

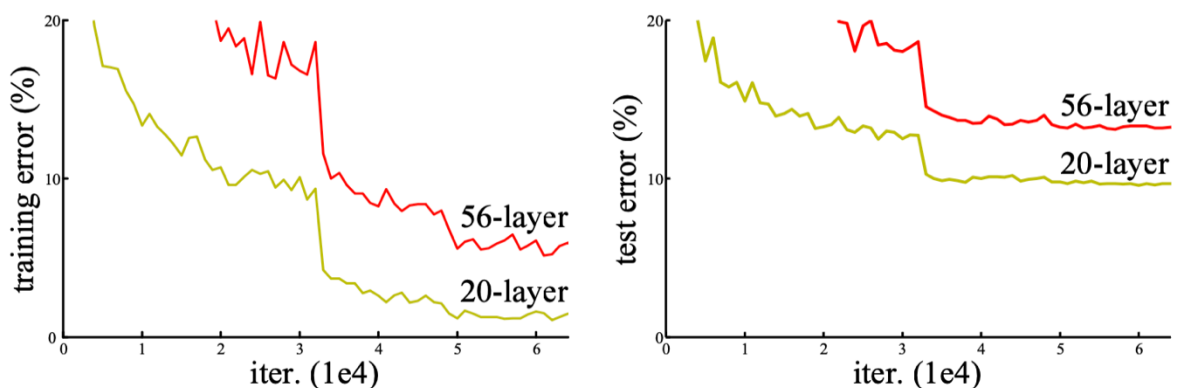
# Deep Residual Learning for Image Recognition – Review

학과 : 소프트웨어학과  
학번 : 20202020827  
학년 : 4 학년  
이름 : 김경민

## 1. 서론

- 연구 동기(왜 이 논문이 쓰였는가?)
  - 딥러닝 분야에서 네트워크의 깊이(depth)를 늘리는 것은 모델의 표현력을 향상시키는 대표적인 방법이다. 그러나 층이 깊어질수록 기울기 소실(vanishing gradient), 학습 불안정성, 그리고 네트워크가 깊어짐에도 성능이 오히려 떨어지는 Degradation 문제가 발생한다.
  - 이러한 문제를 해결하기 위해 He et al.는 Residual Learning(잔차 학습)을 제안하였고, 이를 통해 매우 깊은 신경망에서도 효율적이고 안정적으로 학습할 수 있는 ResNet(Residual Network) 구조를 고안하였다.
- 논문이 다루는 핵심 목표
  - 네트워크가 깊어질수록 발생하는 Degradation 문제를 해소할 수 있는 새로운 구조(Residual Block) 제안
  - 기존의 단순(plain) 네트워크와 잔차 연결(ResNet)을 적용한 네트워크를 비교하여, 깊이가 깊어질 때도 학습 오류가 증가하지 않고 오히려 성능이 향상되는 현상을 실험적으로 입증
  - ImageNet, CIFAR-10 등 다양한 데이터셋에서 모델을 학습·평가함으로써, ResNet 이 기존 모델 대비 더 우수한 결과를 낸다는 점을 검증

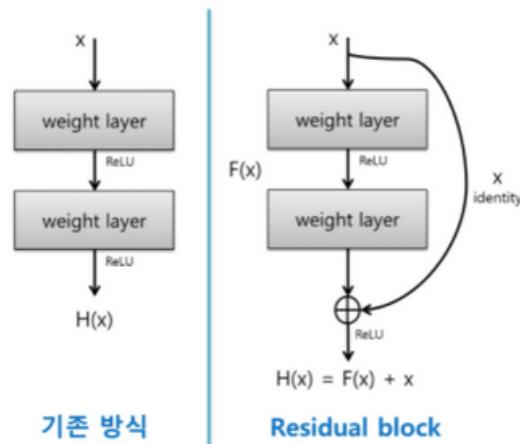
## 2. Degradation 문제: 20-layer vs. 56-layer Plain Network



- 논문에서는 “Degradation 문제”가 네트워크 깊이를 단순히 늘렸을 때 발생한다고 지적함.
- 특정 깊이를 넘어서면, 오히려 더 깊은 네트워크가 학습 난이도와 기울기 소실 문제로 성능이 떨어지는 현상이 발생.
- 위 그림은 56-layer plain 네트워크가 20-layer 보다 높은 오류율을 보이는 모습을 시각적으로 보여줌.
- 깊이가 늘어나면 모델 표현력은 증가하지만, 기울기 소실(vanishing gradient) 문제로 학습이 제대로 이루어지지 않을 수 있음.

