Paper Review 1

Deep Residual Learning for Image Recognition

https://arxiv.org/abs/1512.03385

본 논문은 깊이가 깊어짐에 따라 훈련을 용이하게 하는 residual learning framework를 설명하고, 이를 적용하여 이미지 인식을 위한 딥러닝 모델로써 ResNet(Residual Network) 구조를 제안하고 있다.

1. Introduction

"Deep Convolution Neural Networks"는 이미지 분류에서 큰 발전을 가져왔다. 딥러닝 모델의 네트워크의 깊이(Network depth)는 매우 중요하고, 최근 연구들에서는 매우 깊은 모델들이 큰 성과를 가져왔다. 하지만 기존 모델들에서는 깊이가 증가함에 따라 딥러닝 모델을 훈련시키기가 어려웠다.

깊이가 깊어질수록 성능이 저하되는 문제인 degradation problem을 해결하기 위해 본 연구는 residual learning framework를 제안한다. 이는 네트워크의 각 층이 직접 목표 함수를 학습하는 대신 입력값과의 차이인 잔차 함수(residual function)을 학습하도록 층을 재구성한다.

2. Residual Learning Framework

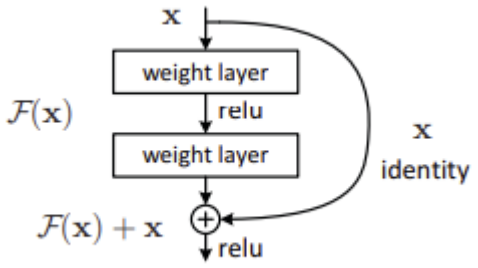


Figure 2. Residual learning: a building block.

잔차 함수인 F(x) = H(x)−x를 학습하고, 최종적으로 F(x) + x를 출력한다.

이 방법은 네트워크가 더 깊어질수록 더 높은 정확도를 얻을 수 있다. 또한, shortcut connections를 통해 layer를 연결하여 계산 복잡도를 덜었다. 이는 입력 x를 출력으로 직접 연결하여 추가적인 매개변수나 계산 복잡도 없이 학습하도록 한다. Identity shortcut 과 projection shortcut 간에는 차이가 크지 않으므로, 메모리 및 시간 복잡도를 줄이기 위해 identity shortcut을 사용하는 것이 효율적이다.

3. Residual Network (ResNet)

Residual learning framework를 활용하여 설계된 딥러닝 모델로, 일반적인 네트워크에 shortcut connection을 추가하여 변환한 구조이다.

입력과 출력의 차원이 동일하면 identity shortcut 을 사용하고,

차원이 증가하면

(a) zero-padding을 통해 추가적인 매개변수 없이 차원을 증가시키는 identity mapping을 하거나,

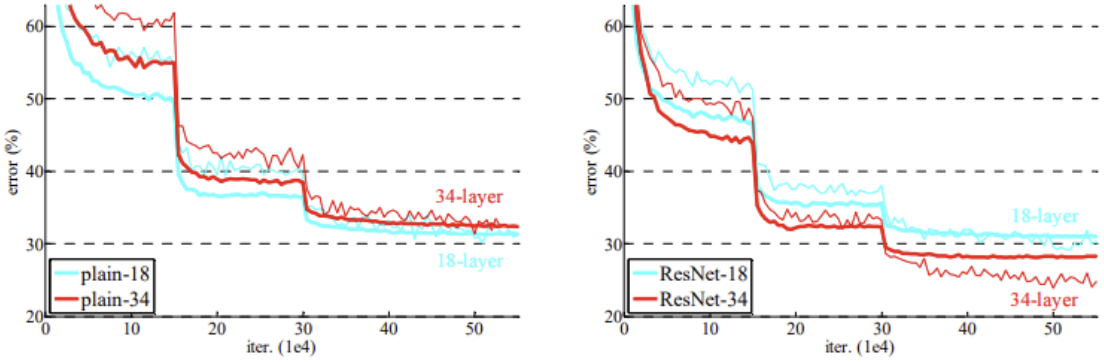
(b) 1×1 합성곱을 통해 projection shortcut을 이용할 수 있다.

