

Cs231n Lecture 9 Summary

CNN Architectures

- 1. Review & Case study**
- 2. Comparing complexity**
- 3. Other Architectures to know...**

1. Review

1) LeNet-5

LeNet 은 산업에 성공적으로 적용된 최초의 ConvNet 이다. 이미지를 입력으로 받아서 stride 가 1 인 5x5 필터를 거치고 Conv layer 와 pooling layer 를 거친다.

2) AlexNet

AlexNet 은 2012 년에 나온 모델이다. 최초의 Large scale CNN 이라고 할 수 있다. AlexNet 의 Architecture 는 다음과 같다. LeNet 과 유사하지만 layer 의 개수가 더 많아진 것이 특징이다.

Architecture:
CONV1
MAX POOL1
NORM1
CONV2
MAX POOL2
NORM2
CONV3
CONV4
CONV5
Max POOL3
FC6
FC7
FC8

AlexNet 의 ImageNet 으로 학습시키는 경우 입력의 크기는 227 x 227 x 3 이다. 첫 번째 layer 는 11 x 11 크기의 필터가 stride = 4 이기 때문에 총 96 개가 존재한다.

Q1. What is the output volume size? (출력 사이즈)

A1. 출력값의 차원 = (전체 이미지 크기-필터 크기) / stride + 1

따라서 $(227-11)/5 = 55$ 이다. 즉 55 x 55 x 96

Q2. What is the total number of parameters in this layer?

A2. $11 \times 11 \times 3 \times 96 = 35k$ 개

두 번째 layer 는 pooling layer 이다. 여기에는 stride = 2 인 3 x 3 크기의 필터가 있다.

Q3. What is the output volume size?

A3. $(55-3)/3 = 17$, depth 는 똑같으므로 17 x 17 x 96

Q4. What is the total number of parameters in this layer?

A4. 0 (zero). Pooling layer 에는 파라미터가 존재하지 않는다.

->why?

파라미터는 우리가 학습시키는 가중치를 뜻한다. Conv layer 에는 학습할 수 있는 가중치가 있지만 pooling layer 는 단지 특정 지역에서 큰 값을 뽑아내는 역할만을 수행한다. 따라서 가중치가 존재하지 않아 파라미터가 없다.

3) ZFNet

ZFNet 은 AlexNet 의 하이퍼 파라미터를 개선한 모델이다. AlexNet 과 레이어 개수가 같고 구조도 같다. 다만 stride size 나 filter 의 개수를 조절해 Error rate 를 개선하였다.

4) VGGNet

VGGNet 은 더 깊어진 네트워크, 더 작은 필터라는 말로 요약할 수 있다. 항상 3x3 필터만 사용하고 주기적으로 pooling 을 수행하여 네트워크를 구성한다.

필터의 크기가 작은 이유는 depth 를 더 키울 수 있기 때문이다.

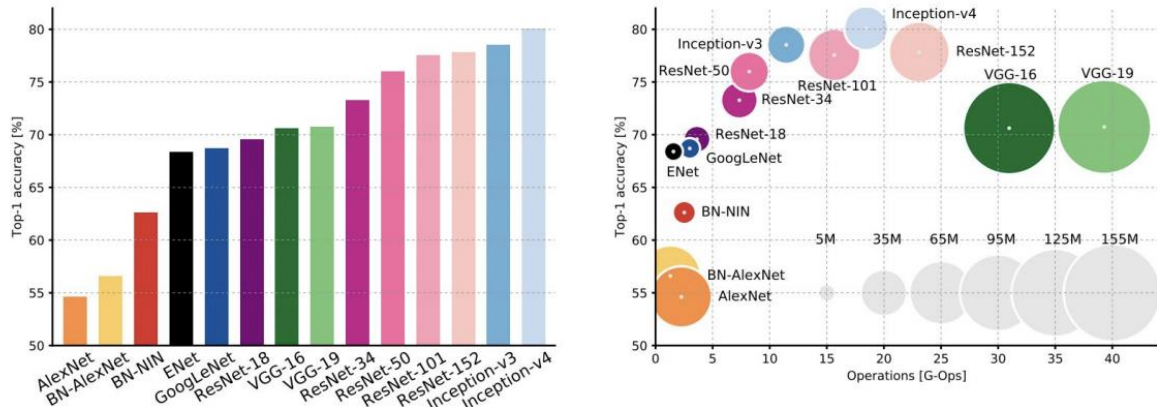
5) GoogLeNet

GoogLeNet 은 2014 년 Classification Challenge 에서 우승한 모델이다. 총 22 개의 layer 를 가지고 있다. 또한 inception module 을 사용하고 FC-layer 가 없다. 파라미터를 줄이기 위해서이다. 전체 파라미터의 수는 약 5 백만개인데, AlexNet 보다 적은 것을 알 수 있다.