- “깊으면 무조건 좋다”는 단순 가정이 틀렸음을 보여주는 사례로, 네트워크 구조 자체에 개선이 필요함

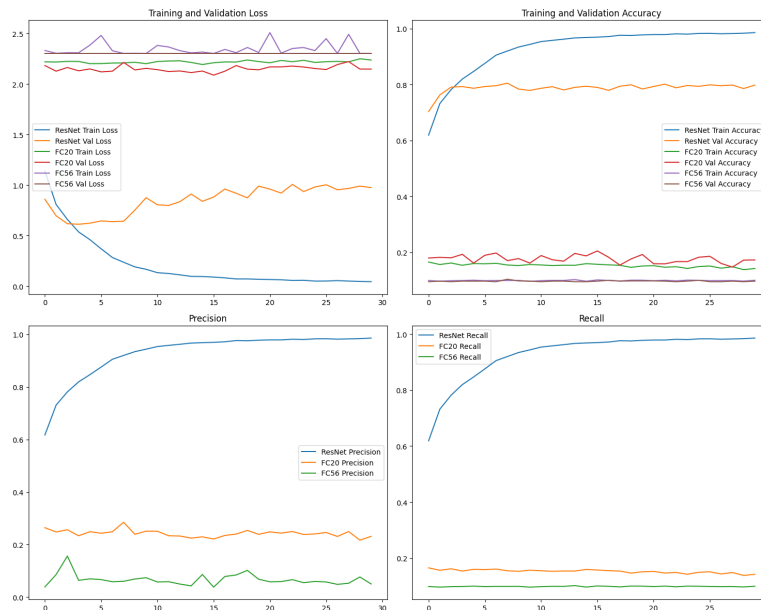
#### 4. Residual Block 구조: $y=F(x)+x$



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

- 기존의 신경망에서는 목표가  $H(x) = y$  인 함수가 되어, 학습 과정에서  $H(x) - y$  의 값을 최소화하는 방식으로 학습이 진행되었다. Residual Network (ResNet)에서는 이와 다른 접근, ResNet 은  $H(x) = x$  라는 목표를 두고,  $H(x) - x = F(x)$ 로 두어  $F(x)$ 를 최소화하려고 한다. 여기서  $F(x)$ 는  $H(x)$ 가  $x$ 에 대한 변형된 값을 나타내는 함수로,  $F(x) = 0$ 이라는 목표를 두고 학습을 진행합니다. 이때  $F(x)$ 가 0으로 수렴하도록 유도하는 과정에서 학습이 더 용이해지며, 기울기 소실 문제를 해결할 수 있다.
- Skip Connection 은 입력값  $x$ 가 여러 층을 건너뛰어 출력값에 더해지는 역할을 한다. 이 방식으로 입력값이 그대로 출력에 더해지며, 네트워크의 깊이가 깊어져도 기존 입력값이 손실되지 않고 효과적으로 전달된다. 이를 통해 모델이 깊어져도 초기 정보가 소실되지 않으며, 기울기 역전파가 원활하게 이루어진다. 결국  $H(x) = F(x) + x$ 라는 식으로, 네트워크는  $x$ 를 직접 더하는 방식으로 정보를 전달하고, 이를 통해 기울기 소실 문제를 완화하고 학습의 안정성을 향상시킬 수 있다.
- ResNet에서는 입력  $x$ 가 각 층을 거칠 때마다 변환된 값에 추가되므로, 초기 입력 정보가 그대로 전달되어 학습이 원활하게 이루어진다. 따라서 네트워크가 깊어져도 기울기 소실 문제를 해결할 수 있으며, 더 깊은 네트워크에서 학습이 안정적으로 이루어지도록 돕는다.

