Res-SE-Net: Boosting Resnets by Enhancing Bridge-connections

Varshaneya V[1], Balasubramanian S[2], and Darshan Gera[3]

Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam

[1]varshaneya.v@gmail.com
[2]sbalasubramanian@sssihl.edu.in
[3]darshangera@sssihl.edu.in

mind 2019

Introduction

- "AI is the new electricity."
- Popular deep neural networks – Inception net (2014), Resnet (2016), SE-Net (2018).
- Ours is a fundamental research work in the line of proposing a new architecture for improved performance.

mind 2019

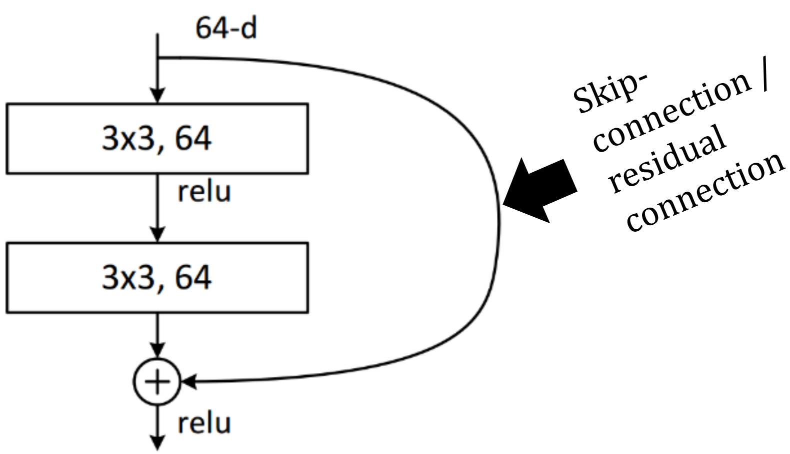
Important papers from literature

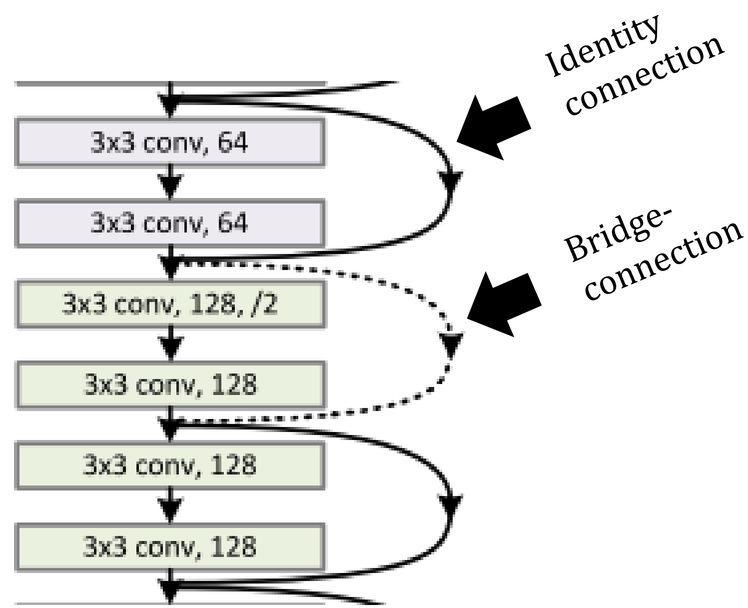
- K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition **(CVPR)**, **2016**.
    - Won ILSVRC challenge on ImageNet dataset in 2015.

- J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," IEEE Conference on Computer Vision and Pattern Recognition **(CVPR)**, **2018**.
    - Won ILSVRC challenge on ImageNet dataset in 2017.

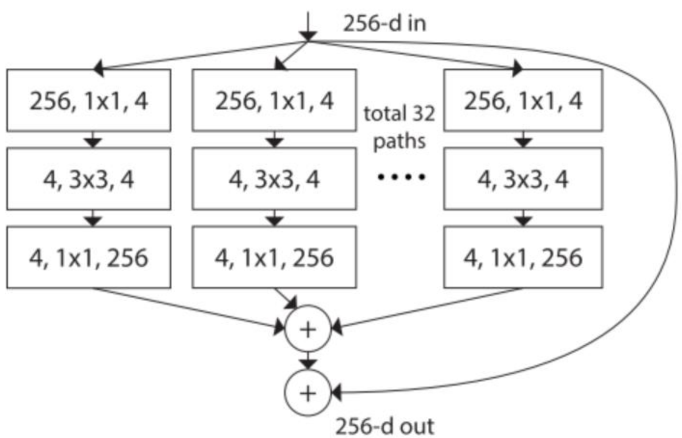
Idea behind Residual Connections

- Deep neural networks perform well for many computer vision tasks.
- Training is very difficult – vanishing gradient problem.
- Residual connections increase the gradient flow.
- Learning a perturbation of input.

