

9차 일석이조 조별보고서	
작성일: 2023년 11월 1일	작성자: 김동규
조 모임 일시: 2023년 10월 31일 8교시	모임장소: 구글미트
참석자: 이준용, 유정훈, 김동규, 이학빈, 탁성재	조원: 이준용, 유정훈, 김동규, 이학빈, 탁성재
구 분	내 용
<p>학습 범위와 내용</p> <p>(조별 모임 전에 조장이 공지)</p>	<p>1. 컨볼루션 신경망 사례연구</p> <p>4.2.1 AlexNet</p> <p>4.2.2 VGGNet</p> <p>4.2.3 GoogLeNet</p> <p>4.2.4 ResNet</p>
<p>논의 내용</p> <p>(모임 전 공지된 개별 학습 범위에서 이해된 것과 못한 것들)</p>	<p>Q(1) Google Net에 대해 궁금하여 팀원들과 함께 www.axiv.net 에서 논문을 검색하여 더 깊게 탐구해보았다.</p> <p>A(1) GoogLeNet은 2014년 ILSVRC대회에서 1위를 차지한 모델이고, 가장 큰 변화는 네트워크 층의 깊이이다.</p>

1409.4842v1 [cs.CV] 17 Sep 2014

Going deeper with convolutions

Christian Szegedy
Google Inc.

Wei Liu
University of North Carolina, Chapel Hill

Yangqing Jia
Google Inc.

Pierre Sermanet
Google Inc.

Scott Reed
University of Michigan

Dragomir Anguelov
Google Inc.

Dimitru Erhan
Google Inc.

Vincent Vanhoucke
Google Inc.

Andrew Rabinovich
Google Inc.

Abstract

We propose a deep convolutional neural network architecture codenamed Inception, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

CNN의 성능을 향상하는 가장 직접적인 방법은 네트워크의 크기를 늘리는 방법이다. 2013년까지는 CNN 네트워크 깊이가 10개 미만이었으나, 2014년 GoogLeNet과 VGGNet이 각각 22 layer, 19 layer로 2배 이상 커졌다. 이로 인해 Top-5 에러율도 각각 6.7%, 7.3% 낮아졌다.

하지만, 네트워크가 깊어진다고 무조건 다 좋은 건 아니다. 그에 따라 파라미터 수가 증가하게 되고 그 결과로 Overfitting문제가 발생하거나 연산량이 급격히 늘어날 수 있다.

위에 설명한 AlexNet만 보더라도 파라미터 수가 6천만 개이고 GPU 2개를 사용해서 네트워크만 깊게 만들면 학습시간이 늘어나 일주일 이상 학습을 시켰다고 한다.

따라서, 이러한 문제를 해결하기 위해서는 네트워크 구조의 변화가 필요하다고 생각했고 Google에서 Inception이라는 모듈로 구성된 GoogLeNet으로 이를 해결했다.

Inception 모듈은 싱가포르 국립대학의 Min Lin이 2013년에 발표한 “Network in Network”의 구조를 더 발전시킨 형태이다.
1 X 1 Convolution으로 채널의 수를 줄여서 파라미터 수를 줄이는 것이 Network in Network의 핵심 내용이다.
초기에는 위 그림의 좌측처럼 1x1/3x3/5x5 Convolution, 3x3 Max pooling을 나란히 놓는 구조를 만들었다. 하지만 3x3/5x5에서도 파라미터를 줄이기 위해 figure 2 (b)처럼 사용하였고 Bottleneck 구조라고도 불린다.
GoogleNet에서는 이를 반영해서 네트워크를 깊게 만들면서도 연산량의 수가 늘지 않는 Inception이라는 모듈을 만들었다. 이로 인해 AlexNet 보다 12배나 파라미터수가 감소하였다 GoogleNet은 총 9개의 Inception으로 구성되어 있다고 한다.