## 📌 관련 실험 (FC20 vs. FC56 vs ResNet-18 )



Mode	Train Loss	Val Loss	Train Accuracy	Val Accuracy
FC20	2.2377	2.1482	0.1419	0.1727
FC56	2.307	2.3029	0.0999	0.0965
ResNet	0.0043	0.9452	0.9145	0.7982

- CIFAR-10 에서 ResNet-18 모델을 사용해 본 결과:

- FC20/FC56 대비 훨씬 빠른 손실 감소와 안정적 학습 곡선
- 훈련 정확도 **90% 이상**, 검증 정확도 약 79%로, 단순 FC 모델에 비해 월등히 높은 성능
- 이는 논문에서 말한 “잔차 연결로 인해 깊어져도 성능이 떨어지지 않는다”는 점을 실제 코드로 재현
- CIFAR-10 데이터셋에서 FC20(20 층 완전 연결)과 FC56(56 층 완전 연결)을 비교.
  - FC20: 약 **16%** 내외의 낮은 정확도
  - FC56: 깊이가 더 깊음에도 불구하고 정확도가 **10% 이하**로, 오히려 더 낮아짐
- 이는 논문에서 말한 Degradation 문제와 동일한 현상. 단순히 층만 늘리면 성능이 올라가지 않고, 최적화 어려움으로 학습이 불안정해지는 결과를 확인.

## 4. VGG-19 vs. 34-layer Plain vs. 34-layer Residual

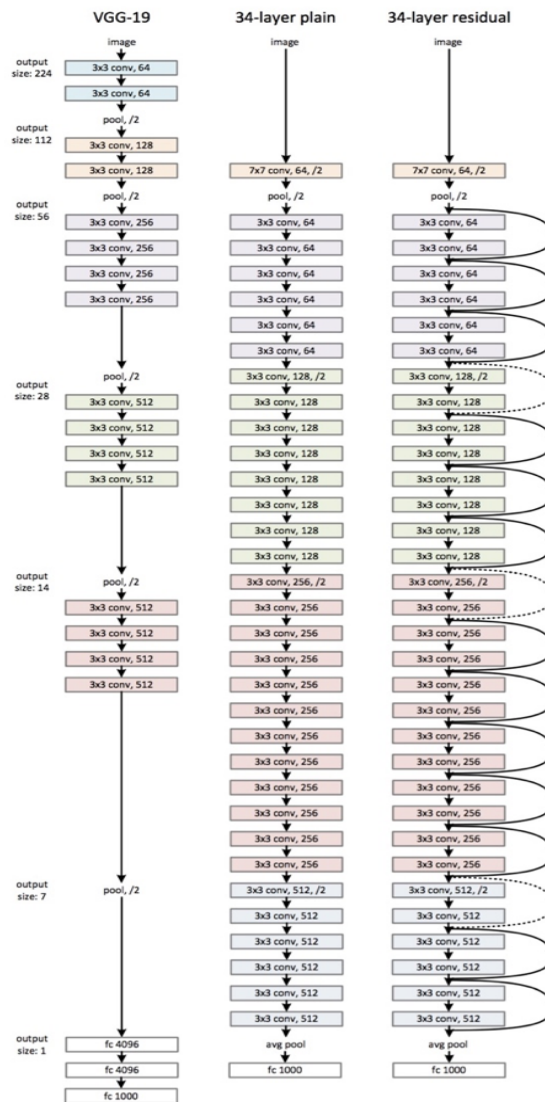


Figure 3. Example network architectures for ImageNet. **Left:** the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

(왼쪽) - VGG-19: 모든 합성곱을  $3 \times 3$  필터로 통일해 층을 깊게 구성

(가운데) - 34-layer plain: 34 층까지 단순히 쌓은 구조 (잔차 연결 없음)

(오른쪽) - 34-layer residual: 동일한 34 층이지만, 잔차 연결을 도입한 ResNet

- 논문에서는 VGG 와 Plain 네트워크, 그리고 Residual 네트워크를 구조적으로 비교함.
- 결과적으로 Plain 모델은 깊이가 깊어질수록 Degradation 문제로 성능이 낮아지고, Residual 모델은 오히려 깊어질수록 성능이 개선됨을 강조.