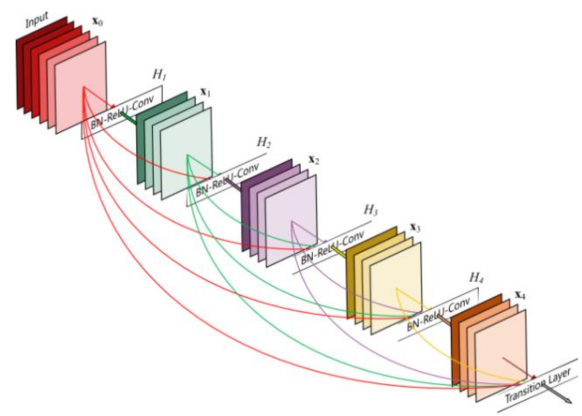




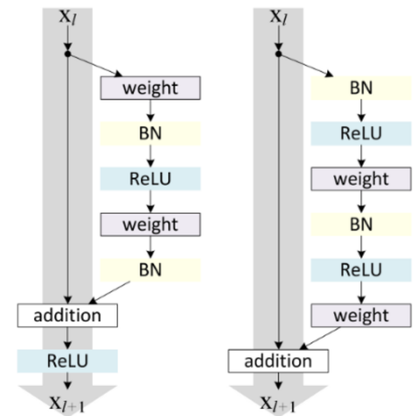
Variants of Resnet



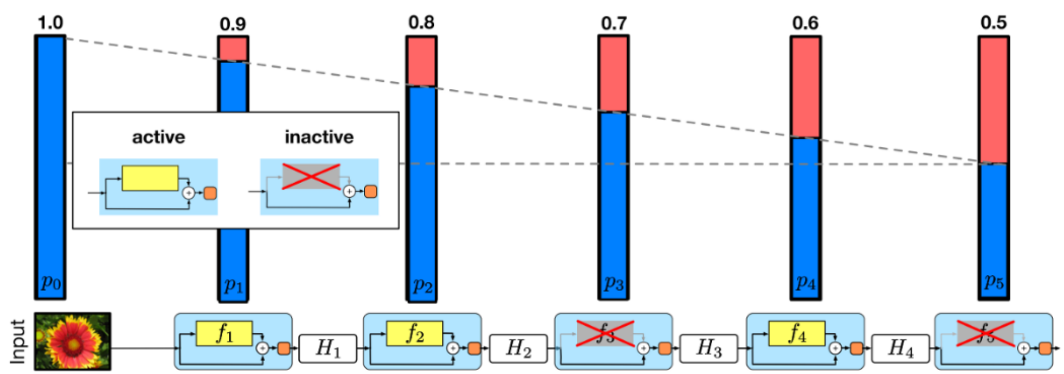
Building block of ResNeXt



Densenet



Pre-activated Resnet



Resnet with stochastic depth

Are Bridge-connections influential?

Architecture	CIFAR-10(Acc%)		CIFAR-100(Acc%)	
	Top-1	Top-5	Top-1	Top-5
Res-20	91.4	99.74	67.37	91.06
Res-32	92.32	99.73	69.8	91.25
Res-44	93.57	99.81	73.15	92.9
Res-56	93.16	99.82	73.8	92.99
Res-110	93.66	99.77	73.33	92.7

Performance of Baseline Resnets

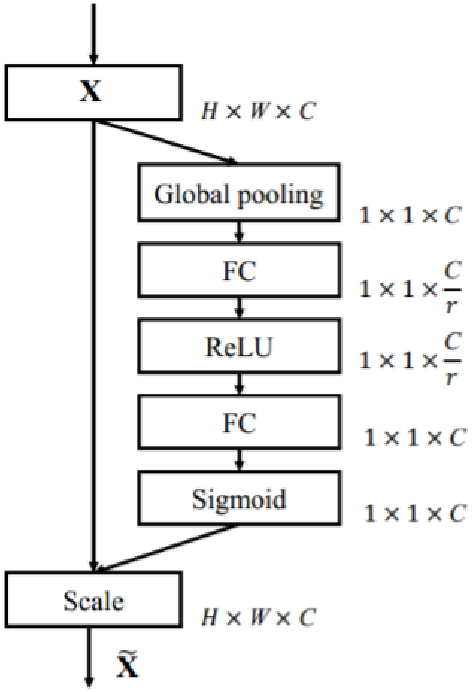
Architecture	CIFAR-10(Acc%)		CIFAR-100(Acc%)	
	Top-1	Top-5	Top-1	Top-5
Res-20	90.48	99.71	67.37	91.06
Res-32	90.38	99.62	69.8	91.25
Res-44	87.62	99.51	73.15	92.9
Res-56	77.46	98.55	48.65	76.67
Res-110	92.42	99.83	1.00	5.00

