ResNET Review

오연정

oyj2679@g.skku.edu

Computer Vision

2023/03/06

TRAIN AND TEST

Contents

TRAIN AND TEST

Abstract

ResNET

: Residual Network

reformulate the layers as learning residual functions with reference to the layer inputs

TRAIN AND TEST

Abstract

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Table 1. Architectures for ImageNet. Building blocks are shown in brackets (see also Fig. 5), with the numbers of blocks stacked. Down-sampling is performed by conv3_1, conv4_1, and conv5_1 with a stride of 2.

TRAIN AND TEST

Introduction

Problem & Theory

More layer -> gradients 가 vanish/explode

-> 초기 10개의 레이어에 stochastic gradient descent with back-propagation 함수를 사용함

F = H − x -> H = F + x : identity mapping 이 이상적이라면 residual을 0으로 만들 것.

-> Shortcut connections

Theory with shortcut

(1) resnet이 plain net에 비해 optimize하기가 쉽다

(2) 이전 네트워크들에 비해 depth(레이어 수)가 커져도 정확성 향상에 더 도움이 된다

TRAIN AND TEST

Deep Residual Learning

Residual Learning

F = H(x) -x : identity mappings -> reformulation

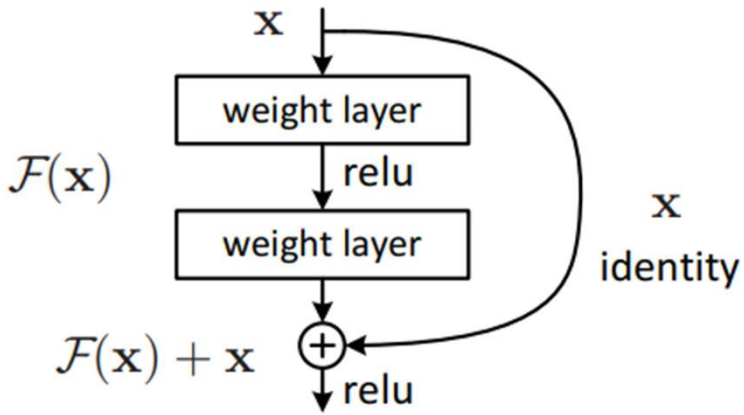


Figure 2. Residual learning: a building block.