## (2) 해석

- VGG 는 깊이가 깊어지면 여전히 기울기 소실 문제를 겪을 수 있으나, ResNet 의 잔차 연결은 이 문제를 해결해 더 깊은 층에서도 안정적인 학습을 가능케 함.

### 관련 실험 (Custom-PlainCNN vs. ResNet vs. VGG, CIFAR-100)

#### • 모델 구성

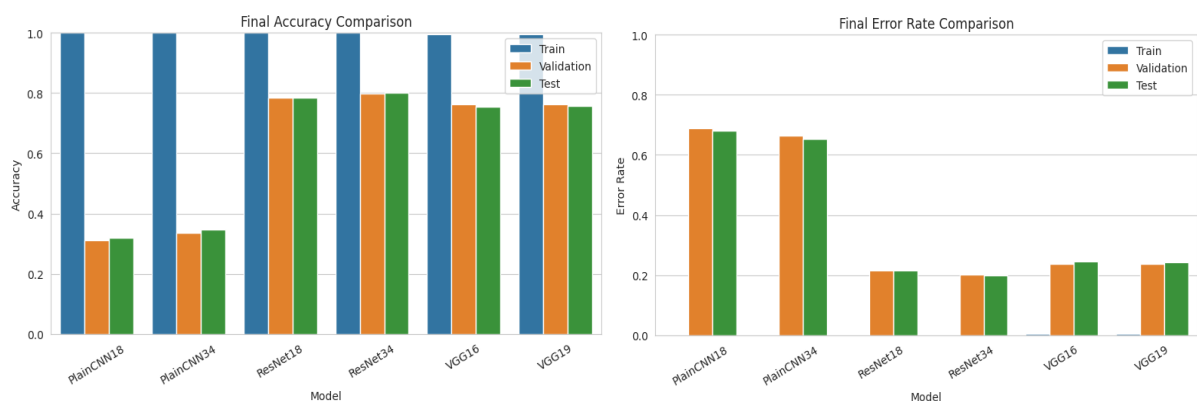
- PlainCNN18/34: 단순 CNN 구조, convolution 레이어와 FC 레이어를 통해 100 개 클래스로 출력
- ResNet/VGG: 사전 모델로, 마지막 출력층을 CIFAR-100 에 맞게 수정

#### • 데이터 전처리 및 구성

- 이미지 리사이즈:  $224 \times 224$
- 텐서 변환 및 ImageNet 통계(mean, std)로 정규화
- CIFAR-100 데이터셋을 80:10:10 비율로 학습/검증/테스트 분할

#### • 학습 설정

- 손실 함수: CrossEntropyLoss
- 옵티마이저: SGD (lr=0.001, momentum=0.9)
- 총 에폭 수: 40
- 에폭별로 손실, 정확도, 에러율 출력 및 기록