Performance of Baseline Resnets without Bridge-connections

Squeeze- Excitation (SE) Block

- Weighting the feature-maps – strengthens their contribution.

- Does not allow globally insignificant features to propagate through the network.



mind 2019

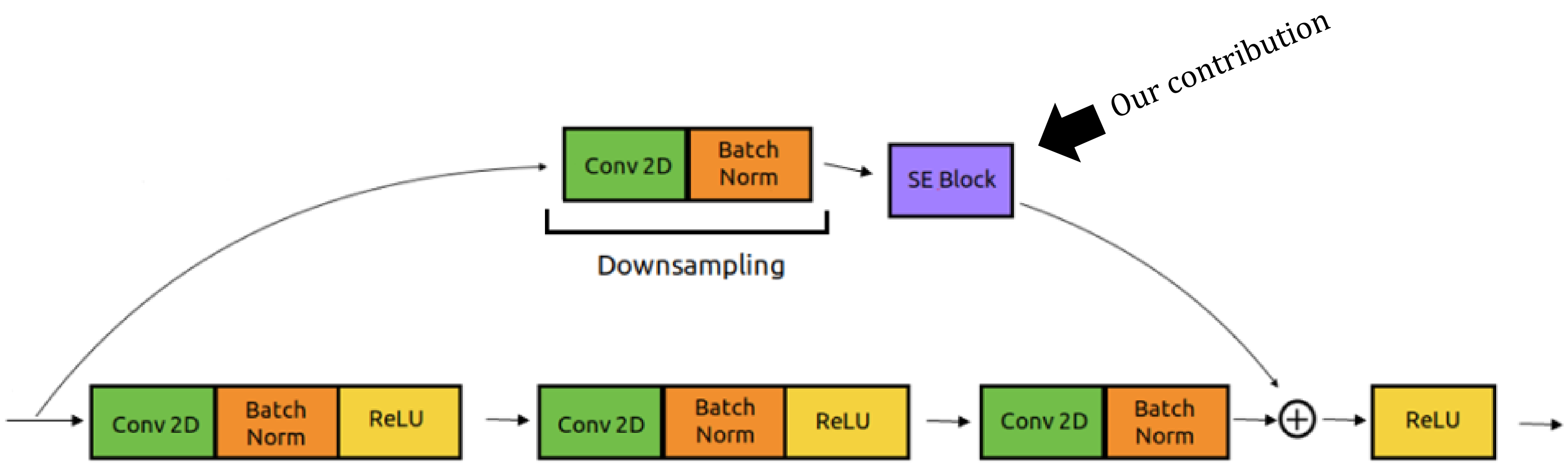
Gaps in Resnet and SE-Resnet

- Resnet equally weights all feature-maps on all connections. Some feature-maps may be redundant.

- SE blocks are not plugged into bridge-connections. Still, feature-maps may be redundant.

We fill these gaps by introducing SE block on bridge-connections.

Proposed architecture: Res-SE-Net



Res-SE-Net: Dataset

1. **CIFAR-10**
    - RGB images and belonging to 10 classes.
    - 50000 training images or 5000 training images per class.
    - 10000 test images or 1000 test images per class.

2. **CIFAR-100**
    - RGB images and belonging to 100 classes.
    - 50000 training images or 500 training images per class.
    - 10000 test images or 100 test images per class.

