논문 스터디

“Resnet”

Deep Residual Learning for Image Recognition

발표자 : 신동훈

목차

A table of Contents

Part 1,

Abstract

Part 1

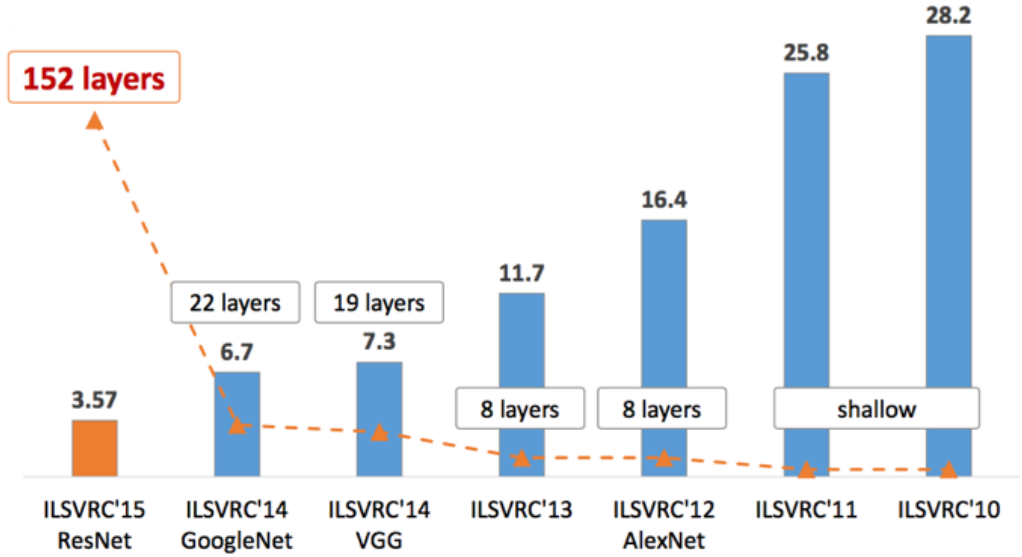
Abstract

- 본 논문은 심층 신경망의 학습이 어렵기 때문에 “**residual learning framework**”를 제안한다.

- 제안된 방법은 **최적화**하기도 쉽고, 깊이가 상당히 증가된 네트워크에서도 **정확도**를 얻을수 있다는 것을 실험 결과로 증명한다.

- 본 연구에서는 ImageNet dataset을 통해 최대 **152개의 Layer**로 구성된 residual network를 평가한다.
  - 이는 VGG 네트워크보다 **8배**는 더 깊지만 복잡성은 낮다.

- 두 개의 모델을 앙상블하여 Image Net test set에서 **3.57%**의 top -5 error를 보였고, ILSVRC 2015 classification 부문에서 **1위**를 차지했다.



ILSVRC 2014의 모델들과 ResNet의 비교 결과

Part 1

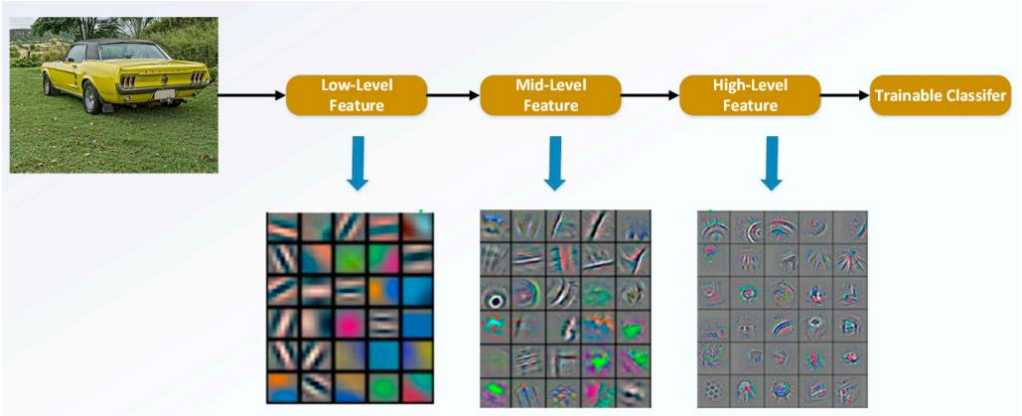
Introduction

- DCNN은 이미지 분류 Task에서 큰 발전을 가져왔다.

