Deep Residual Learning for Image Recognition (2015)

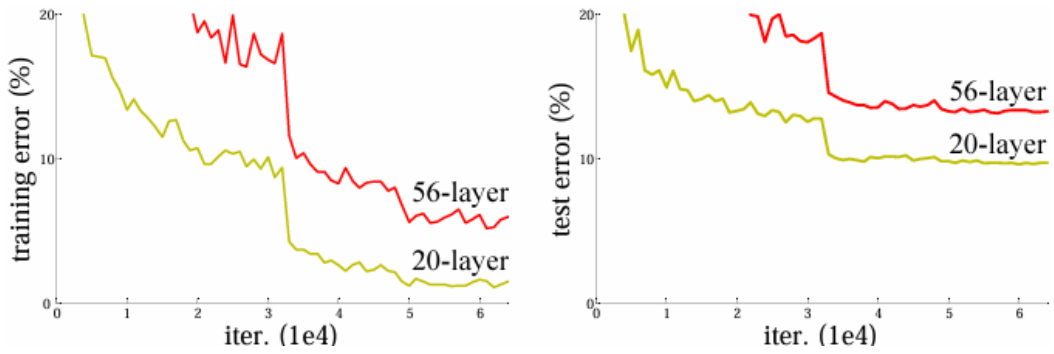
태그 ResNet

Deep Residual Learning for Image Recognition 논문 리뷰

1. 서론

딥러닝은 컴퓨터 비전 분야에서 혁신적인 성과를 이루었습니다. 특히 네트워크의 깊이가 증가할수록 모델의 표현력이 커지고, 복잡한 패턴을 더 잘 학습할 수 있다는 믿음이 있었습니다. 실제로 VGGNet, GoogLeNet 등은 네트워크를 더 깊게 쌓으면서 뛰어난 성능을 보였습니다. 하지만 네트워크가 일정 이상 깊어지면 오히려 학습이 잘 되지 않고, 성능이 떨어지는 현상(Degradation Problem)이 나타납니다. 이는 단순히 과적합 때문이 아니라, 깊은 네트워크에서 최적화가 어렵기 때문입니다.

이 논문은 이러한 문제를 해결하기 위해 **잔차 학습(Residual Learning)**이라는 새로운 접근법을 제안합니다. 이 방법은 네트워크가 입력값과 출력값의 차이(잔차)만을 학습하도록 유도하여, 매우 깊은 네트워크에서도 안정적으로 학습할 수 있게 합니다.



2. 관련 연구(배경)

기존의 딥러닝 네트워크는 대부분 컨볼루션 레이어를 여러 개 쌓는 구조였습니다. VGGNet, GoogLeNet 등은 네트워크를 깊게 쌓으면서 성능을 높였지만, 네트워크가 너무 깊어지면 오히려 학습이 어렵다는 한계가 있었습니다. 이는 기존의 최적화 방법이 깊은 네트워크에 적합하지 않기 때문입니다.

이 논문은 네트워크의 깊이와 성능 간의 관계를 실험적으로 분석하며, 단순히 네트워크를 깊게 쌓는 것만으로는 한계가 있음을 보여줍니다. 특히 CIFAR-10 데이터셋에서 20층과 56층 plain 네트워크를 비교하면, 더 깊은 네트워크가 오히려 더 높은 오류율을 보이는 현상을 확인할 수 있습니다.

3. 잔차 학습(Residual Learning)의 개념

논문은 네트워크가 입력값을 직접 학습하는 것이 아니라, 입력값과 출력값의 차이(잔차)만을 학습하도록 유도하는 방법을 제안합니다.

즉, 네트워크가 $H(x)$를 직접 학습하는 것이 아니라, $F(x)=H(x)-x$를 학습하도록 만듭니다.

이를 수식으로 표현하면 다음과 같습니다. → $y=F(x)+x$

여기서 x는 입력값, $F(x)$는 여러 레이어를 거쳐 계산된 잔차, y는 최종 출력입니다.

이 구조를 **잔차 블록(Residual Block)**이라고 부릅니다.