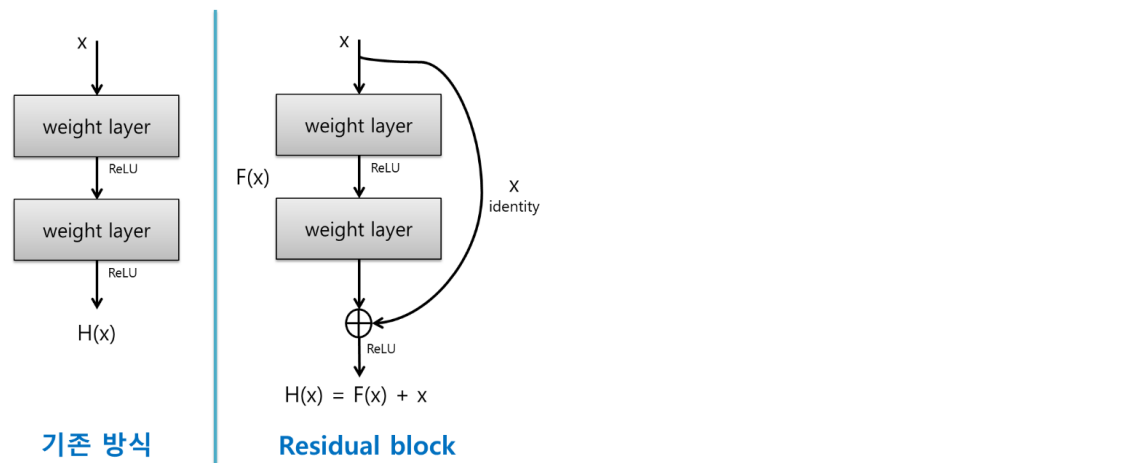
Gradient descent (경사 소실 문제)

하지만 GoogleNet은 층의 수가 많기 때문에 파라미터가 앞으로 갈수록 업데이트가 되지 않는 Gradient descent 문제가 발생할 가능성이 높다.
이문제를 해결하기 위해 추가 분류기를 사용한다.
아키텍처 중간에 일부 추가 분류기 분기를 사용했으며 이러한 분기는 훈련 중에만 사용된다. 추가 분류기를 통해 종속변수 예측치를 도출한다. 그리고 이 예측치를 통해서 새로운 비용 함수를 계산해서 앞쪽 층의 파라미터 업데이트를 실시한다.

Q2 Resnet에 대하여

A2

VGGNet, GoogleNet과 같이 많은 은닉층을 통한 계산으로 ‘깊은’ 구조를 추구하는 경우, 일정 층수가 넘어가면 성능 향상에 효율이 급감하는 지점이 존재하였다. 따라서 ResNet은 잔류학습이라는 개념을 적용해 성능 저하를 피했다.
잔류학습은 현재 은닉층의 결과를 다음 은닉층으로 넘겨 계산하는 형태가 아닌, 다음 은닉층으로는 일부의 결과를 넘기고 계산이 끝나면 현재 은닉층의 결과값을 더해 각 층의 계산을 줄이는 형태의 아이디어를 말한다.



잔류학습을 통하면 각각의 은닉층은 추가로 계산할 부분만 계산하면 되므로 계산량이 줄게 되고, 레이어가 깊어질수록 추가되는 정보가 줄어들게 되므로 최종적으로는 최소값으로 수렴하게 된다.

잔류학습은 이전 계산값 + 추가된 일부 값에 대한 새로운 계산값이므로 수식을 취하면

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i)$$

이때 각 값의 변화율이 곧 학습률이므로 각각의 값에 편미분을 취하는 경우, 바로 직전 레이어의 변화율*(1+ 추가된 값들에 대한 변화율)의 형태를 가지게 된다.

$$\frac{\partial \epsilon}{\partial x_l} = \frac{\partial \epsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \epsilon}{\partial x_L} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, W_i) \right)$$

이때 변화율은 -1을 가질 수 없으므로 기울기 소멸이 발생하지 않으므로, 학습률을 계속 높게

유지할 수 있다.

Q3 효과적인 모델 선택에 도움이 될 수 있게 다양한 컨볼루션 신경망 아키텍처 사례들을 조사하고 비교한 결과입니다.

A3

1. VGGNet:

장점:

간결한 디자인으로 이해하기 쉽다.

깊이와 필터 크기가 일정하므로 모델 구조를 쉽게 조절할 수 있다.