- Deep-Network는 모델을 end-to-end multi-layer 방식으로 통합이 가능하며, 각 Feature들의 Level은 해당 Feature를 추출하기 위해 거친 layer의 depth에 따라 달라진다.

- 깊이의 중요성이 커지면서 "Is learning better networks as easy as stacking more layers?" 라는 의문점이 생겼다.
  - Gradients vanishing/exploding 문제가 있었지만, 이는 'normalize-d initialization'과 'intermediate normalization' 기법을 통해 수렴이 가능해졌다.

- 또, 다른 문제는 성능 저하인데 깊이에 따른 정확도 저하 현상이다.
  - 이는 과적합이 아니라고 주장하고 있다.



각 Level에 해당하는 Feature

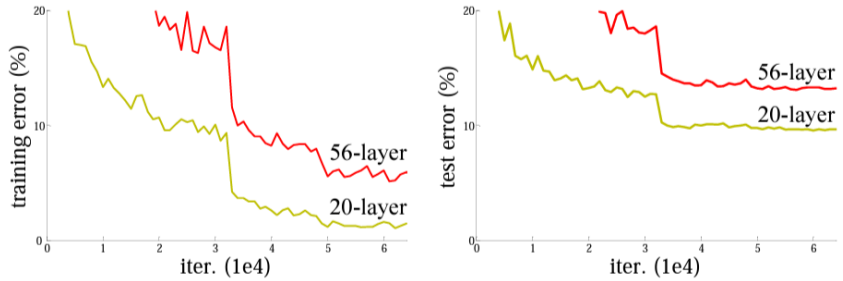


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

Part 1

Introduction

- 성능 저하 문제에 있어 깊이에 따라 최적화 난이도는 바뀐다.
- 하지만 깊은 모델도 제한적인 상황에서 최적화 가능한 솔루션이 있다.
  - 이는 추가된 Layer는 identity mapping, 다른 Layer는 shallower model에서 학습된 Layer로 구성하는 것이다.
- 해당 솔루션의 의미는 deeper 가 shallower 보다 결과가 좋아야 한다는 것을 의미한다.
- 따라서 본 논문에서는 "성능 저하 문제를 해결하기 위해 "Residual Learning Framework"를 제안한다.
  - 이는 few stacked layer가 직접 underlying mapping을 학습하게 하지 않고, residual mapping을 학습하도록 하는게 좋다.

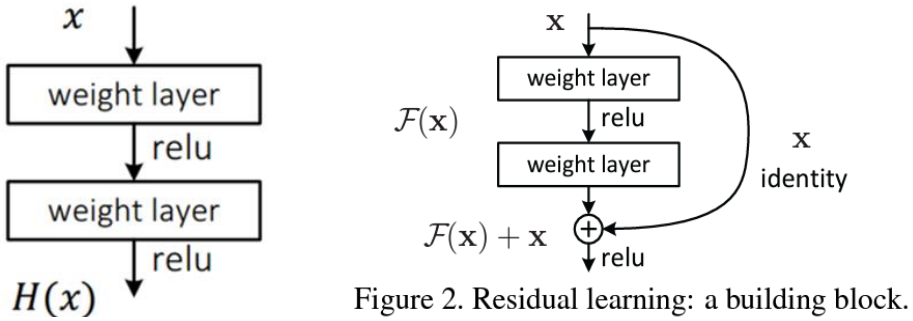


Figure 2. Residual learning: a building block.