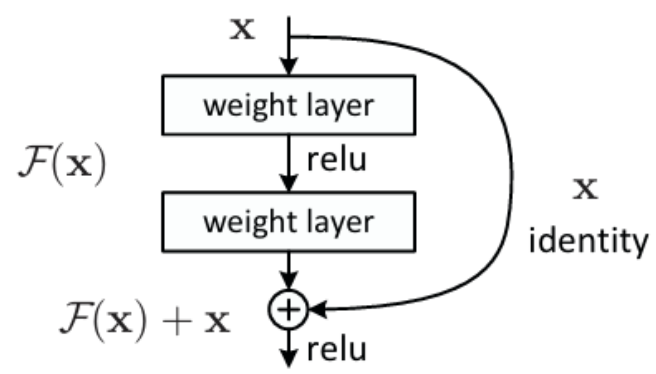
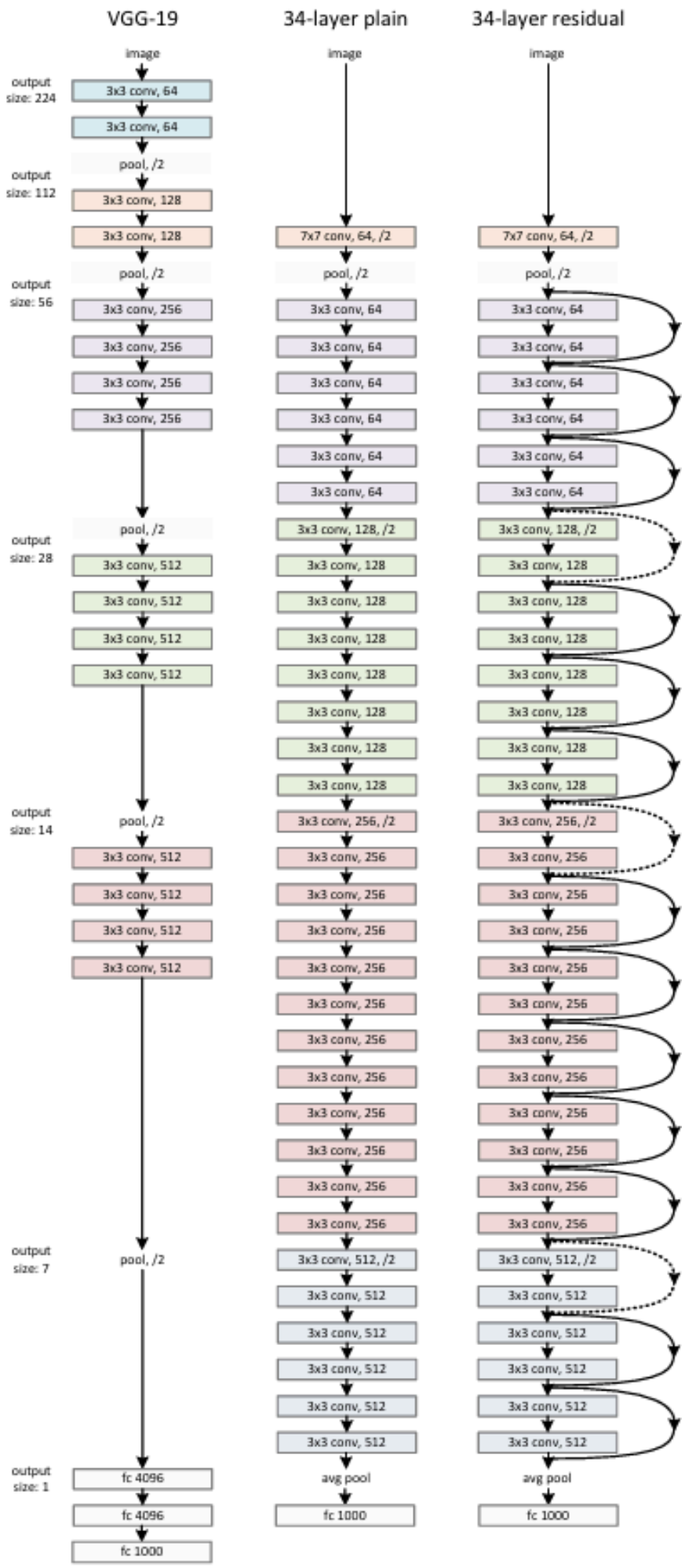


Figure 2. Residual learning: a building block

4. 네트워크 구조



논문에서는 여러 가지 깊이의 네트워크(예: 34, 50, 101, 152층)를 제안합니다.

