

ResNet: Deep Residual Learning for Image Recognition

* 더 나은 네트워크를 학습하는 것이 더 많은 층을 쌓은 것만큼 쉬운가에 대한 질문에 답을 하는 것은 vanishing/exploding gradient의 문제 때문에 힘들었다.

→ 문제는 normalized initialization과 intermediate normalization layers에 의해 해결 가능

* degradation 문제를 deep residual learning framework를 소개함으로써 해결함

* shortcut connection은 단순히 identity mapping을 실행하고 그들의 결과는 stacked layer의 결과에 더함

* highway network은 gating function과 함께 short connection을 보여줌

→ 이 gate들은 data-dependent이며 파라미터들을 가지고 있음

→ identity shortcut들이 parameter-free인것과 대비됨

* gated shortcut이 닫혀 있을 때(0에 접근할 때, highway 네트워크의 레이어들은 non-residual 기능을 대표함

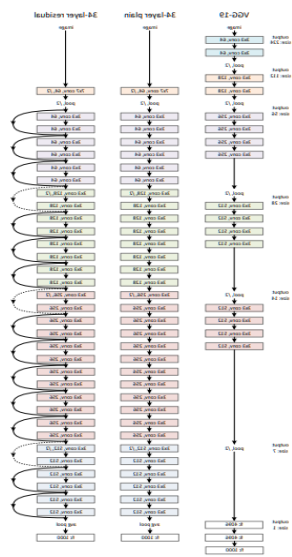
→ 반면, 논문의 식은 항상 residual 기능을 학습함. 또한 identity shortcut은 절대 안 닫히고 모든 정보는 학습되어야 될 추가적인 잔여 기능과 함께 통과됨

→ 추가적으로 highway network는 굉장히 증가된 깊이와 함께 정확성 증가를 입증하지 않음

* multiple nonlinear layer들이 점근적으로 복잡한 함수들과 비슷하게 할 수 있다고 가정하면, 그것은 그들이 residual function들과 근사적으로 비슷하게 할 수 있다고 가정하는 것과 같음

→ stacked layer들이 $H(x)$ 를 비슷하게 하는 것을 예상하는 것보다 이 layers가 residual function($F(x) := H(x)-x$)을 명시적으로 비슷하게 함 → original function은 $F(x)+x$ 가 됨

* 만약 optimal function이 zero mapping보다 identity mapping에 가깝다면, function을 새로 배우는 것보다 solver가 identity matrix와 참조하여 작은 변화를 찾는 것이 쉬움



* ImageNet의 네트워크 아키텍처

- 왼쪽: reference로써 VGG-19 모델
- 가운데: 24 파라미터의 레이어를 가진 plain network
- 오른쪽: 24 파라미터 레이어를 가진 residual network
- 점선의 지름길은 차원을 증가시킴.