H(x) – x

F(x) = H(x) - x

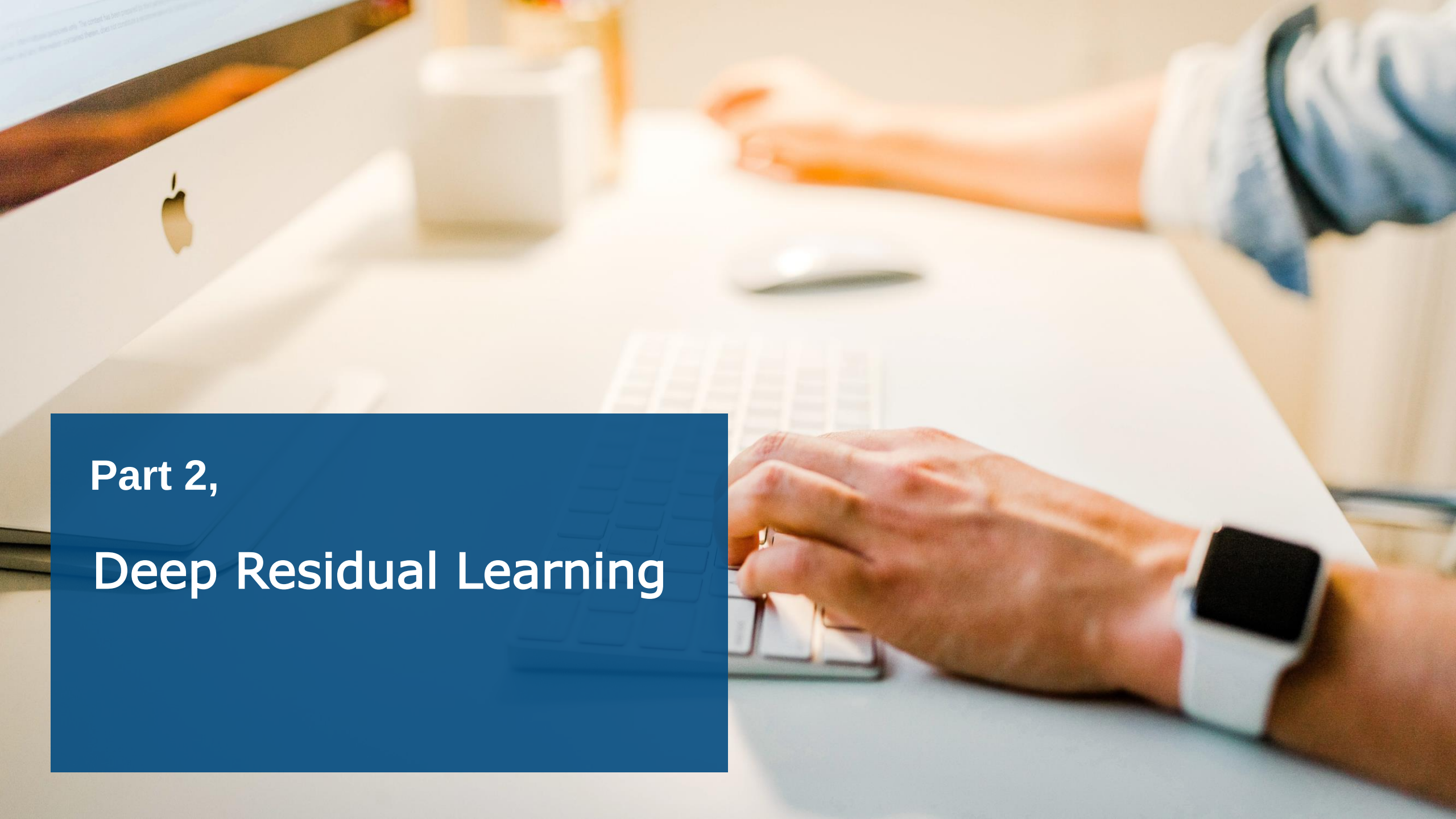
H(x) = F(x) + x

Part 1

Introduction

- 본 논문에서는 ImageNet 데이터에 대한 두 가지 실험 결과를 제공하는데,
  - plain network(simply stack layers)는 depth가 깊어짐에 따라 더 높은 training error를 보이는 것에 반해, 제안한 deep residual network는 쉽게 최적화가 가능하다.
  - deep residual network는 아주 깊어진 depth에서 성능의 이득을 가졌으며, 이전에 연구됐던 네트워크에 비해 훨씬 향상된 결과를 보인다.

