ResNet Paper - Deep Residual Learning for Image recognition

Authors: Kaiming He, Xiangyu Zhang, Shaoquing Ren, Jian Sun - Microsoft research

Introduction

Usually when we consider a deep neural network layer, deeper the layer, higher the complexity and higher the difficulty to train the model. In this paper they represent a network which eliminates this problem by introducing residual function and mapping. This network is easier to optimize and can gain higher accuracy with increased depth.

Problem with deep networks

Network depth is an important factor for the result but is learning better networks is as easy as stacking layers?. And there is the vanishing gradient problem which can be solved by having normalization layers which makes the network converge. When deeper networks converge the accuracy can be saturated and then degrades rapidly (Degradation problem). This problem is not because of overfitting and it also leads to higher training error. This problem also indicates that not all models are easy to optimize. This problem can be addressed by identity mapping but then come the time constraints.

Proposed solution

The degradation problem can be solved by a deep *residual learning framework.* The layers will have a residual mapping instead of direct underlying mapping [H(x)]. They hypothesize that this is easier to optimize the residual mapping. This forms a shortcut connection in feedforward neural networks without any additional parameters or computational complexity.

Main results:

- Easy to optimize.
- Increase in depth can increase accuracy.

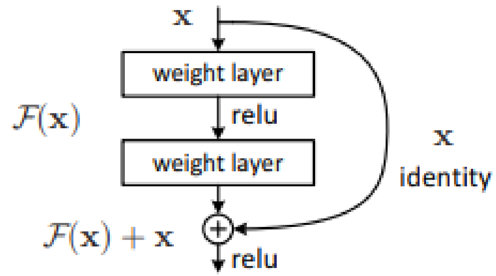


Figure 2. Residual learning: a building block.

Residual Representation

In Low level vision and computer graphics for solving partial differential equations the widely used method is the Multigrid method, but as an alternative a hierarchical basis preconditioning can be used which relies on variables that represent residual vectors which helps the solver to converge at a much higher rate.

Shortcut Connections

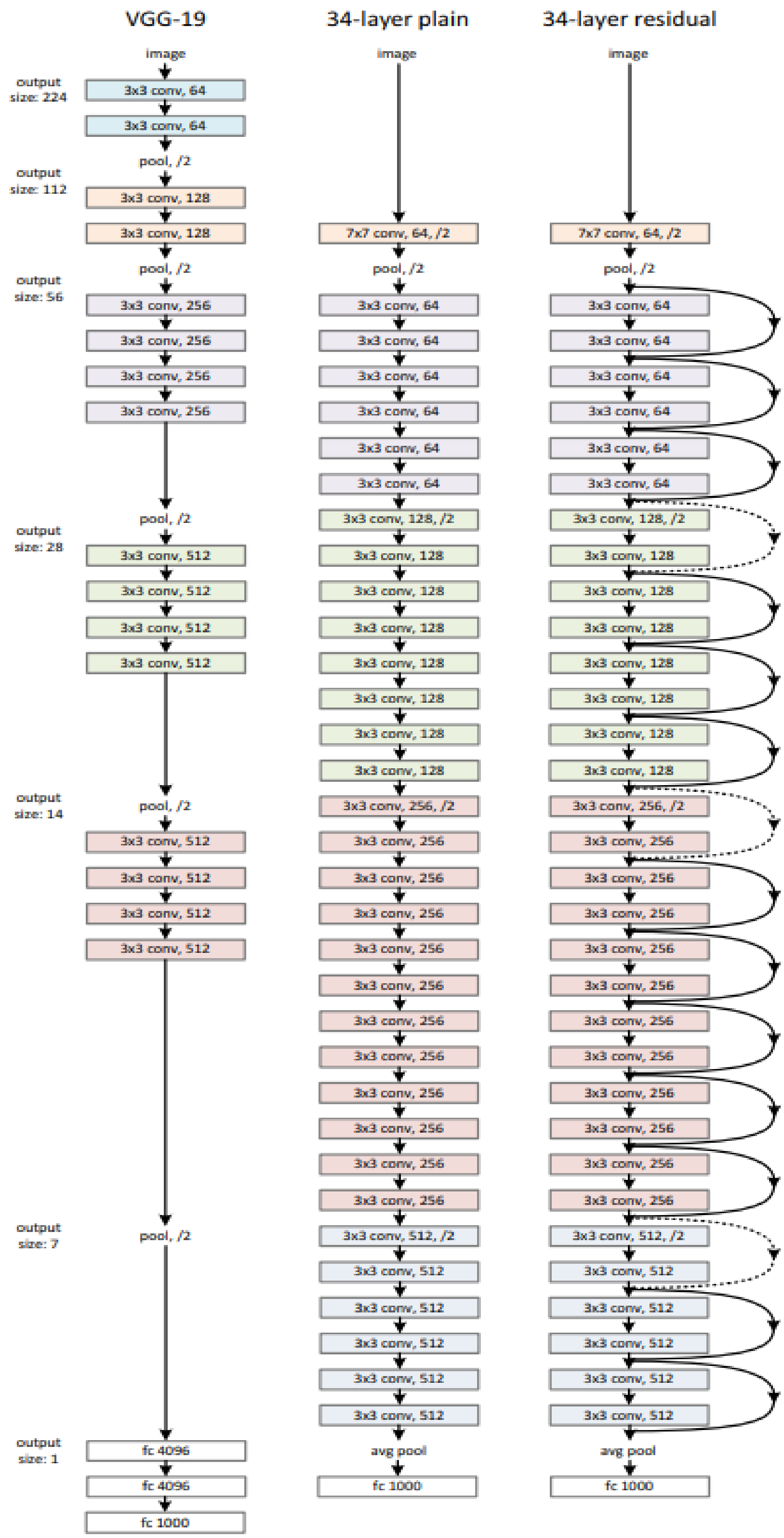
In the earlier days most of the MLPs were linearly connected which gave rise to the vanishing gradient problem. So some proposed an inception layer composed of shortcut and deeper (few) branches with gating functions, but these proposed gating mechanisms have parameters and if it is closed it cannot represent residual functions. This paper proposes a never closed *identity shortcuts* which can pass all the information.

Residual Learning

The underlying mapping tries to approximate the H(x), they explicitly let the layer to approximate a residual function F(x): H(x) - x and thus the original function becomes F(x) + x. Both of these can approximate the function but the difference comes in the ease of learning.

Network Architecture comparison

ResNet has fewer filters and lower complexity.



Experiments and Results

model	top-1 err.	top-5 err.
VGG-16 [40]	28.07	9.33
GoogLeNet [43]	-	9.15
PReLU-net [12]	24.27	7.38
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
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

Table 3. Error rates (%, **10-crop** testing) on ImageNet validation. VGG-16 is based on our test. ResNet-50/101/152 are of option B that only uses projections for increasing dimensions.

method	top-1 err.	top-5 err.
VGG [40] (ILSVRC'14)	-	8.43†
GoogLeNet [43] (ILSVRC'14)	-	7.89
VGG [40] (v5)	24.4	7.1
PReLU-net [12]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

method	top-5 err. (**test**)
VGG [40] (ILSVRC'14)	7.32
GoogLeNet [43] (ILSVRC'14)	6.66
VGG [40] (v5)	6.8
PReLU-net [12]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

Table 5. Error rates (%) of **ensembles**. The top-5 error is on the test set of ImageNet and reported by the test server.

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 appendix 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 appendix for better results.