PADERBORN UNIVERSITY
The University for the Information Society

Communications Engineering

Deep Residual Learning

Seminar Presentation

Varun Nandkumar Golani

Department of Communications Engineering – Paderborn University
Prof. Dr.-Ing. Reinhold Haeb-Umbach

30.03.2021

Introduction

Deep Residual Learning (DRL) [He+16]

New technique to train deeper networks by introducing shortcut connections in the Deep Neural Network(s) (DNN) architecture

Introduction

Deep Residual Learning (DRL) [He+16]

New technique to train deeper networks by introducing shortcut connections in the Deep Neural Network(s) (DNN) architecture

- Can't we use the same architecture as that of a Neural Network(s) (NN)/DNN? - Degradation Problem
- Degradation Problem: Insignificant accuracy gains from the depth or the depth affects the accuracy negatively
- Why do we need DRL? - Solves the degradation problem
- State-of-the-art results on various tasks like image recognition [He+16], speech recognition [HDHU16] etc.

Table of contents

Residual Learning (RL) [He+16]

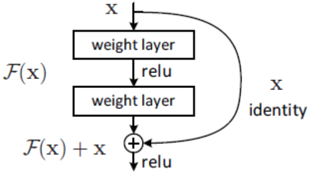

Figure: A building block for RL.
Image Source: [He+16].

- Unknown mapping function $\mathcal{H}(x)$
- Learn residual function $\mathcal{F}(x) = \mathcal{H}(x) - x$ instead of $\mathcal{H}(x)$
- Obtain $\mathcal{H}(x) = \mathcal{F}(x) + x$
- Easier to optimize $\mathcal{F}(x)$ than $\mathcal{H}(x)$
- Two cases for adding $\mathcal{F}(x)$ and x
  - $y = \mathcal{F}(x, \{W_i\}) + x$
  - $y = \mathcal{F}(x, \{W_i\}) + W_s x$

Convolutional Neural Networks (CNN) [WZL18]

Convolution Layer

- Outputs new feature maps by conv. of input images
- Filter size: 3×3, 5×5, 7×7
- Learned filters can extract diff. structures for modeling

Pooling Layer

- Aggregates the input feature maps
- Types: maximum, average
- Can capture large distance correlations in images

Activation Layer

- Nonlinear transformation of the input feature map
- E.g. ReLU, sigmoid etc.
- Extracts more complex correlations in images

Batch Normalization Layer

- Normalizes the data inputs present in a batch $\mathcal{B}$
- Faster speed of convergence
- Not sensitive to parameter initialization

Table of contents

Image Recognition [He+16]

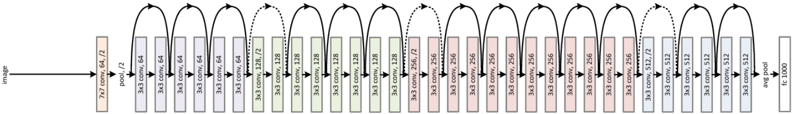

Figure: Architecture of ResNet-34. Image Source: [He+16].

Image Recognition [He+16]

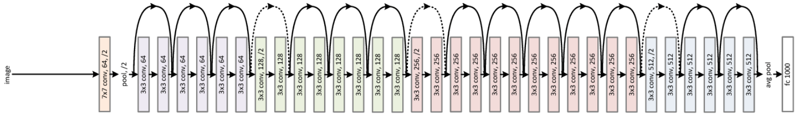

Figure: Architecture of ResNet-34. Image Source: [He+16].

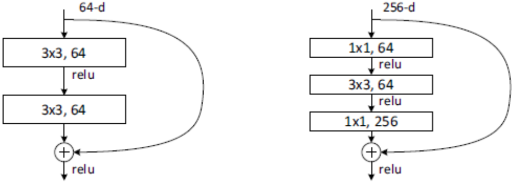

Figure: Left: A non-bottleneck Residual Block (RB). Right: A bottleneck RB used for ResNet-50/101/152. Image Source: [He+16].

Speech Recognition [HDHU16]

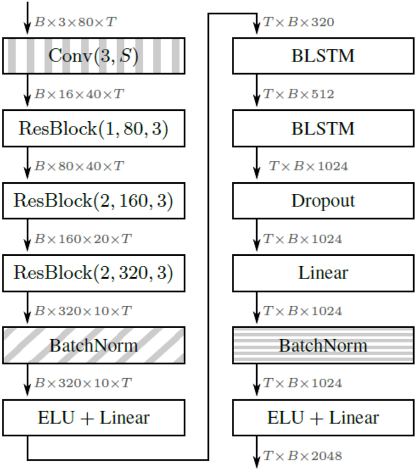

Figure: Wide Residual BLSTM Network(s) (WRBN). B denotes the mini-batch size and T denotes the number of frames. Image Source: [HDHU16].