- 또한, CIFAR-10 dataset에 대해서는 다음 실험의 결과를 제공한다.
  - 성능 저하 문제 및 제안하는 방법의 효과가 특정 dataset(ImageNet)에만 국한되지 않음을 보임
  - 제안하는 방법의 사용 여부에 따른 layer response의 std 분석
  - 1000개 이상의 layer로 이루어진 모델에 대한 실험



Part 2,

Deep Residual Learning

Deep Residual Learning

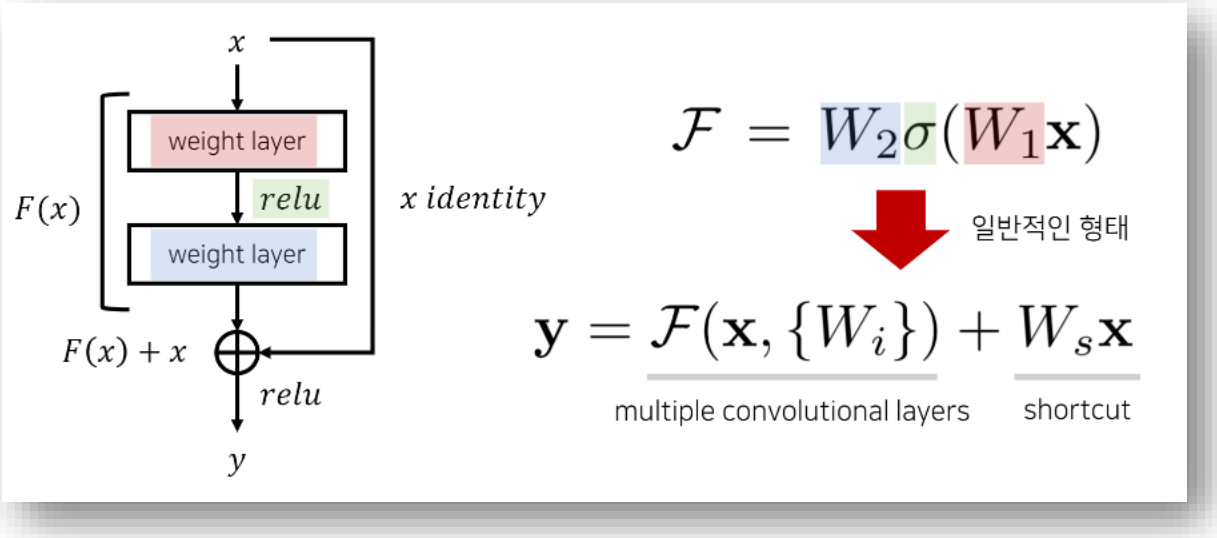
3.1 Residual Learning

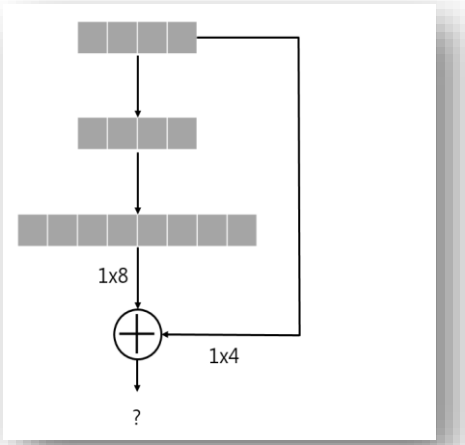
- 실제로는 Identity mapping이 최적일 가능성이 낮다고 한다.
  - 하지만, ResNet에서 제안하는 재구성 방식은 문제에 pre-conditioning을 추가하는데 도움을 준다.
  - 따라서 pre-conditioning으로 인해 Optimal function이 zero mapping보다 identity mapping에 더 가깝다면, solver가 identity mapping을 참조하여 작은 변화를 학습하는 것이 새로운 function을 학습하는 것보다 더 쉬울 것이라고 마이크로소프트팀은 주장한다.

