

ResNet

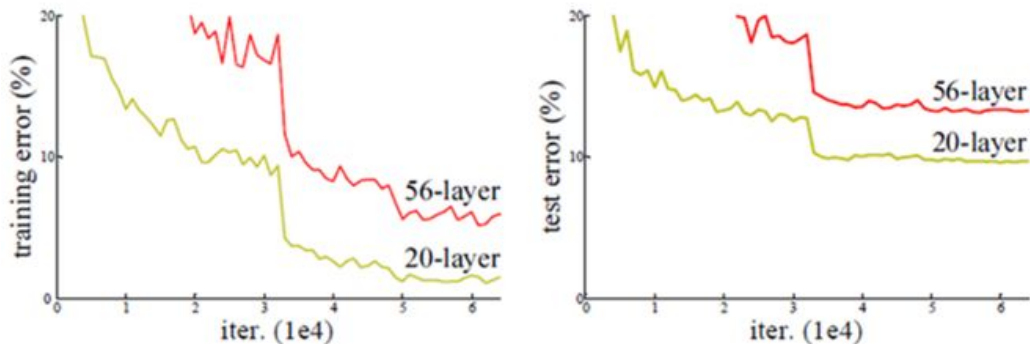
Deep Residual Learning for Image Recognition

Abstract

- 딥러닝에서 **neural network**가 깊어질수록 훈련하기 더 어려움
- 실질적으로 더 깊은 **network** 훈련을 용이하게 하기 위한 **residual learning framework**를 제시함
- ImageNet dataset을 사용해서 최대 **152개의 layer**를 쌓아 **3.57%의 error**를 달성함
 - VGGnet 보다 8배 깊으면서도 복잡성은 여전히 낮음

Introduction

- Network의 layer를 많이 쌓을수록 좋은 결과를 가지는가?



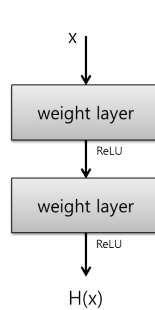
- degradation 문제 발생
 - layer가 깊을수록 성능이 상승하다가 gradient vanishing 때문에 성능이 더 안좋아짐

=> ResNet 방법이 고안됨

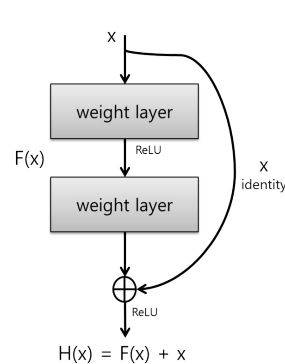
Residual Block

- Residual Learning

$$H(x) = F(x) + x$$
$$H'(x) = F'(x) + 1$$



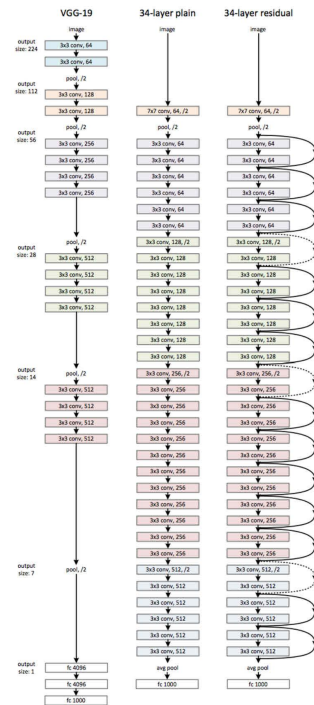
기존 방식



Residual block

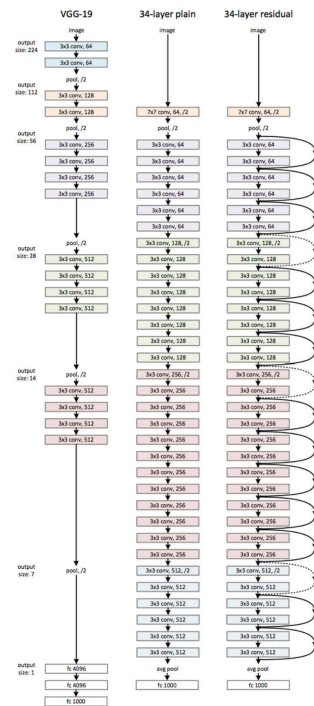
ResNet의 구조

- VGG-19 구조를 뼈대로 함
- 34-layer Plain Network
 - 각 layer들을 같은 크기의 output feature map을 만들고, 동일한 filter의 개수를 가짐
 - feature map size가 반으로 줄어들면, filter는 2배가 됨(layer 당 time complexity를 유지하기 위함)