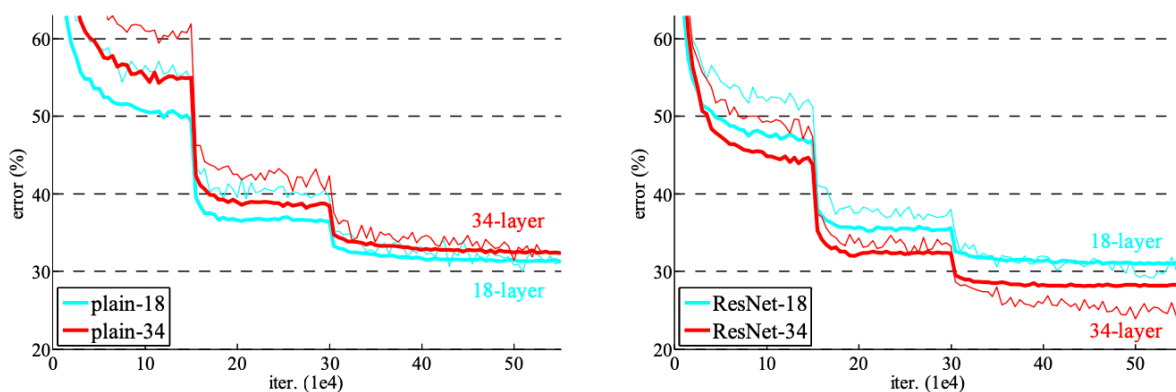
위 표는

Model	Val Accurcay	Test Accurcay	Test error
PlianCNN18	31.24%	32.04%	68.00%
PlainCNN34	33.64%	34.72%	65.28%
VGG16	76.34%	75.56%	24.44%
VGG19	76.32%	75.70%	24.30%
ResNet18	78.52%	78.40%	21.60%

ResNet34	79.82%	80.02%	20.00%
----------	--------	--------	--------

- CIFAR-100 에서 PlainCNN(18, 34), ResNet(18, 34), VGG(16, 19)를 비교 실험.
  - PlainCNN 모델들은 클래스가 100 개로 늘어나자 **에러율이 높고 학습이 불안정**.
  - VGG16/19 는 나쁘지 않은 성능이나, **ResNet18/34** 가 에러율 면에서 더 우수한 결과를 보임.
- 이는 논문에서 지적한 대로, **깊은 네트워크에서도 잔차 연결이 핵심적인 역할**을 한다는 사실을 재확인해 준다.

## 5. ImageNet 성능 비교



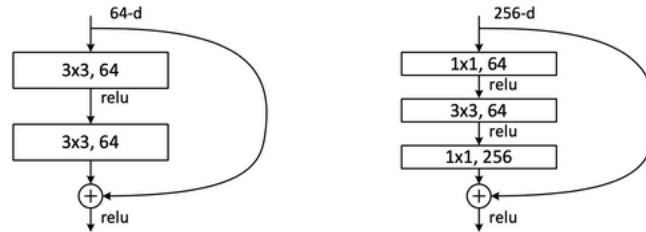
	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	<b>25.03</b>

- Plain-18 vs. ResNet-18: Top-1 오류율이 27.94% vs. 27.88% (큰 차이는 아니지만 ResNet 이 근소하게 우수)
- Plain-34 vs. ResNet-34: 28.54% vs. 25.03%로, ResNet-34 가 확실히 우수
- 파라미터 수가 크게 늘지 않았음에도 불구하고, 잔차 연결로 인해 성능이 향상됨을 시사
- ImageNet 처럼 대규모 데이터셋에서도 Residual Learning 이 효과적임을 수치로 입증.
- 깊이가 34 층이 될 때, Plain 네트워크는 오류율이 급격히 올라가지만, ResNet 은 25%대의 오류율로 상대적으로 낮은 편.

### (2) 해석

- 논문 전반의 메시지를 요약해주는 표로, “깊이가 같아도 구조적 차이(잔차 연결)로 인해 성능이 달라진다”는 점을 수치로 확인 가능.
- 특히 34 층에서의 3% 이상 차이는 매우 의미가 크며, 이후 50 층·101 층·152 층까지 확장되면서 ResNet의 우수성이 더 명확히 드러남.

## 6. Residual Block 구조: $y=F(x) + x$ (Residual Connection)



위의 Residual Block 구조를 더욱 깊게 확장한 모델에서는 Bottleneck 레이어가 사용됩니다. Bottleneck 구조는 1x1 합성곱을 사용하여 파라미터 수를 줄이면서도 네트워크의 성능을 유지할 수 있다. Bottleneck 레이어를 통해 연산량을 줄이면서도 네트워크 깊이를 증가시킬 수 있는 방법을 제공

- Bottleneck 레이어는 3x3 필터와 1x1 필터를 결합하여, 파라미터 수를 효율적으로 줄이며, 깊이를 더 깊게 할 수 있도록 도와줍니다. 예를 들어, 64-d와 같은 구조에서 3x3 필터가 두 개 연속으로 사용되며, 이를 통해 더 작은 연산으로도 더 많은 비선형성(non-linearity)을 추가할 수 있다.
- 1x1 컨볼루션을 사용하여 채널 수를 줄이는 방식은 연산량을 줄이면서도 성능 손실을 최소화한다. 이는 기존 모델에서 발생할 수 있는 기울기 소실(vanishing gradient) 문제를 해결하는 데에도 도움을 준다. Bottleneck 구조를 채택함으로써, 모델은 더 많은 파라미터를 사용하지 않고도 깊이를 늘리고, 효율적인 학습을 할 수 있게 된다.

