논문리뷰: Deep Residual Learning for Image Recognition

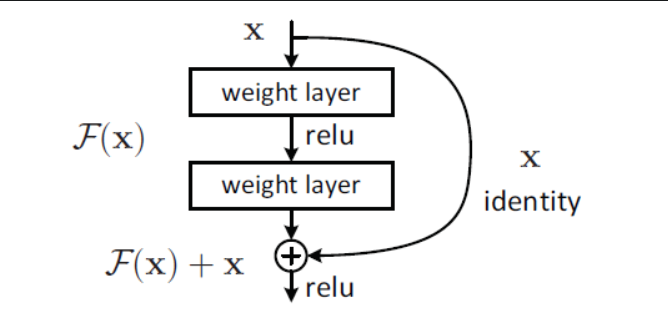
14기 김상현

1. Introduction

깊은 합성곱 신경망은 좋은 이미지 분류 능력을 보인다. 이에 연구진은 ‘더 많은 층을 쌓은 망이 학습이 잘 될까?’ 라는 의문을 갖게 되었다. 여기에는 vanishing/exploding gradient라는 문제가 있었지만, 정규화 기법을 통해 해결했다. 하지만 depth가 증가하면서 degradation problem이 발생했다. 이는 overfitting으로 인한 것이 정확도가 포화상태에 도달하게 되면서 성능이 급격하게 떨어지는 현상을 말한다.

기존에 학습된 shallower 구조에 identity mapping인 layer만 추가하여 deeper 구조를 만든 constructed solution 이상의 성능을 갖는 solution을 찾을 수 없는 것을 실험을 통해 확인했다.

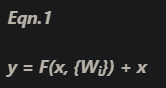
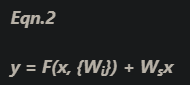
이 논문에서는 degradation problem을 다루기 위해 ‘deep residual learning framework’를 도입했다. 이 방법은 few stacked layer를 underlying mapping을 직접 학습하지 않고, residual mapping에 맞추어 학습하도록 한다. 여기서 original mapping H(x) = F(x)+x. F(x)=0이 되게 하는 residual mapping을 optimize하는 것이 더 쉬울 것이라는 가정을 한다. F(x)+x를 구현하기 위해 shortcut connection을 사용한다. Shortcut connection은 하나 이상의 layer를 건너뛴다.



이와 같은 identity shortcut connection은 단순히 더하기를 하므로 연산과정에서 parameter나 computational complexity가 추가되지 않는다. 또한 기존의 SGD에 따른 역전파로 end-to-end 학습이 가능하다.

1. Deep Residual Learning

H(x)를 점근적으로 근사화 시키는 것은 F(x) = H(x)-x인 residual function을 근사화 시키는 것과 동치로 생각할 수 있고 F(x)+x로 reformulation 하는 것은 degradation problem을 해결하기 위한 것이다. 이 reformulation은 문제에 precondition을 제공하는 효과를 준다. 만약 optimal function이 zero mapping보다 identity mapping에 더 가깝다면, solver가 identity mapping을 참조하여 작은 변화 F(x)를 학습하는 것이 새로운 것을 학습하는 것보다 더 쉬울 것이다. 또한 identity mapping은 합리적인 preconditioning을 제공한다.