ResNet의 구조

- 34-layer Residual Network
 - plain Network에 shortcut connection 추가
 - dimension이 증가할 때 두 가지 옵션 고려
 - identity shortcut connection을 계속 실행하는 경우, dimension을 증가시키기 위해 나머지를 zero padding함
 - projection shortcut connection은 dimension을 맞추기 위해 1x1 convolution을 사용함



Implementation

- image는 [256, 480] 중에 가까운 방향으로 **resize**됨
- 픽셀 당 평균을 뺀 이미지 또는 좌우반전된(**horizontal flip**) 이미지로부터 무작위로 샘플링한 **224x224 crop**
- **standard color augmentation** 사용
- **convolution** 이후 **activation** 이전의 **batch normalization** 적용
- **mini-batch size 256**으로 **SGD** 적용
- **learning rate**는 0.1부터 시작, **local minimum**을 만나거나 **loss**가 진동하는 경우 1/10씩 감소함
- **iteration** : $60 * 10^4$
- **weight decay** : 0.00001, **momentum**: 0.9
- **dropout** X

Experiments

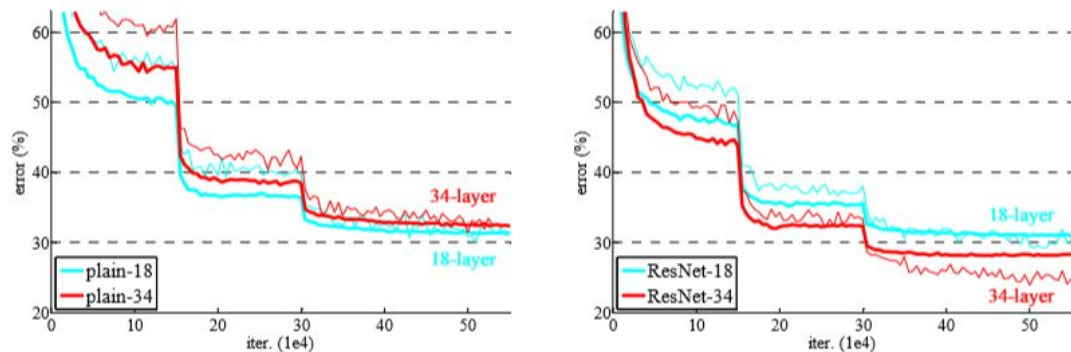


Figure 4. Training on ImageNet. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	25.03

Table 2. Top-1 error (% , 10-crop testing) on ImageNet validation. Here the ResNets have no extra parameter compared to their plain counterparts. Fig. 4 shows the training procedures.

Experiments

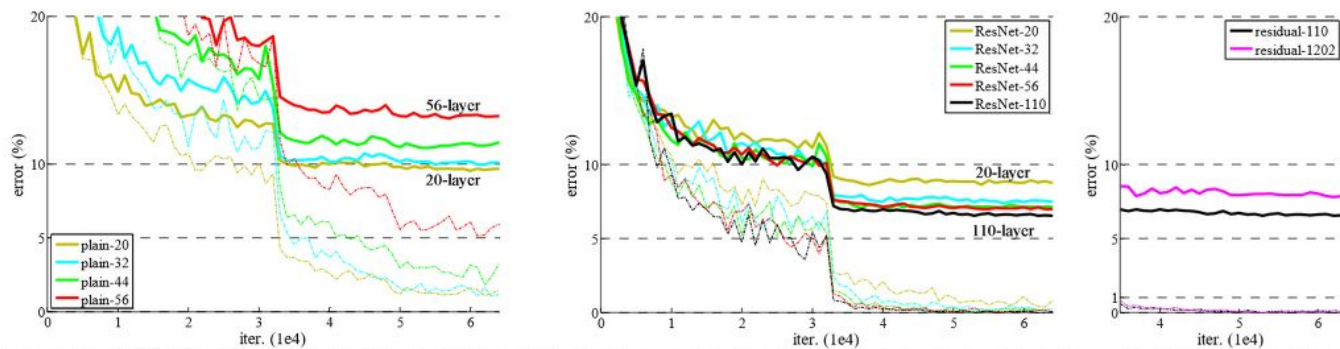


Figure 6. Training on **CIFAR-10**. Dashed lines denote training error, and bold lines denote testing error. **Left:** plain networks. The error of plain-110 is higher than 60% and not displayed. **Middle:** ResNets. **Right:** ResNets with 110 and 1202 layers.