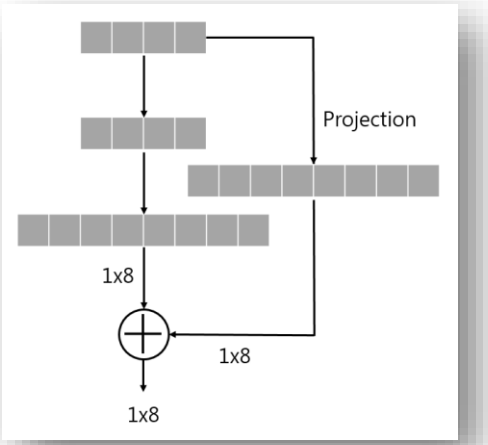
Deep Residual Learning

3.2 Identity Mapping by Shortcuts

- **Shortcut connection은 파라미터나 연산 복잡성을 추가하지 않는다.**
  - 이때, F + x 연산을 위해 x와 F의 차원이 같아야 하는데, 이들이 서로 다를 경우 linear projection인 Ws 를 곱하여 차원을 같게 만들 수 있다. 여기서 Ws 는 차원을 매칭 시켜줄 때에만 사용한다.







Part 2

Deep Residual Learning

3.3 Network Architectures

Plain Net

기본적인 네트워크는 Vgg Net을 기반으로 설계한다. Conv layer은 대부분 3X3 필터를 가진다.

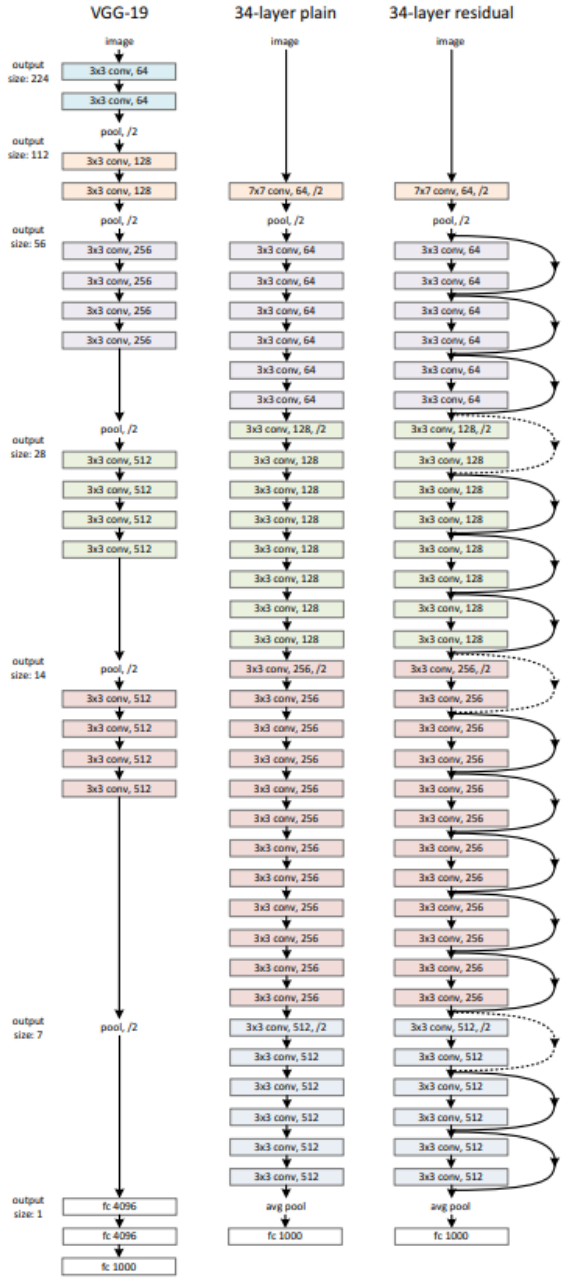
ResNet

Plain net을 기반으로, Shortcut Connection을 삽입하여 Residual Version의 네트워크를 만든다.