ResNet의 기본 구조는 다음과 같습니다.

- **입력 이미지**: 224x224
- **초기 컨볼루션**: 7x7, stride 2, 64 채널
- **Max Pooling**: 3×3, stride 2
- **잔차 블록**: 3x3 컨볼루션을 여러 번 반복 (총 34층)
- **평균 풀링(Average Pooling)**
- **완전 연결층(Fully Connected Layer)**
- **Softmax**

논문에서는 VGG-19, plain 34층, residual 34층 네트워크의 구조를 비교합니다.

ResNet은 plain 네트워크와 동일한 파라미터 수와 연산량을 유지하면서, shortcut connection을 추가해 성능을 크게 높였습니다.

---

5. 구현 및 학습 세부사항

논문에서는 ImageNet 데이터셋에 대해 다음과 같은 구현 방식을 사용합니다.

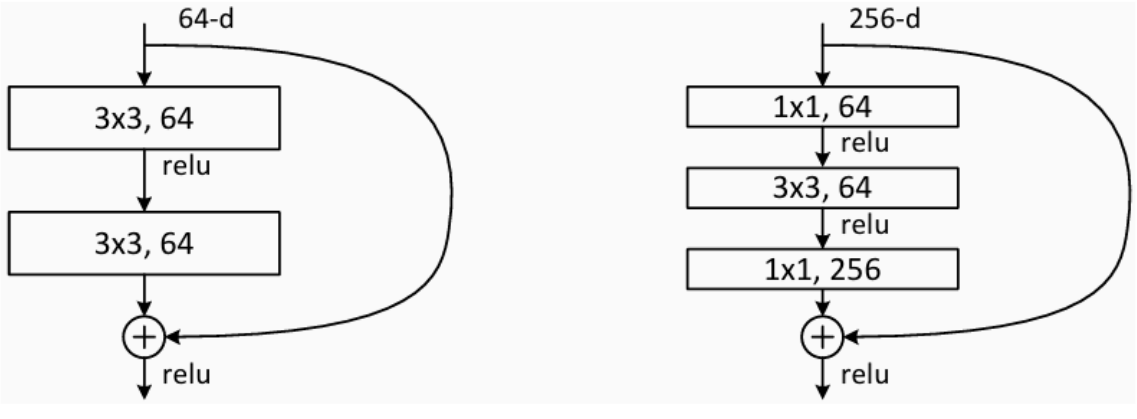
- **이미지 전처리**: 224x224 크롭, 랜덤 플립, 색상 증강
- **배치 정규화**: 모든 컨볼루션 후 적용
- **최적화**: SGD, 미니배치 256, 초기 학습률 0.1, 오류가 수렴하면 10으로 나눔
- **가중치 감쇠**: 0.0001, 모멘텀: 0.9
- **드롭아웃 미사용**

---

6. Bottleneck 구조

ResNet-50/101/152에서는 Bottleneck 구조를 사용합니다.

이 구조는 1x1 → 3×3 → 1×1 컨볼루션을 사용하여 연산량을 줄이고, 매우 깊은 네트워크에서도 안정적으로 학습할 수 있게 합니다.



