

Deep residual learning for image recognition

Abstract

본 논문은 이전의 연구에서 다룬 네트워크보다 훨씬 더 깊은 네트워크의 용이한 학습을 위해 *residual learning framework*를 제안한다. 이 논문에서 말하는 *residual function*이란 layer의 입력을 참고하는 형태이다. 이러한 *Residual network*는 optimize가 쉽고, 깊은 네트워크에서도 정확도를 얻을 수 있다. VGG보다 8배 깊어진 152 layer를 사용한 residual network를 ImageNet으로 평가하였다. VGG보다 좋은 성능을 내고 복잡도는 감소하였고, 3.57%의 error를 달성해서 ILSVRC 2015에서 1등을 차지했다.

1.Introduction

Network들의 깊이가 깊어질수록 좋은 성능을 보여줬다. 이에 따라 'model의 layer를 단순히 쌓으며 depth을 늘리는 것이 학습에 유리한가?' 라는 궁금증이 생기게 되었다. 하지만 Deep network의 깊이가 어느정도 깊어지면, 수렴을 방해하는 vanishing/exploding gradient문제가 있다는 것을 알 수 있다. 이는 normalized initialization(->*weight initialization*)과 intermediate normalization(->*batch normalization*)을 통해 해결되었다. 그러나 또다른 문제가 발생한다.

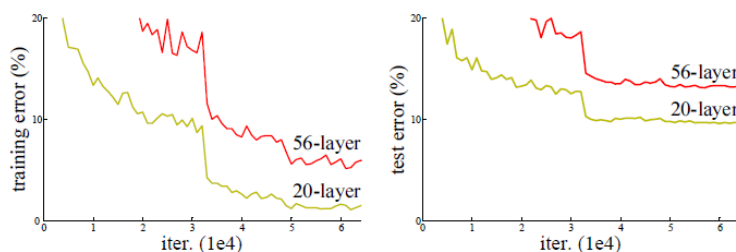


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer “plain” networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

Figure 1에서 볼 수 있듯, layer가 깊은 network의 training error와 test error가 layer가 얇은 network보다 높게 나타난다 (*Degradation Problem*). 이것은 overfitting의 문제가 아니라 최적화가 쉽지 않기 때문에 나타나는 현상이며, 모든 시스템의 최적화가 유사하지 않음을 나타낸다.

본 논문에서는 이 문제를 *Residual Learning Framework*를 도입해서 해결한다.

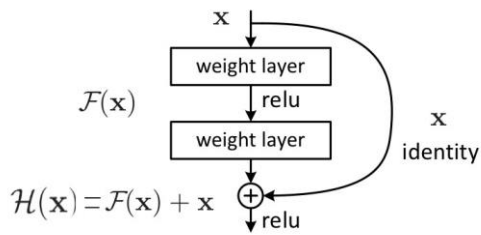


Figure 2. Residual learning: a building block.

이는 기존의 방식으로 $H(x)$ 를 optimization하는 것이 아닌 $F(x) = 0$ 이 되도록 optimize를 하는 것이다. Residual Learning의 residual(잔차)은 $H(x)$ 와 x 의 잔차인 $F(x)$ 를 의미하는 것이다.

$H(x) = F(x) + x$ 는 *shortcut connection(skip connection)*으로 구현할 수 있다. *Shortcut connection*이란 하나이상의 layer를 skip(건너뛰)하는 것이다. 본 논문에서는 Figure2와 같이 shortcut connection이 identity mapping을 수행하고, 그 출력을 stacked layer의 출력에 더해진다. Input인 x 가 output에 더해지는 것이기 때문에 추가적인 파라미터가 필요하지도 않고 연산복잡도도 증가하지 않는다. 자세한 Network 설명은 **3.Deep Residual Learning**에서 설명하도록 한다.