일반적으로 입력과 출력이 동일한 경우(solid line shortcuts)identity shortcut을 직접 사용이 가능하며, 차원이 증가하는 경우(dotted line shortcuts)에는 다음과 같은 두가지 옵션을 고려한다.

1. zero entry를 추가로 padding하여 dimension matching 후 identity mapping을 수행한다. (별도의 parameter가 추가되지 않음)

2. Eqn.2의 projection shortcut을 dimension matching에 사용한다.



Part 3,

Experiments

Part 3

Experiments

4.1 ImageNet Classification

본 논문에서는 실험을 위해 1000개의 class로 구성 된 ImageNet 2012 classification dataset 사용한다. 테스트 결과는 top-1 error와 top-5 error를 모두 평가한다

Plain Net

첫 번째는 18-layer와 34-layer을 가지는 Plain net의 평가를 진행한다. 34-layer 네트워크는 그림 3의 중간에 해당하며 18-layer는 비슷한 형태를 가지며 자세한 내용은 표 1을 살펴보면 된다.

결과를 나타내는 표 2를 살펴 보면 18-layer network에 비해, 34-layer network의 validation error가 높은 것으로 보인다. 여기서 degradation의 문제를 관측했다.

이유를 확인하기 위해 그림 4를 살펴보면 layer가 더많은 plain-34가 더 높은 Training error를 보이는 것을 알 수 있다.

	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.

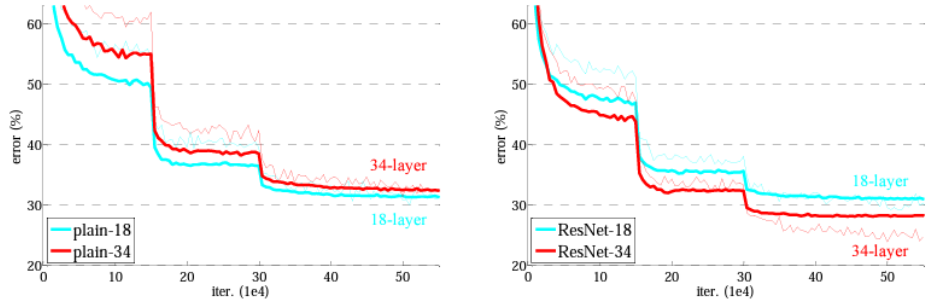


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

Part 3

Experiments

4.1 ImageNet Classification

ResNet

두 번째 실험 모델은 ResNet이며 18-layer와 34-layer를 비교 분석 한다. 앞서 3.3 절에서 dimension이 증가할 때 short connection을 하는 방법에 두 가지 옵션이 있었는데, 해당 실험에서는 zero padding해 주는 방식을 사용한다.

실험에서는 3가지 주요 사실이 관찰 되었다.

1. 상황이 반전되어 34가 2.8% 정도 더 좋은 성능을 보이며, 34가 상당히 낮은 training error를 가지는 것으로 보이며, 이에 따라 상대적으로 좋아진 validation성능이 포착되었다. 이는 degradation problem를 피할 수 있다는 근거가 되며 증가된 depth에서도 합리적인 정확도를 얻을 수 있다는 것을 나타낸다.

2. plain 모델과 비교해보면 residual learning을 적용하면 더 깊게 망을 쌓고, 더 좋은 성능을 낼 수 있다는 점을 보여준다.

3. plain 모델과 resnet 모델을 비교했을 때(그림 4), 각각의 성능은 비슷하고, 비교적 정확하지만 초기 단계에서 resnet 모델이 좀 더 빨리 수렴하여 최적화를 용이하게 한다.

