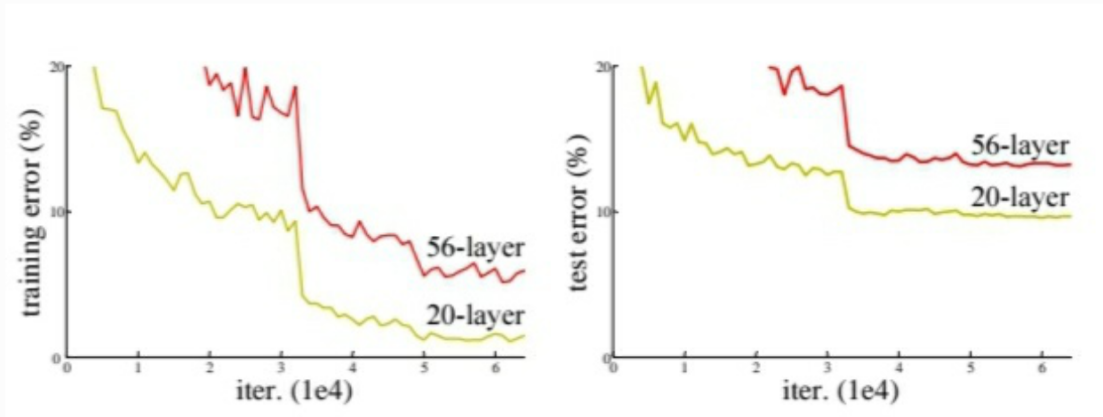


## 문제 제기

⇒ 많은 층으로 구성되어 있는 네트워크가 항상 좋은 성능을 내냐? **No!!**



⇒ 과거에는 Overfit 문제 때문에 좋은 성능을 못냄, 현재는 weight 초기화를 잘하고 좀더 효율적인 Normalization 으로 인해 overfit 어느 정도 해결

⇒ Gradient vanishing, exploding 문제 때문에 depth가 증가해도 어느 정도 성능에 도달하면 성능↓ (Degradation)

· Degradation = train, test 둘 다 성능↓

⇒ 이를 해결 하기 위해 ResNet 제안

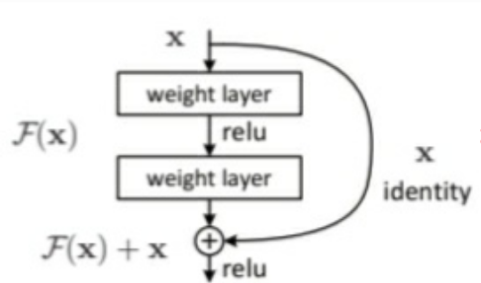
## Residual : 잔차

Idea: Original mapping 보다 residual mapping이 더 최적화 시키기 좋을것

Original Mapping:  $H(x) = F(x) + x$

residual Mapping:  $F(x) = H(x) - x$  **초점**

Target과 입력값이  
차이만 확실히겠다.



= Residual block

$F(x)$  = 잔차,  $x$  = 입력값,  $H(x)$ : 기존의 Neural Network

⇒  $F(x) = 0$  이 최적화이기 때문에 이전 Model보다 optimize가 용이

Shortcut connection : 한꺼번에 layer를 skipping 하는것

$$y = F(x, \{w_i\}) + x$$

Identity shortcut

$y$ : output layer  
 $x$ : input layer  
 $w$ : weight

조건:  $x$ 와  $y$ 의 Dimension이 항상 같아야됨  
( $\because$  + 연산하기 위해)