TRAIN AND TEST

Deep Residual Learning

Identity mapping Shortcuts

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}. \quad (1)$$

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s\mathbf{x}. \quad (2)$$

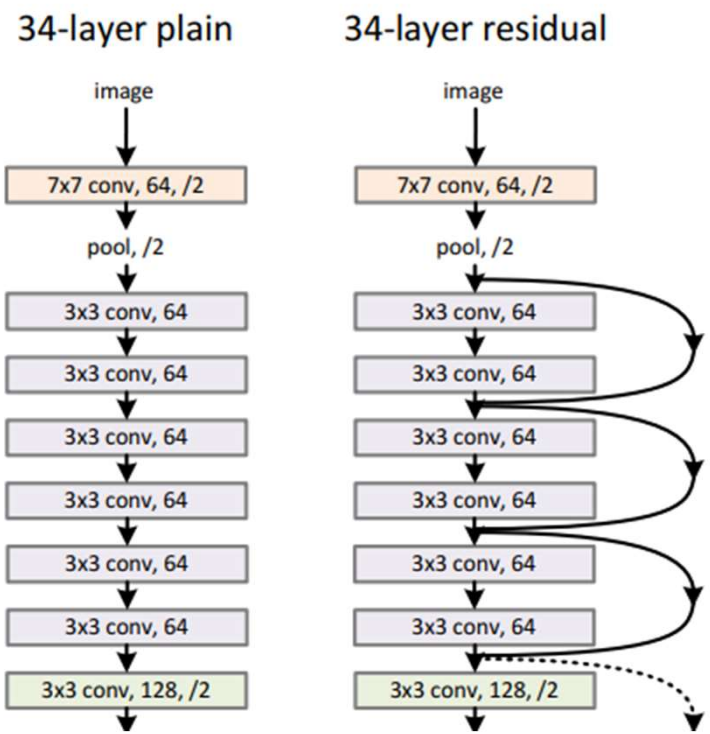
y = W1x + x -> no effect

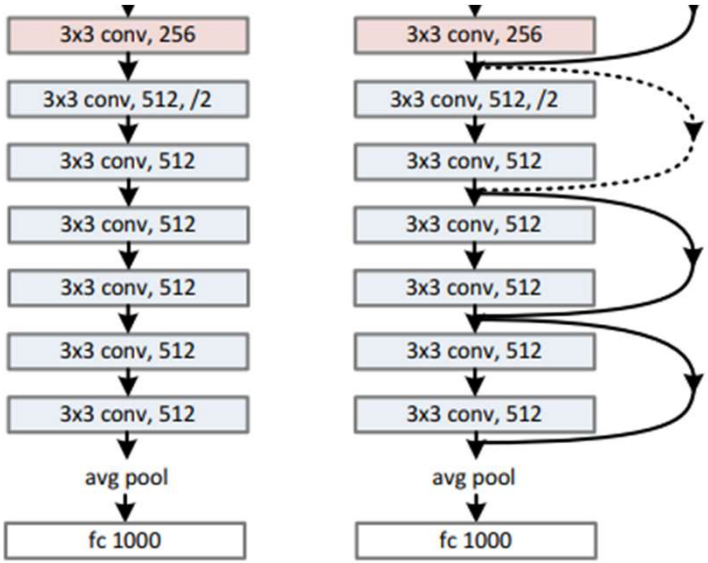
convolutional layer 반영

TRAIN AND TEST

Deep Residual Learning

Network Architectures





Same dimension -> just use
Different dimension -> zero padding or projection

Experiments

Implementation - ImageNet

Background

-224x224, [256,480] scale

-batch normalization(convolution 직후, activation 이전)

-SGD, mini-batch 256size, learning rate 0.1, 10 에러한계, 60x 10^4 iteration

-1000 클래스를 가진 데이터셋, 128만 training, 5만 validation, 10만 test images

(1) plain networks(18, 34 layer 비교) : depth error, degration prob 둘다 나타남

(2) Residual Networks(18, 34 layer 비교) : (zero padding 방법 이용)

[1] 34 layer 가 18 layer 보다 train, vaild error가 낮다.

[2] ResNet 이 plain net 보다 낮다.

[3] 18 layer 사용시 SGD가 괜찮다.(Resnet 이 좀더 빠르게 converge 할 뿐)