Bottleneck 블록 구조

---

7. 실험 결과

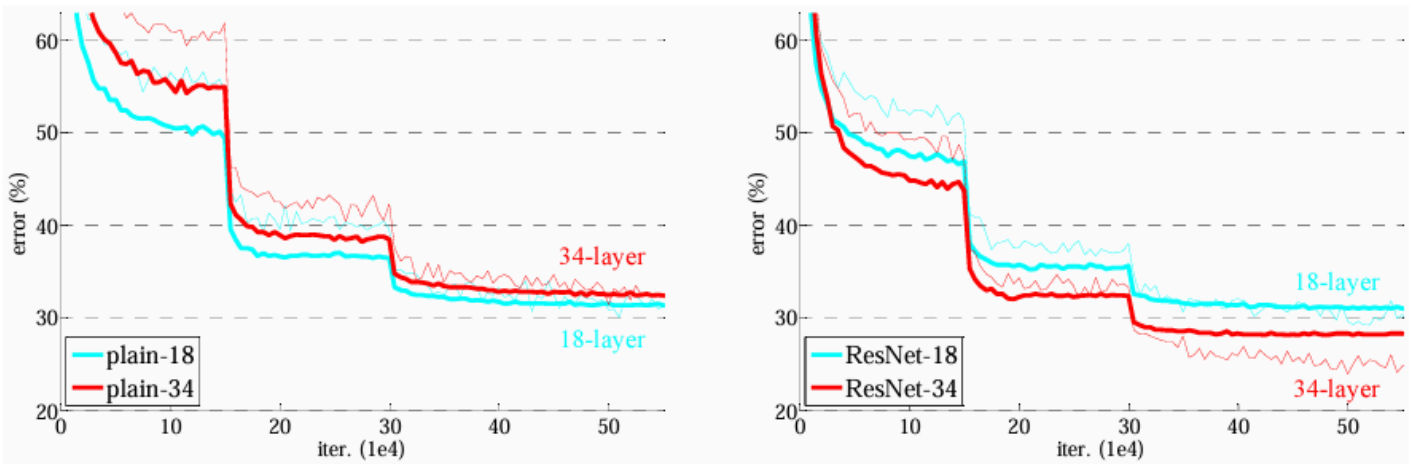
논문에서는 다양한 깊이의 ResNet(예: ResNet-34, ResNet-50, ResNet-101, ResNet-152)을 실험했습니다.

주요 결과는 아래 표와 같습니다.

모델	층 수	ImageNet Top-5 오류율	FLOPs (10억)
VGG-16	16	7.5%	19.6
Plain Net-34	34	10.2%	3.6
ResNet-34	34	**7.5%**	3.6
ResNet-152	152	**4.49%**	11.3

ResNet-34는 동일 층 수의 일반 네트워크(Plain Net-34)보다 **2.7%p** 더 좋은 성능을 보였습니다.

ResNet-152는 ImageNet에서 4.49%의 오류율을 기록하며, 당시 최고 성능을 달성했습니다.



ImageNet에서 plain/residual 네트워크의 학습/검증 오류율 비교 그래프

---

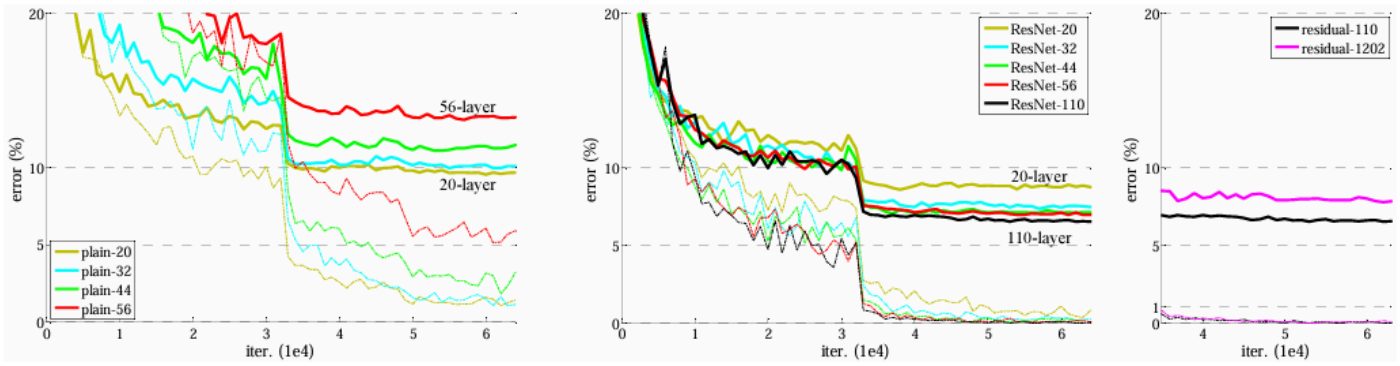
8. 추가 실험 및 분석

8.1. Identity vs. Projection Shortcut

논문에서는 shortcut connection에 대해 **identity mapping(항등 연결)**과 **projection(투영 연결)**을 비교합니다. 실험 결과, identity mapping이 충분히 효과적이며, projection은 추가 파라미터를 도입해 성능은 약간 높아질 수 있지만, complexity가 증가합니다.