Identity shortcut:  $x$ 와  $y$ 의 차원이 같을때

$$y = F(x, \{w_i\}) + W_s x$$

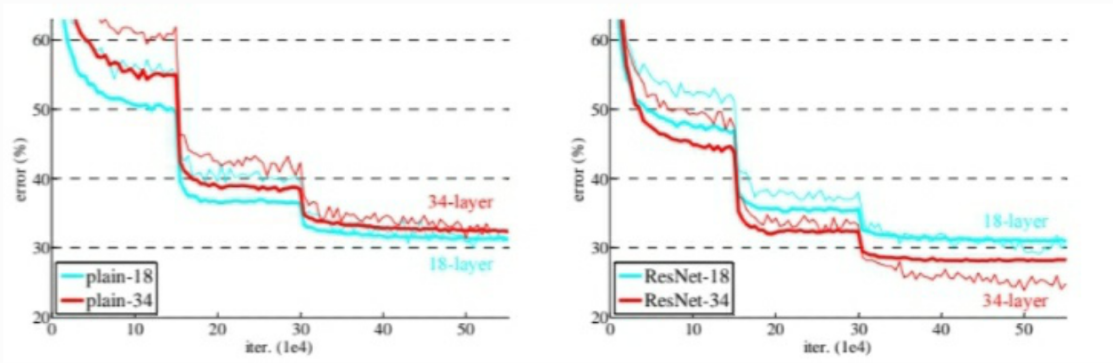
Projection shortcut

$y$ : output layer  
 $x$ : input layer  
 $w$ : weight  
 $s$ : square matrix

Projection shortcut:  $x$ 와  $y$ 의 차원이 다를때  
⇒  $W_s$ 로  $x$ 와  $y$ 의 차원을 맞춰줌

- Experiments

• Plain Network (기존의 NN) vs ResNet



plain 34 vs plain 18

ResNet 18 vs ResNet 34

=> 34 가 더 높은 Error 가림

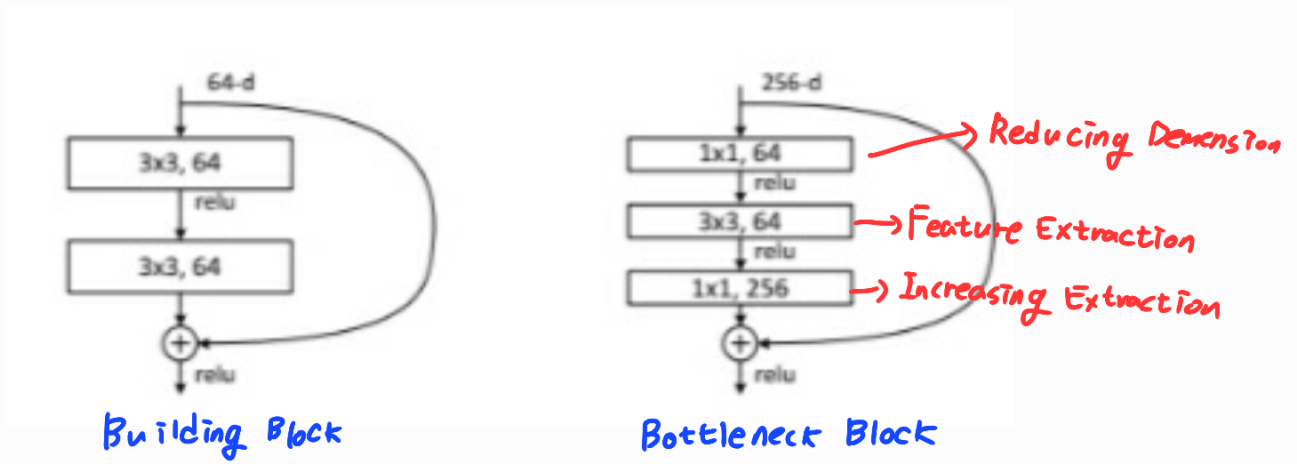
=> 34층이 더 낮은 error

• Identity shortcut 과 Projection shortcut

	ResNet-A	ResNet-B	ResNet-C
Increasing Dimension	Zero padding	Projection shortcut	Projection shortcut
Normal	Identity shortcut	Identity shortcut	Projection shortcut
Error Top 1	25.03	24.52	24.19
Error Top 5	7.76	7.46	7.40

=> ResNet-C 가장 좋은 성능, But, ResNet-B 와 성능차이 미세, memory/time 복잡도 ↑ 포함 X

• Depth 가 늘어날때 Degradation 문제가 나는지 안나는지



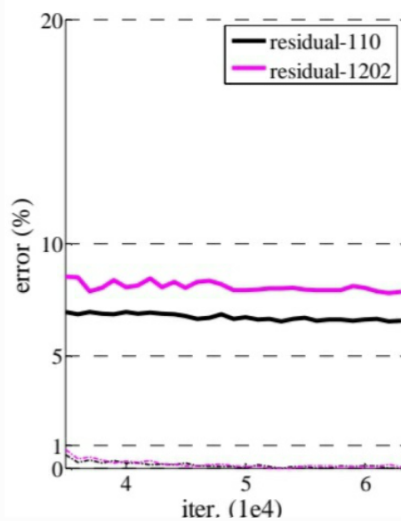
• Building Block : ResNet 34 에서만

• Bottleneck Block : ResNet 50, 101, 152

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	<b>21.43</b>	<b>5.71</b>

→ depth 가 깊어질수록 error rate ↓

• CIFAR-10 ⇒ layer를 100개 이상 쌓았을 때



⇒ • overfitting 문제

• maxout 이나 dropout 정규화 사용 X

• 따라서 maxout이나 dropout 사용하면 성능 향상 가능성