Deep Residual Learning for Image Recognition summary

Summarized by: 202355514 강지원

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

Microsoft Research

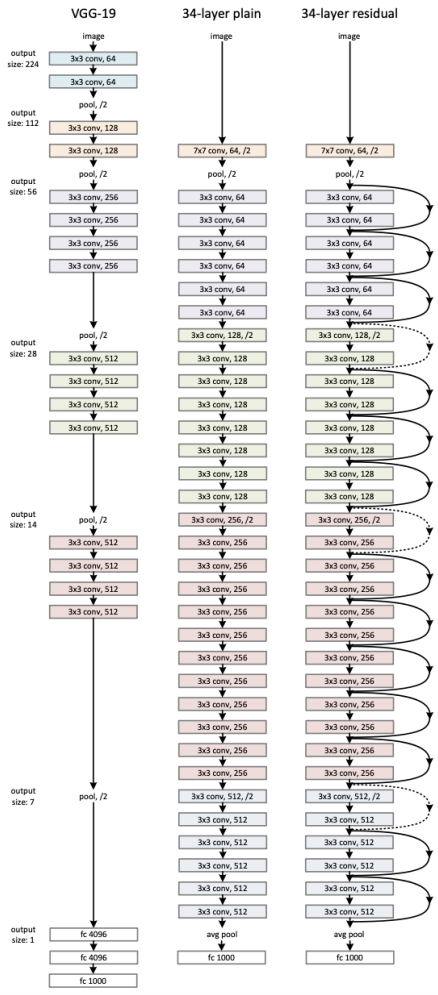

Figure 1: Example network architectures for ImageNet.

DCNN(Deep Convolutional Neural Networks)[1]는 CNN을 층으로 깊게 쌓은 네트워크 구조로, 층을 깊게 쌓을수록 표현력이 강화된다. 최근 연구들은 네트워크 깊이(Depth)가 성능 향상에서 핵심적인 역할을 한다는 것을 증명했으며, 실제로 ImageNet과 같은 대규모 데이터셋에서도 16~30층 이상의 깊은 모델들이 우수한 성능을 보여주었다. 이에 본 논문은 "층을 단순히 깊게 늘리면 더 나은 네트워크가 되는가?"라는 질문에서 출발한다.

기존에는 깊은 네트워크 학습에서 기울기의 소실 및 폭발이 학습의 수렴을 방해하는 큰 문제였으나, 최근 Normalized Initialization과 Batch Normalization 기법 덕분에 깊은 네트워크 학습이 가능해졌다. 그러나 네트워크가 깊어질수록 성능저하(Degradation)가 발생하는데, 이는 깊이가 증가함에 따라 정확도는 오르다가 일정 시점 이후 포화상태에 머무르고, 이후에는 오히려 떨어지게 되는 현상이다. 이러한 성능저하는 단순 과적합 문제가 아니라, 깊이가 깊어질수록 훈련 자체가 어려워지기 때문이다.

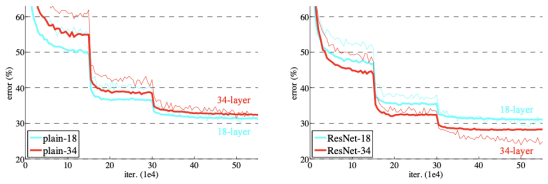

Figure 2: Training on ImageNet. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

이를 해결하기 위해 논문은 Residual Learning을 제안한다. 원하는 함수 $H(x)$를 직접 학습하는 대신, 잔차 함수 $F(x) = H(x) - x$를 학습하도록 재구성한다. 이는 입력을 그대로 전달하는 항등(Identity) Shortcut Connection을 통해 구현되며, 추가적인 연산이나 파라미터 없이 학습 난이도를 낮추고 성능을 향상시킨다. (기존의 깊은 모델 VGG-19 구조, 34-layer Plain 구조, Plain Net에 Shortcut Connection을 추가한 ResNet 구조 예시 이미지는 Fig 1을 통해 참고할 수 있다.)

실험 결과, 단순히 층을 쌓은 Plain Net은 깊어질수록 training error가 증가하는 반면, Residual Net은 34-layer ResNet이 18-layer ResNet보다 2.8% 더 높은 성능을 보였고, plain 34-layer 대비 top-1 error를 3.5% 감소시켰다. 이는 ResNet이 깊어질수록 정확도와 수렴 속도에서 이점을 가지며 최적화가 잘 이루어짐을 보여준다. (해당 실험의 결과는 Fig 2에서 확인할 수 있다.)

기존 연구로 제안된 Highway Network는 게이트 함수가 닫히면 Residual Learning이 이루어지지 않아 깊은 네트워크에서 효과적이지 않았다. 반면, ResNet은 단순하고 파라미터가 없는 Identity Shortcut Connection을 도입해 깊은 네트워크 학습 문제를 근본적으로 해결하였다. 또한 차원이 맞지 않을 경우 Linear Projection을 활용해 효율적으로 구현할 수 있으며, bottleneck block[2]을 통해 연산량을 줄이면서도 깊이를 늘릴 수 있는 구조를 설계할 수 있다.

더 나아가, 152-layer ResNet 모델은 VGG를 제치고 적은 연산량과 높은 정확도로 ILSVRC 대회에서 top-5 error 3.57%를 기록하며 1위를 차지했다. 이로써 ResNet은 ImageNet과 CIFAR-10 등 특정 데이터셋에 국한되지 않고, 다양한 데이터셋에서 압도적 성능을 입증했다. 또한 분류뿐 아니라 객체 탐지, 세분화 등 다양한 비전 과제에서 탁월한 성능을 보이며, 범용적인 원리로서의 가능성을 보였다.

1000층이 넘는 매우 깊은 네트워크 모델도 최적화는 가능했으나 작은 데이터셋에서는 과적합이 발생해 오히려 성능이 저하되었다. 따라서 향후에는 dropout, maxout 등 기존의 BN보다 더 강력한 정규화 기법이 요구될 것으로 보인다.

본 보고서는 Deep Residual Learning for Image Recognition[3] 을 요약한 것이다.

[1] DanielCS. Deep convolutional neural networks 설명. https://danielcs.tistory.com/114, 2020.

[2] Lighthouse97. Cnn의 bottleneck에 대한 이해. https://velog.io/@lighthouse97/CNN%EC%9D%98-Bottleneck%EC%97%90-%EB%8C%80%ED%95%9C-%EC%9D%B4%ED%95%B4, 2021. Velog Blog.

[3] Xiangyu Zhang, Kaiming He, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. 2016.