## 7. CIFAR-10 추가 결과: FitNet, Highway Network, ResNet 등 비교



method			error (%)
Maxout [9]			9.38
NIN [25]			8.81
DSN [24]			8.22
	# layers	# params	
FitNet [34]	19	2.5M	8.39
Highway [41, 42]	19	2.3M	7.54 (7.72±0.16)
Highway [41, 42]	32	1.25M	8.80
ResNet	20	0.27M	8.75
ResNet	32	0.46M	7.51
ResNet	44	0.66M	7.17
ResNet	56	0.85M	6.97
ResNet	110	1.7M	<b>6.43</b> (6.61±0.16)
ResNet	1202	19.4M	7.93

(CIFAR-10 분류 에러)

(FitNet, Highway Network, Maxout, DSN 등 기존 모델과 ResNet 을 비교, ResNet-110, ResNet-1202 등이 오류율 6~7%대로 다른 모델 대비 성능 우위)

- 논문은 CIFAR-10 에서도 다양한 기존 모델(Highway Network, FitNet, DSN 등)과 비교 실험을 진행.
- ResNet-110: 약 6.43~6.61%의 오류율로 기존 모델보다 낮은 에러율을 달성.
- 극단적으로 ResNet-1202 층까지 쌓아도, 여전히 오류율이 7%대 수준으로 준수한 성능을 보여줌.

## (2) 해석

- ResNet 은 단순히 ImageNet 뿐 아니라, 소규모/중규모 데이터셋(CIFAR-10/100)에서도 좋은 성능을 낸다는 점을 강조.
- Highway Network 처럼 직접 연결을 활용한 모델들도 있지만, ResNet 의 skip connection 방식이 구현이 단순하면서 효과가 크다고 볼 수 있음.

## 8. 결론

이번 논문을 통해 Degradation 문제와 Residual Block 의 역할을 깊이 이해할 수 있었다. 단순히 네트워크 깊이를 늘린 Plain(CNN) 모델은 기울기 소실(vanishing gradient) 문제로 인해 오히려 성능이 저하될 수 있음을 확인했다. 특히, FC20 vs. FC56 실험과 20-layer vs. 56-layer 그래프에서, 네트워크의 깊이가 증가할수록 성능이 떨어지는 현상을 확인했다.

반면, Residual Block 을 사용한 모델에서는  $y=F(x)+x$  구조가 기울기 전파를 원활하게 돕고, 네트워크가 깊어져도 학습이 안정적으로 진행되는 것을 확인했다. ResNet-18, ResNet-34, ResNet-50 등의 모델에서 잔차 연결을 통해 성능 향상과 더불어 안정적인 학습이 가능함을 확인했다.

또한, ImageNet 과 CIFAR-10/100 과 같은 다양한 데이터셋에서 ResNet 의 성능이 뛰어난 결과를 보였으며, 이를 통해 ResNet 이 기존 모델들보다 우수한 성능을 보였다.

FC 모델과 PlainCNN 모델, VGG 모델과 비교한 실험에서도 **잔차 연결**이 있는 **ResNet**이 더 나은 성능을 보여주었습니다. 이로써, "층을 깊게 쌓는 것"만으로는 한계가 있다는 점을 다시 한번 확인할 수 있었다.

이번 논문을 읽으며 특징 맵의 크기 축소(downsampling)와 **풀링층** 및 **stride**의 역할을 이해함으로써, 기울기 손실을 방지하는 방법을 학습할 수 있었다.