	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.

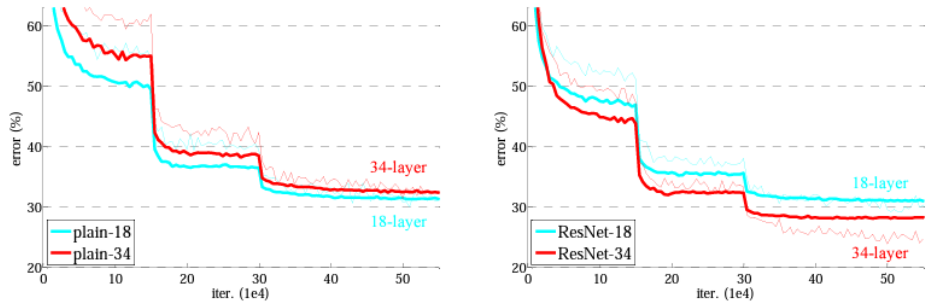


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

Part 3

Experiments

4.1 ImageNet Classification

Identity vs. Projection Shortcuts.

앞선 절에서 Identity Shortcuts을 통해 좋은 성능을 가진 다는 것을 실험으로 입증했다. 앞선 실험에서는 차원이 증가 할 때 zero padding을 이용했는데, 본 실험에서는 zero padding 이외의 다른 옵션들을 실험한다.

1. zero-padding shortcut는 dimension matching에 사용되며, 모든 shortcut는 추가적인 Weight 사용이 없기 때문에 parameter-free하다(앞선 실험에 사용됨).

2. projection shortcut는 dimension을 늘릴 때만 사용되며, 다른 shortcut은 모두 identity다. 이때는 차원을 늘려줘야 하기 때문에 늘려주고 싶은 차원 수 만큼 1x1 conv filter 를 사용한다. 그래서 plain 모델과 비교했을 때 parameter는 약간 증가한다.

3. 모든 shortcut은 projection이다. 이때 파라미터 증가가 커진다.

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	**21.43**	**5.71**

Table 3. Error rates (%, **10-crop** testing) on ImageNet validation. VGG-16 is based on our test. ResNet-50/101/152 are of option B that only uses projections for increasing dimensions.

Part 3

Experiments

4.1 ImageNet Classification

Deeper Bottleneck Architectures

다음으로는 training time을 고려하여 building block을 bottleneck design으로 수정한다.

그림 5에서는 2-layer stack과 3-layer stack의 디자인을 보여준다. 둘은 유사한 time complexity를 갖는다 [5]. 여기서 parameter-free인 ideneity shortcut은 이 architecture에서 특히 중요하다.

만약 그림 5의 오른쪽 다이어그램에서 identity shortcut이 projection으로 대체된다면, shortcut이 두 개의 high dimensional 출력과 연결되므로 time complexity와 model size가 두 배로 늘어난다. 따라서 identity shortcut은 이 bottleneck design을 보다 효율적인 모델로 만들어준다.

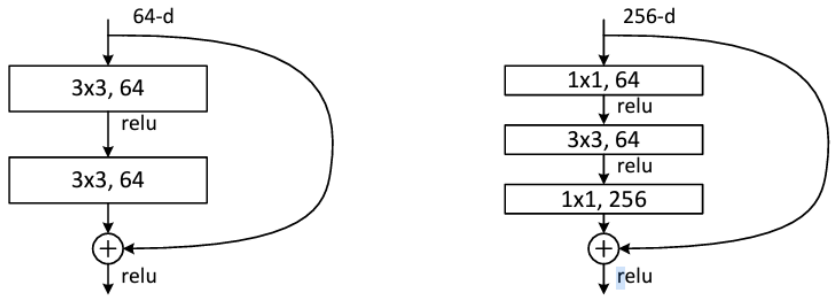


Figure 5. A deeper residual function $\mathcal{F}$ for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a "bottleneck" building block for ResNet-50/101/152.

경청해주셔서 감사합니다.