)

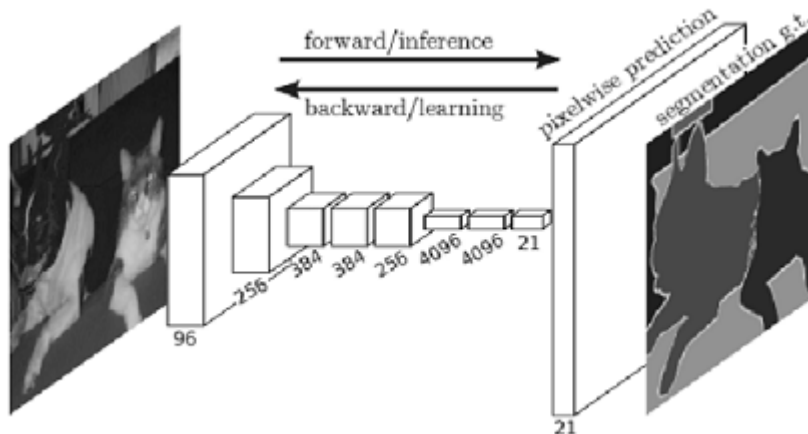


사진 캡션 생성

- NIC(Neural Image Caption): 심층 CNN과 자연어를 다루는 순환신경망(Recurrent Neural Network)으로 구성된다.
- RNN: 순환적 관계를 갖는 신경망으로 자연어나 시계열 데이터 등의 연속된 데이터를 다룰 때 많이 활용한다.
- 멀티모달 처리(multimodal processing): 사진이나 자연어와 같은 여러 종류의 정보를 조합하고 처리하는 것

딥러닝의 미래

Deep Q-Network(강화학습)

컴퓨터도 사람과 같이 시행착오 과정에서 스스로 학습하게 하려는 분야가 있는데 이를 강화 학습이라 한다. 강화학습에서는 에이전트라는 것이 환경에 맞게 행동을 선택하고 그 행동에 의해서 환경이 변한다는 게 기본적인 틀이다. 강화학습의 목적은 더 나은 보상을 받는 쪽으로 에이전트의 행동 지침을 바로잡는 것이다.

그림 8-26 강화학습의 기본 틀 : 에이전트는 더 좋은 보상을 받기 위해 스스로 학습한다.



강화학습에서 주의할 점은 보상이 정해진 것이 아니라 예상 보상이라는 점이다. 딥러닝을 사용한 강화학습 중 Deep Q-Network라는 방법이 있다. 이는 Q학습이라는 강화학습 알고리즘을 기초로 한다. Q학습에서는 최적 행동 가치 함수로 최적인 행동을 정한다. 이 함수를 딥러닝(CNN)으로 비슷하게 흉내 내어 사용하는 것이 DQN이다.