Res-SE-Net: Training Details

Padding with 4 pixels → Random crop to 32 x 32 → Random horizontal flip → Standard Normalization

Hyperparameter	Value
Max No of Iterations	64000
Initial Learning Rate	0.1
Learning rate after 32000 iterations	0.01
Learning rate after 48000 iterations	0.001
Weight Decay	0.0001
Momentum	0.9
Batch size	128

Res-SE-Net: Implementation Details

- Experiments done with Res-SE-Net of layers 20, 32, 44, 56 and 110.
- Implemented using PyTorch 0.4.1 framework.
- Trained on NVIDIA Titan X GPU.
- Code available at https://github.com/varshaneya/Res-SE-Net

Res-SE-Net: Results

Architecture	CIFAR-10(Acc%)		CIFAR-100(Acc%)	
	Top-1	Top-5	Top-1	Top-5
Res-SE-Net-20	91.9	99.75	67.99	91.03
Res-SE-Net-32	92.79	99.82	69.93	91.95
Res-SE-Net-44	94.08	99.83	73.83	93.54
Res-SE-Net-56	93.64	99.87	74.29	93.45
Res-SE-Net-110	94.53	99.87	74.93	93.48

Performance of the proposed model

Architecture	CIFAR-10(%)		CIFAR-100(%)	
	Top-1	Top-5	Top-1	Top-5
Res-SE-Net-20	0.5	0.01	0.62	-0.03
Res-SE-Net-32	0.47	0.09	0.13	0.7
Res-SE-Net-44	0.51	0.02	0.68	0.55
Res-SE-Net-56	0.48	0.05	0.49	0.46
Res-SE-Net-110	**0.87**	**0.1**	**1.6**	**0.78**
Average Improvement	0.566	0.504	0.704	0.492

Improvement over Resnet

Architecture	CIFAR-10(%)		CIFAR-100(%)	
	Top-1	Top-5	Top-1	Top-5
Res-SE-Net-20	-0.2	0.05	-0.32	0.34
Res-SE-Net-32	-0.29	-0.01	-0.16	-0.17
Res-SE-Net-44	**0.37**	**0.06**	**2.63**	**1.76**
Res-SE-Net-56	0.0	0.11	2.87	2.17
Res-SE-Net-110	**0.74**	**0.13**	**1.94**	**0.77**
Average Improvement	0.124	0.068	1.392	0.974

Improvement over SE-Resnet

Res-SE-Net: Highlights

- Res-SE-Net-110 outperforms all Resnets, SE-Resnets and the other Res-SE-Nets on both CIFAR-10 and CIFAR-100 datasets.

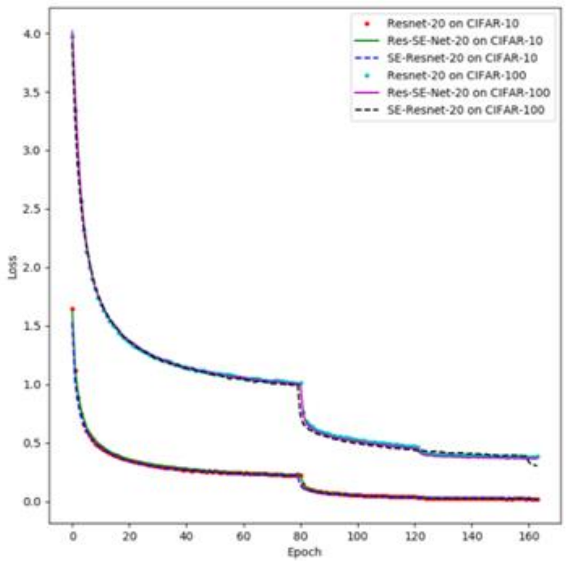
- The average improvement of Res-SE-Net (across different depths) over baseline Resnet is by 0.566% on CIFAR-10 and by 0.704% on CIFAR100 datasets.

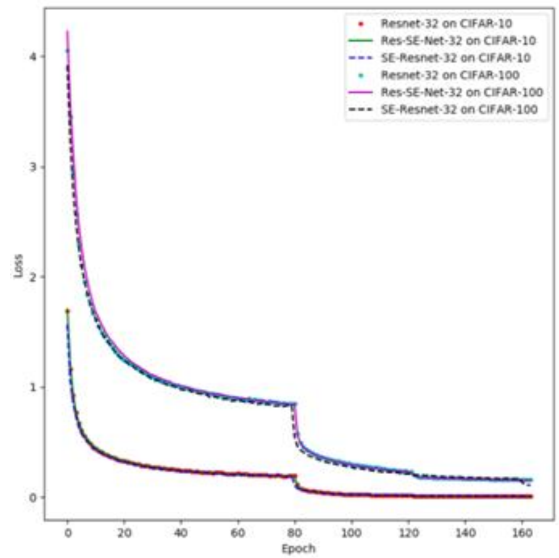
- The average improvement of Res-SE-Net (across different depths) over SE-Resnet is by 0.124% on CIFAR-10 and by 1.392% on CIFAR100 datasets.