Speech Recognition [HDHU16]

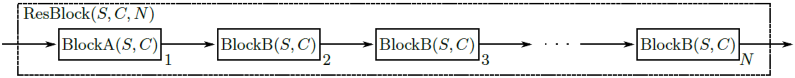

Figure: ResBlock(S,C,N). $S \rightarrow$ stride, $C \rightarrow$ number of output channels, $N \rightarrow$ number of blocks. Image Source: [HDHU16].

Speech Recognition [HDHU16]

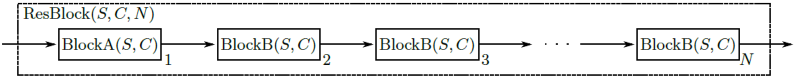

Figure: ResBlock(S,C,N). $S \rightarrow$ stride, $C \rightarrow$ number of output channels, $N \rightarrow$ number of blocks. Image Source: [HDHU16].

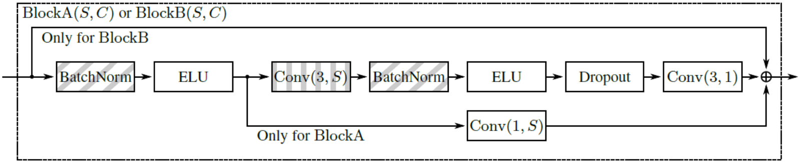

Figure: For Conv(A,S) blocks, $A \rightarrow$ filter size, $S \rightarrow$ consecutive striding, zero padding of size $(A - 1)/2$ in both the directions. Image Source: [HDHU16].

Table of contents

CIFAR-10 Classification [Kri09; He+16]

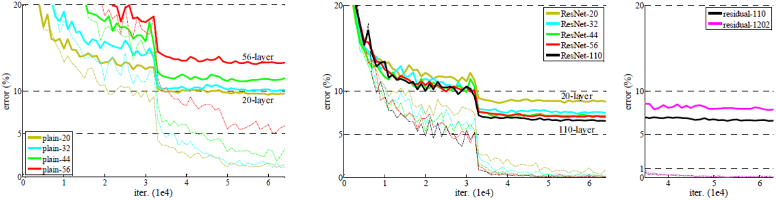

Figure: Dashed lines → training error, bold lines → testing error. Left: Plain Networks. Middle: ResNets. Right: ResNet-110 and ResNet-1202. Image Source: [He+16].

ImageNet Classification [Rus+15; He+16]

Table: **Single-model** results (% error) of ResNets with other baselines on the ImageNet validation set. $\dagger \rightarrow$ reported results are on the ImageNet test set. Table Source: [He+16].

method	top-1 err.	top-5 err.
VGG [SZ15] (ILSVRC'14)	-	$8.43^{\dagger}$
GoogLeNet [Sze+15] (ILSVRC'14)	-	7.89
VGG [SZ15] (v5)	24.4	7.1
PReLU-net [He+15]	21.59	5.71
BN-inception [IS15]	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**

ImageNet Classification [Rus+15; He+16]

Table: Results (% error) of **ensembles**. The top-5 error is communicated by the test server after evaluating the trained model on the ImageNet test set. Table Source: [He+16].

method	top-5 err. (**test**)
VGG [SZ15] (ILSVRC'14)	7.32
GoogLeNet [Sze+15] (ILSVRC'14)	6.66
VGG [SZ15] (v5)	6.8
PReLU-net [He+15]	4.94
BN-inception [IS15]	4.82
ResNet (ILSVRC'15)	**3.57**

Conclusion

- In general, DNN using DRL are easier to optimize as compared to the normal DNN (plain networks) [He+16]
- DNN using DRL can exploit the depth of the DNN which results into more accurate models
- Too deep DNN using DRL can also overfit the data (especially if the dataset is small)
- For instance, ResNet-1202 (7.93%) performs worse (in terms of testing error) than ResNet-110 (6.43%) on the CIFAR-10 dataset [Kri09] and both have almost the same training error [He+16].

Bibliography I

J. Heymann, L. Drude, and R. Haeb-Umbach. "Wide Residual BLSTM Network with Discriminative Speaker Adaptation for Robust Speech Recognition". In: *Computer Speech and Language*. 2016.

K. He et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1026–1034.

K. He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.

Bibliography II

S. Ioffe and C. Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML’15. Lille, France: JMLR.org, 2015, 448–456.

A. Krizhevsky. *Learning multiple layers of features from tiny images*. Tech. rep. 2009.

O. Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *Int. J. Comput. Vision* 115.3 (Dec. 2015), 211–252. ISSN: 0920-5691. DOI: 10.1007/s11263-015-0816-y. URL: https://doi.org/10.1007/s11263-015-0816-y.

K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].

Bibliography III

C. Szegedy et al. "Going Deeper with Convolutions". In: *Computer Vision and Pattern Recognition (CVPR)*. 2015. URL: http://arxiv.org/abs/1409.4842.

S. Wu, S. Zhong, and Y. Liu. "Deep Residual Learning for Image Steganalysis". In: *Multimedia Tools Appl.* 77.9 (May 2018), 10437–10453. ISSN: 1380-7501. DOI: 10.1007/s11042-017-4440-4. URL: https://doi.org/10.1007/s11042-017-4440-4.