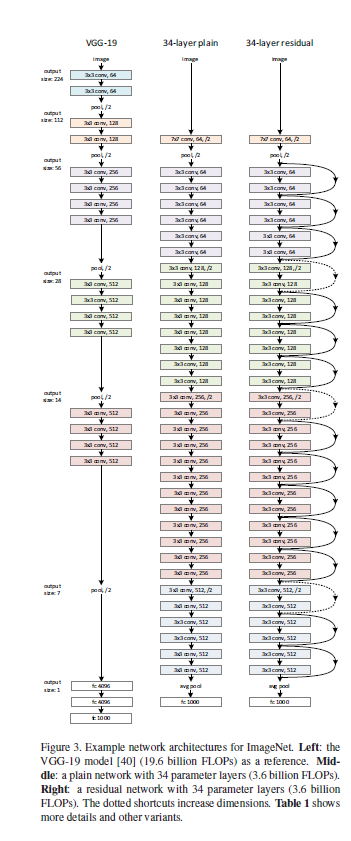
Shortcut을 이용한 identity mapping을 위해 위의 방정식을 이용한다. 방정식 1은 F(x)+x 즉 shortcut connection을 보여주고 방정식 2는 F(x)+x를 할 때 둘의 차원이 맞지 않는 경우 W\_s를 x연산해서 dimension matching을 수행한다. 여기서 W\_s는 dimension matching용도로만 사용된다.

1. Network Architectures

다양한 Plain/Residual network에 대해 ImageNet dataset을 이용해 실험을 했다

Plain – VGG net을 기반으로 만들었다. 두 가지 규칙에 따라 구성했다. 1)동일한 output feature map size에 대해 layer는 동일한 수의 filter를 갖는다. 2)feature map size가 절반인 경우 layer당의 time complexity를 보전하기 위해 filter의 수를 2배로 한다. 네트워크 마지막에는 GAP(Global Average Pooling)와 activation이 ‘softmax’인 1000-way FC layer로 구성된다. 이 네트워크는 VGG-19보다 더 깊은 층을 갖지만 VGG-19의 18%의 연산량(FLOPs기준)을 갖는다.

Residual – plain network를 기반으로 shortcut connection을 삽입하여 만든다 input과 output의 dimension이 같은 경우의 identity shortcut은 Eqn1을 사용하고, dimension이 증가한 경우 2가지 경우를 고려한다 1)zero entry로 패딩해서 dimension matching을 한 후 identity mapping 수행 2)Eqn2의 projection shortcut을 이용



1. Experiments

18/34-layer plain network의 train/validation error를 확인해보면 layer가 깊을수록 학습과정 전반에서 error가 높게 나타나는 degradation problem을 관찰할 수 있다.

18/34-layer residual network를 이용해 학습을 수행하고 plain/residual network들의 결과를 확인하면 Resnet은 layer가 깊어져도 더 좋은 성능을 보였고, 같은 layer 개수를 갖는 plain 보다 좋은 성능을 보인다. 그리고 마지막으로 Resnet이 유사한 성능을 보이는 plain과 비교할 때 빠른 수렴속도를 확인할 수 있다.

Resnet에서 Identity shortcut과 Projection shortcut을 비교했는데 3가지 옵션들 간의 성능 차이가 미미했고 이는 projection shortcut이 degradation problem 해결에 필수적이지 않다는 것을 알 수 있다.

Bottleneck 구조를 사용해서 50/101/152-layer residual network를 구성했고, 152층을 갖는 신경망이 가자 좋은 성능을 보여줬다. 즉 residual network를 이용하면 깊은 망을 쌓을 때 생기는 degradation problem을 해소할 수 있다.

이후 다른 깊이의 6개의 모델들을 ensemble을 수행한 결과 top-5 error가 3.57%라는 굉장한 성능을 보여줬다.

32x32 image를 갖는 CIFAR10 dataset을 이용해서 residual network를 분석했다. 위의 결과와 비슷하게 110 layer를 갖는 network가 가장 적은 error를 보여줬다. 또한 1202 layer의 경우 학습은 110 layer와 비슷한 error를 보이면서 잘 됐지만, test 성능이 좋지 못한 것을 보았을 때 overfitting 문제가 일어났다.

Reference

(ResNet) Deep residual learning for image recognition 번역 및 추가 설명과 Keras 구현(<https://sike6054.github.io/blog/paper/first-post/>)

[논문읽기] 02. Deep Residual Learning for Image Recognition

(<https://leechamin.tistory.com/184>)

CRNN 학습 결과