(a)와 (b)에서는 shortcut이 두 가지 크기의 feature map을 가로지를때 stride 2를 사용한다.

4. Experiments

- ImageNet classification dataset

1.28M의 training image, 50k의 validation image, 100k의 test image를 1000개의 class로 분류하는 실험.

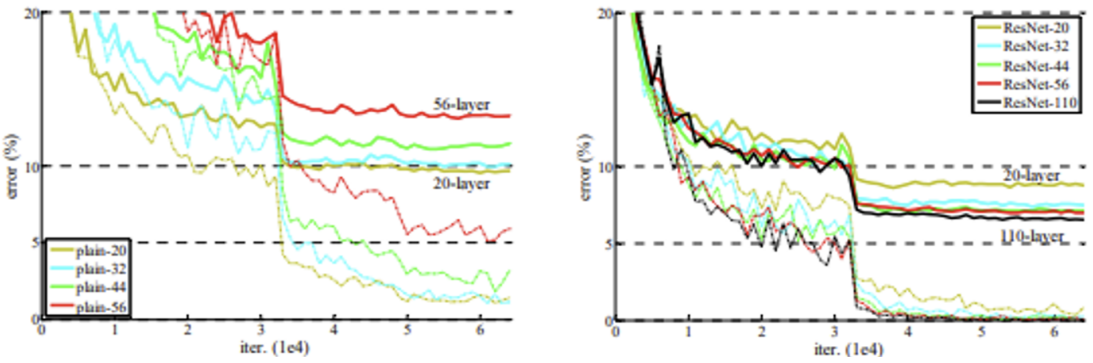


Bottleneck Architecture를 사용하였다. 기존 블록은 잔차 함수 F를 2개의 layer로 구성하는 것에 비해 bottleneck design 블록은 F를 3개의 layer로 구성한다. 각 layer는 1x1, 3×3, 1×1의 합성곱 층으로 이루어져 있다.

50 / 101 / 152개의 layer를 가진 각각의 ResNet을 비교하였을 때 degradation problem이 나타나지 않았다. 152-layer의 ResNet을 학습하였더니 3.57%의 낮은 test error를 기록하였다.

- CIFAR-10 dataset

50k의 training image와 10k의 testing image를 10개의 class로 분류하는 실험.



100개 layer의 ResNet은 learning rate 조정 후 좋은 성능을 보였으며, 1000개의 layer에서도 마찬가지로 높은 성능을 보였다. 일반 네트워크(FitNet, Highway)에서는 깊이가 깊어질수록 training error가 증가하는 degradation problem이 발생하였지만, ResNet에서는 이를 극복하여 깊이가 깊어질수록 정확도가 증가하였다.

110-layer ResNet은 test error로 6.43%, 1202-layer ResNet은 training error는 0.1%보다 낮고 test error는 7.93%를 기록하였다.

- PASCAL and MS COCO dataset

객체 탐지(object-detection) 실험.

training data	07+12	07++12
test data	VOC 07 test	VOC 12 test
VGG-16	73.2	70.4
ResNet-101	**76.4**	**73.8**

Table 7. Object detection mAP (%) on the PASCAL VOC 2007/2012 test sets using **baseline** Faster R-CNN. See also Table 10 and 11 for better results.

metric	mAP@.5	mAP@[.5, .95]
VGG-16	41.5	21.2
ResNet-101	**48.4**	**27.2**

Table 8. Object detection mAP (%) on the COCO validation set using **baseline** Faster R-CNN. See also Table 9 for better results.

각 task에서는 탐지 방법으로 Faster R-CNN 을 사용하였으며, 모두 ResNet이 더 좋은 성능을 보였다. 특히 COCO dataset 에서는 VGG-16 모델을 쓴 것보다 ResNet-101 모델을 사용한 것이 28% 향상된 성능을 보였다.

5. Conclusion

본 연구는 입력값 기반의 잔차 함수를 학습하는 Residual learning framework를 활용하여 degradation problem을 해결할 수 있음을 설명하고 있다. ImageNet classification, CIFAR-10 classification, Pascal / MS COCO object-detection의 task를 통해 layer의 depth가 증가함에 따라 ResNet이 기존 plain net보다 좋은 성능을 나타내는 것을 증명하였다.