단점:

매우 깊은 네트워크로 인해 많은 파라미터가 필요하며, 훈련 및 실행에 시간이 오래 걸릴 수 있다.

메모리 사용량이 크다.

2. GoogLeNet:

장점:

Inception 모듈로 인해 파라미터 공유 및 모델 크기 감소가 가능하다.

계층 간의 다양한 필터 크기 및 유형을 사용하여 효과적인 특징 추출이 가능함

단점:

모델 구조가 복잡하여 이해하기 어려울 수 있다.

모델이 깊고 넓기 때문에 훈련에 높은 계산 리소스가 필요하다.

3. ResNet:

장점:

깊은 네트워크에서도 그래디언트 소실 문제를 해결하고 훈련을 용이하게 한다.

잔여 연결을 통해 효율적인 정보 전달과 그래디언트 흐름을 갖는다.

단점:

모델이 깊어지면 메모리 사용량이 증가하고 계산 비용이 높아질 수 있다.

4. DenseNet:

장점:

밀집 연결 구조로 인해 모든 레이어가 정보를 공유하며 그래디언트 흐름을 최적화한다.

모델이 깊어질수록 파라미터 효율성을 제공한다.

단점:

모델 크기가 크고 메모리 사용량이 높을 수 있다.

5. SqueezeNet:

장점:

작은 모델 크기와 적은 파라미터로 메모리 사용량을 줄이고 모델을 배포하기 용이하다.

계산 및 메모리 효율적인 구조를 제공한다.

단점:

모델의 성능은 일부 효율성을 희생함으로써 달성되므로 더 큰 모델보다 정확성이 낮을 수 있다.

Q4 VGGNet 에 대하여 알아보았습니다.

A4

VGGNet (Visual Geometry Group Network)는 이미지 분류를 위한 딥러닝 아키텍처 중 하나로, 2014년에 개발된 모델입니다. VGGNet은 컴퓨터 비전 작업을 위해 특별히 설계된 합성곱 신경망(Convolutional Neural Network, CNN)의 일종으로, 이미지 인식 및 분류 과제에서 우수한 성능을 보여 줍니다. 특징으로는

1. 깊은 구조: 깊은 신경망 아키텍처로, 16 또는 19개의 층(레이어)로 구성되어 있습니다. 이러한 깊은 구조는 많은 합성곱 레이어와 풀링 레이어로 이루어져 있으며, 레이어의 개수와 복잡성이 높아질수록 모델의 표현력과 성능이 향상됩니다.
2. 작은 커널 크기: 3x3 크기의 작은 커널(합성곱 필터)을 사용하며, 이 작은 커널을 여러 번 적용하여 이미지의 특징을 추출합니다. 이로써 더 깊은 네트워크를 만들면서도 파라미터 수를 줄이고 성능을 향상시킬 수 있습니다.

	<p>3. 맥스 풀링: 맥스 풀링 레이어를 사용하여 이미지 해상도를 줄이고 중요한 특징을 강조합니다. 이로써 네트워크가 더 깊어질 수 있고, 특징 추출과 학습이 효과적으로 이루어집니다.</p> <p>4. 다중 레이어: 다양한 깊이의 레이어를 가지고 있으며, 이러한 다중 레이어는 다양한 추상화 수준의 이미지 특징을 포착할 수 있도록 도와줍니다.</p> <p>5. 사전 훈련된 모델: VGGNet은 대규모 이미지 데이터셋인 ImageNet에서 사전 훈련된 가중치를 가진 모델을 제공하여, 이를 다른 컴퓨터 비전 작업에 전이학습(transfer learning)으로 활용할 수 있습니다.</p> <p>VGGNet은 단순하면서도 효과적인 아키텍처로 이미지 분류, 객체 검출, 시맨틱 분할 등 다양한 컴퓨터 비전 작업에서 사용될 수 있습니다. 그러나 매우 깊은 네트워크로 인해 학습 및 실행에 상당한 계산 리소스가 필요하며, 더 최신의 아키텍처에 비해 더 많은 파라미터를 가질 수 있으므로, 리소스가 제한된 환경에서는 다른 경량 모델도 고려해야 합니다.</p>
질문 내용	없습니다
기타	없습니다