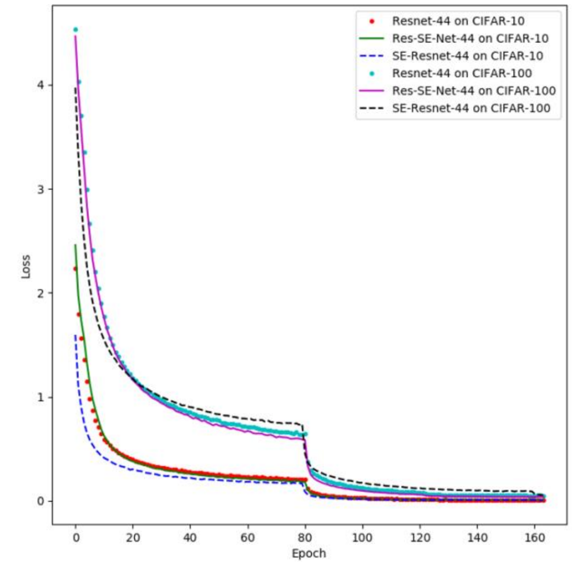
Res-SE-Net: Highlights

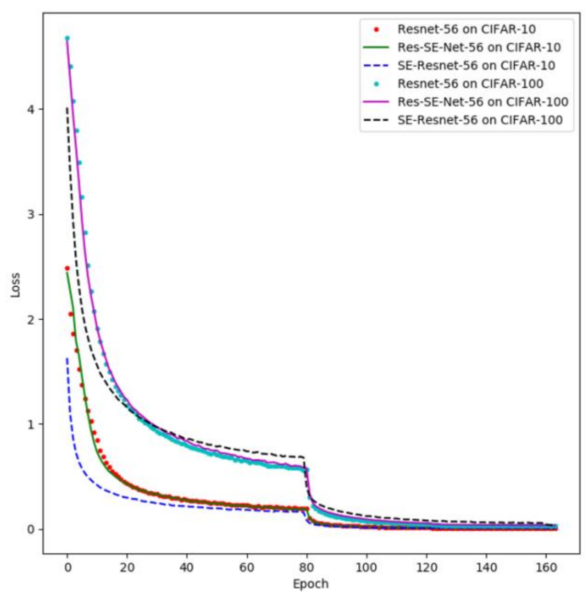
- On CIFAR-10 dataset, Res-SE-Net-44 itself outperforms Resnet-110 and SE-Resnet-110 by a margin of 0.42% and 0.29%.
- On CIFAR-100 dataset, Res-SE-Net-44 itself outperforms over Resnet-110 and SE-Resnet-110 by a margin of 0.5% and 0.84%.
- Res-SE-Net-44 has 61.75% and 62.06% lesser number of parameters compared to Resnet-110 and SE-Resnet-110 respectively.

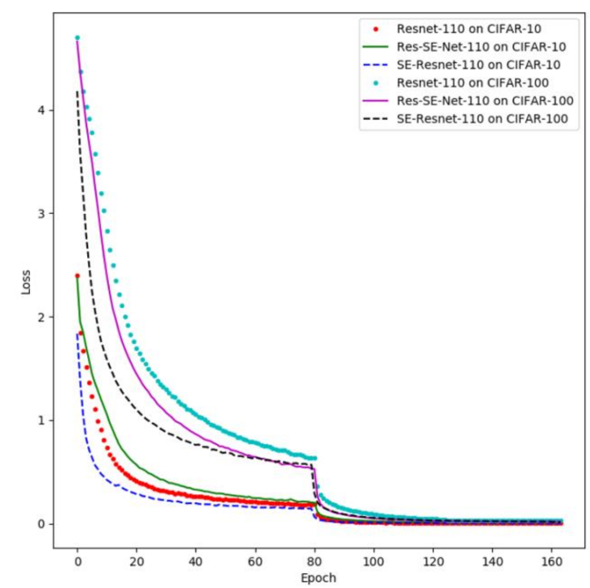
Res-SE-Net: Training Loss











mind 2019

Res-SE-Net: Empirical Observations

- Addition of SE block on the bridge connection before the convolution, did not give good results.
- Addition of SE block to all of the identity-connections degrades the performance.

Conclusion

- Res-SE-Net surpassed the performances of baseline Resnet and SE-Resnet by significant margins on CIFAR-10 and CIFAR-100 datasets.

- Reasonably sized deep networks with positively contributing bridge-connections can outperform very deep networks.

- Addition of SE block on the bridge-connections does not affect the gradient flow during back-propagation.