6) ResNet

ResNet 은 엄청나게 (Very deep) 깊다. 모델을 깊게 쌓기 위해 Residual connection 이라는 방법을 사용한다. 일반 CNN 을 더 깊게 쌓게 된다고 무조건 성능이 좋아지는 것은 아니다. 56 layer 와 20 layer 의 test error 를 비교해 보면 56 layer 가 더 안 좋은 것을 알 수 있다. 따라서 ResNet 의 저자들은 모델이 깊어질수록 optimization 에 문제가 생긴다는 가설을 세웠다. 그래서 더 얇은 모델의 가중치를 깊은 모델의 일부 레이어에 복사하고, 나머지 layer 는 identity mapping 을 하였다. 그리고 이러한 layer 들이 모델에 잘 적용되기 위해 사용한 방법이 바로 Residual mapping 인데, 레이어가 $H(x)$ 를 학습하지 않고 $H(x)-x$ 를 학습한다. 즉 입력 x 에 대한 변화량(delta)만 학습하는 것이다.

2. Comparing complexity



An Analysis of Deep Neural Network Models for Practical Applications, 2017.

이 사진은 모델별 complexity를 비교해 놓은 그래프이다. GoogLeNet-Inception은 버전별로 여러 가지가 있는데 v4가 가장 좋은 모델인 것을 알 수 있다. 오른쪽은 계산 복잡도가 추가된 그래프이고, 원의 크기는 메모리 사용량이다.

가장 효율적인 네트워크는 GoogLeNet이고, VGGNet은 메모리도 크고 계산량도 많지만 성능이 좋다. 초기 모델인 AlexNet은 accuracy가 매우 낮고 메모리 사용도 비효율적인 것을 알 수 있다.

3. Other architectures to know...

1) Network in Network(NiN)

기본적인 아이디어는 MLP conv layer이다. 네트워크 안에 작은 네트워크를 삽입하는 것인데, 각 conv layer 안에 MLP(Multi-Layer Perceptron)를 쌓는다. FC-layer 몇 개를 쌓는 것과 같다. 단순히 conv filter만 사용하지 않고 조금 더 복잡한 계층을 만들어보자는 아이디어이다.

2) Identity Mapping in Deep Residual Networks

Direct path 를 늘려서 정보들이 앞으로 더욱 잘 전달되고 backpropagation 도 더 잘 될 수 있게 개선하였다.

3) Wide Residual Networks

기존의 ResNet 은 깊게 쌓는 것에 중점을 두었지만 사실은 depth 보다 residual 이 더 중요하다고 주장한 연구이다. Residual block 을 더 넓게 만들어 conv layer 필터를 더 많이 추가하였다. 기존의 ResNet 에는 블록당 F 개의 필터만 있었다면 대신 $F * K$ 개의 필터로 구성하였다. 각 레이어를 넓게 구성하였더니 레이어 수가 적어도 기존의 ResNet 보다 성능이 좋다는 것을 입증하였다.

4) RexNeXt

이 논문에서도 residual block 의 width 에 집중해 filter 의 개수를 늘렸다. 각 Residual block 에 multiple parallel pathway 를 추가하였다.

5) Stochastic Depth

네트워크가 깊어질수록 Vanishing gradient 문제가 발생한다. 따라서 train time 에 레이어의 일부를 제거하고 일부 네트워크를 골라 identity connection 으로 만든다. 이렇게 하면 dropout 과 유사하게 되어 gradient 를 더 잘 전달할 수 있다.

6) FractalNet

Residual connection 이 전혀 없고, shallow/deep network 의 정보 모두를 잘 전달하는 것이 중요하다고 생각하여 둘의 경로를 출력에 모두 연결하여 fractal 한 모양의 네트워크를 만든다. 하지만 train time 에는 일부 경로만 이용하여 train 을 진행한다.

7) DenseNet

DenseNet 에는 Dense block 이라는 것이 있다. 한 레이어가 그 레이어 하위의 모든 레이어와 연결되어 있다. 입력 이미지가 모든 layer 의 입력으로 들어가고, 모든 layer 의 출력이 각 layer 의 출력과 concat 되는 것이다. 이러한 dense connection 이 vanishing gradient 문제를 완화시킬 수 있다고 주장한다.

8) SqueezeNet

Fire module 을 도입하였다. Squeeze layer 는 1x1 필터로 구성되고, 이 출력값이 1x1, 3x3 필터들로 구성되는 expand layer 의 입력이 된다. SqueezeNet 은 ImageNet 에서 AlexNet 만큼의 accuracy 를 보이지만 파라미터는 50 배나 더 적다.