TRAIN AND TEST

Experiments

Plain vs ResNET

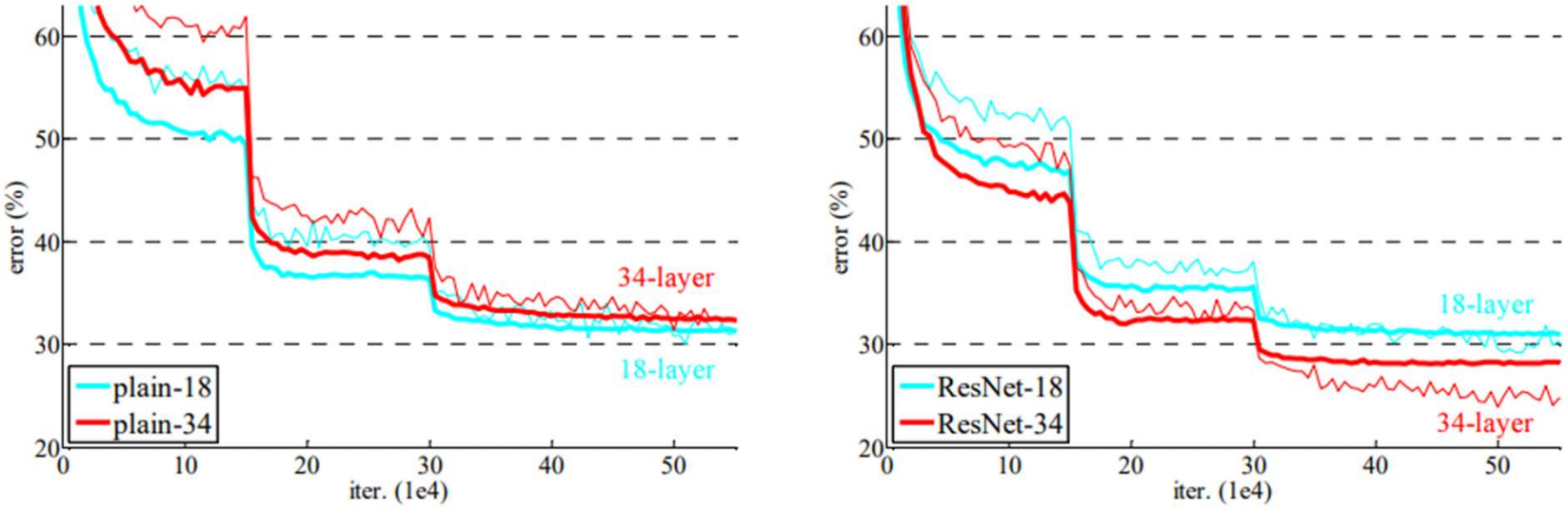


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

TRAIN AND TEST

Experiments

Identity vs Projection

[1] 차원 변하는것만 zero padding, 나머지 동일

[2] 차원 변하는것만 projection, 나머지 동일

[3] 전체 projection

-> 1 < 2 < 3 이지만 거의 비슷함

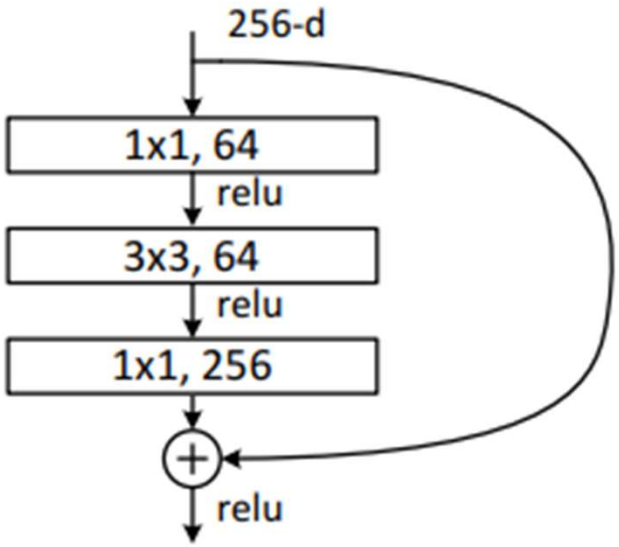
model	top-1 err.	top-5 err.
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40

TRAIN AND TEST

Experiments

Deeper Bottleneck

parameter-free identity shortcuts > projection : Low complexity



model	top-1 err.	top-5 err.
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

TRAIN AND TEST

Experiments

CIFAR-10

method			error (%)
ResNet	20	0.27M	8.75
ResNet	32	0.46M	7.51
ResNet	44	0.66M	7.17
ResNet	56	0.85M	6.97
ResNet	110	1.7M	**6.43** (6.61±0.16)
ResNet	1202	19.4M	7.93

Table 6. Classification error on the **CIFAR-10** test set. All methods are with data augmentation. For ResNet-110, we run it 5 times and show "best (mean±std)" as in [43].

Background

5만개 training, 1만개 test, 10 class

input 32x32, feature map {32,16,8}

layer 2n (n=3,5,7,9) 각각, 총 3n shortcut

weight decaying 0.0001, momentum 0.9, mini-batch 128

learning rate 0.1, 0.01(32k), 0.001(48k), iteration 64k

4px 씩 padding 후 32x32로 랜덤하게 자름

n = 18, 110-layer -> best

TRAIN AND TEST

Experiments

PASCAL & COCO

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.

TRAIN AND TEST

TRAIN AND TEST