

Deep Residual Learning for Image Recognition pre-report

Junseo Park
School of Computer Science and Engineering, Pusan National University.

1 Introduction

딥러닝 모델의 성능은 네트워크 깊이에 크게 의존하지만, 단순히 층을 깊게 쌓는 것은 최적화의 어려움과 **degradation problem**을 유발한다. 이 문제는 깊이가 증가할수록 오히려 학습 오류가 커지고 정확도가 저하되는 현상을 의미한다. 본 논문은 이를 해결하기 위해 **Residual Learning**을 도입하여, 매우 깊은 네트워크에서도 안정적인 학습을 가능하게 하였다. ResNet은 ILSVRC 2015에서 탁월한 성능을 기록하며 새로운 딥러닝 아키텍처 패러다임을 제시하였다.

2 Proposed Method

본 논문은 깊은 네트워크에서 발생하는 퇴화 문제를 해결하기 위해 **Deep Residual Learning** 프레임워크를 제안한다. 기존 방식은 여러 층을 단순히 쌓아 원하는 매핑 $\mathcal{H}(x)$ 를 직접 근사하려 했으나, 본 방법은 각 층이 잔차 함수 $\mathcal{F}(x) := \mathcal{H}(x) - x$ 를 학습하도록 설계한다. 따라서 최종 매핑은 $\mathcal{F}(x) + x$ 의 형태로 재구성된다.

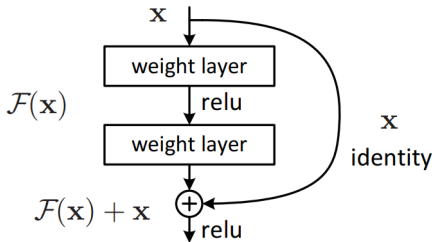


Figure 1: Residual Block: 층이 학습한 잔차 함수 $\mathcal{F}(x)$ 와 입력 x 를 더해 최종 출력 $y = \mathcal{F}(x) + x$ 를 형성한다.

2.1 Shortcut Connections

잔차 블록의 핵심은 **지름길 연결(shortcut connections)**이다. 입력 x 를 항등 매핑으로 그대로 전달하여 층의 출력과 더하며, 다음과 같은 특징을 가진다:

- **추가 매개변수와 계산 복잡도 없음:** 단순히 입력을 나중에 더하는 연산만 수행하므로, 동일한 층 수와 너비를 가진 일반 신경망과 공정한 비교가 가능하다.
- **Projection Shortcut:** 입력과 출력의 차원이 다를 경우 1×1 convolution을 통해 차원을 맞춘다. 이는 최소한의 추가 매개변수로 차원 불일치를 해결한다.

2.2 Key Hypothesis

핵심 가설은 잔차를 최적화하는 것이 원래 매핑을 직접 최적화하는 것보다 더 쉽다는 점이다. 만약 항등 매핑이 최적이라면, 층들의 가중치를 0으로 수렴시켜 $\mathcal{F}(x) \approx 0$ 으로 만들 수 있다. 이 경우 해당 블록은 항등 함수처럼 동작하여, 네트워크는 불필요한 층을 복잡한 최적화 과정 없이 자연스럽게 “건너뛰는(skip)” 효과를 얻게 된다. 따라서 잔차 학습은 매우 깊은 네트워크의 안정적인 학습을 가능하게 한다.

3 Network Architecture

- **Plain Network:** VGG 설계 철학을 기반으로 한 단순한 convolutional network로, 대부분 3×3 convolution 층으로 구성된다. feature map 크기가 절반으로 줄어들면 필터 수를 두 배로 늘리며, 마지막은 global average pooling, fully connected layer, softmax로 연결된다.

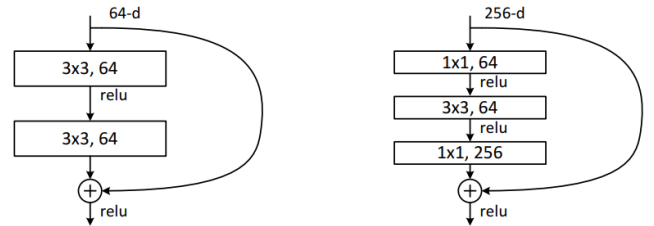


Figure 2: ResNet의 두 가지 블록 구조: Basic Block, Bottleneck Block

- **Residual Network:** Plain network에 지름길 연결을 추가한 구조이다. 얇은 네트워크(18, 34층)에서는 두 개의 3×3 convolution으로 이루어진 Basic Block을 사용하고, 깊은 네트워크(50, 101, 152층)에서는 $1 \times 1 - 3 \times 3 - 1 \times 1$ 구조의 Bottleneck Block을 사용해 계산 효율성을 높인다. 지름길 연결은 차원이 같을 때는 identity 연결, 다를 때는 1×1 convolution을 이용한 projection 연결을 적용한다.

4 Experimental Results

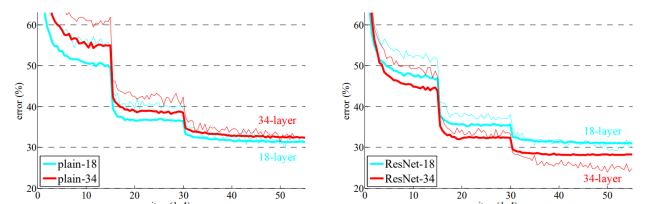


Figure 3: ImageNet 학습 곡선 비교: 동일한 깊이에서 plain network는 퇴화 문제를 보이지만, ResNet은 안정적으로 최적화된다.

- **ImageNet:** Plain network는 깊어질수록 성능이 떨어지는 **degradation** 문제가 나타났지만, ResNet은 깊어질수록 오류가 줄어들었다. 34층 ResNet은 18층보다 더 높은 정확도를 보였으며, 50, 101, 152층 모델에서는 Bottleneck Block을 통해 효율적 학습이 가능했다. 단일 152층 ResNet은 top-5 error 4.49%를 기록하여 당시 최고 성능을 달성하였다.
- **CIFAR-10:** ImageNet에서 관찰된 degradation 문제는 CIFAR-10에서도 동일하게 확인되었다. ResNet은 깊어질수록 오류가 감소하였고, 1000층 이상에서도 training error가 거의 0에 수렴해 **최적화 문제는 해결**되었다. 그러나 1202 layers ResNet에서는 testing error가 오히려 증가하여, 이는 **overfitting** 발생과 정교한 정규화 기법의 부재 때문으로 추정된다.
- **PASCAL and MS COCO:** ResNet은 객체 탐지에도 효과적이었다. Faster R-CNN의 backbone을 VGG-16에서 ResNet-101로 교체했을 때, COCO 데이터셋에서 mAP가 6.0% 증가(상대적으로 28% 향상)하여 **Residual Learning의 범용성**을 입증하였다.

5 Conclusion

ResNet은 깊은 네트워크에서 발생하는 **degradation** 문제를 해결하기 위해 **Residual Learning**을 도입한 아키텍처로 제안되었다. 지름길 연결을 통해 초심층 네트워크의 최적화를 단순화하였으며, ImageNet, CIFAR-10, COCO 등 다양한 벤치마크에서 기존 모델을 능가하는 성능을 보였다. 이로써 ResNet은 매우 깊은 네트워크가 실제로 효과적으로 동작할 수 있음을 보여주며, 딥러닝 모델 설계의 중요한 전환점을 마련하였다.

6 References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [2] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICLR)*, 2015.