8.2. 깊은 네트워크의 학습 곡선

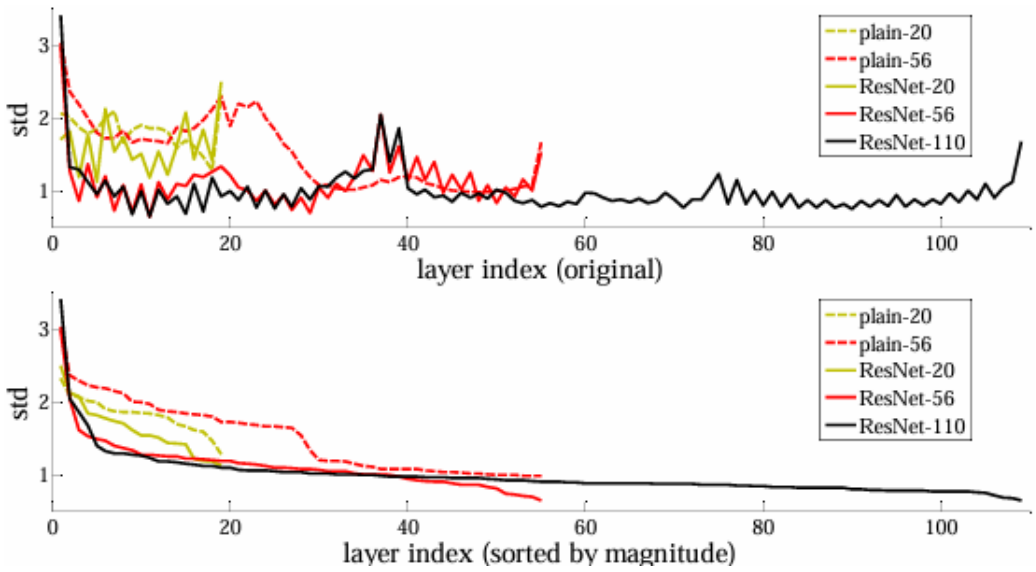
CIFAR-10에서 110/1202층 네트워크도 안정적으로 학습되었으며, plain 네트워크는 깊이가 깊어질수록 학습이 어렵지만, ResNet은 깊이가 깊어져도 학습이 잘 됩니다.



CIFAR-10에서 plain/residual 네트워크의 학습/테스트 오류율 비교 그래프

8.3. Layer Response 분석

논문에서는 각 레이어의 출력 분포를 분석하여, ResNet의 잔차 함수가 작은 값을 가지는 경향을 보임을 확인합니다. 이는 identity mapping이 좋은 preconditioning 역할을 한다는 것을 의미합니다.



CIFAR-10에서 각 레이어의 출력 표준편차 그래프

9. 다양한 태스크에서의 성능

ResNet은 이미지 분류뿐 아니라, **객체 검출, 세그멘테이션** 등 다양한 컴퓨터 비전 태스크에서도 뛰어난 성능을 보입니다. 특히 COCO 객체 검출 벤치마크에서 VGG-16 대비 28%의 상대적 성능 향상을 보였습니다.

10. 결론

"Deep Residual Learning for Image Recognition" 논문은 **딥러닝 네트워크의 학습 난이도를 근본적으로 해결**한 혁신적인 연구입니다. ResNet은 단순한 아이디어로도 큰 효과를 낼 수 있음을 보여주며, 딥러닝 연구자들에게 영감을 주었습니다. 앞으로도 ResNet의 아이디어는 다양한 분야에서 계속 활용될 것입니다.