Squeeze-and-Excitation Networks



Jie Hu[1] , Li Shen[2] , Gang Sun[1]

[1] Momenta [2] University of Oxford

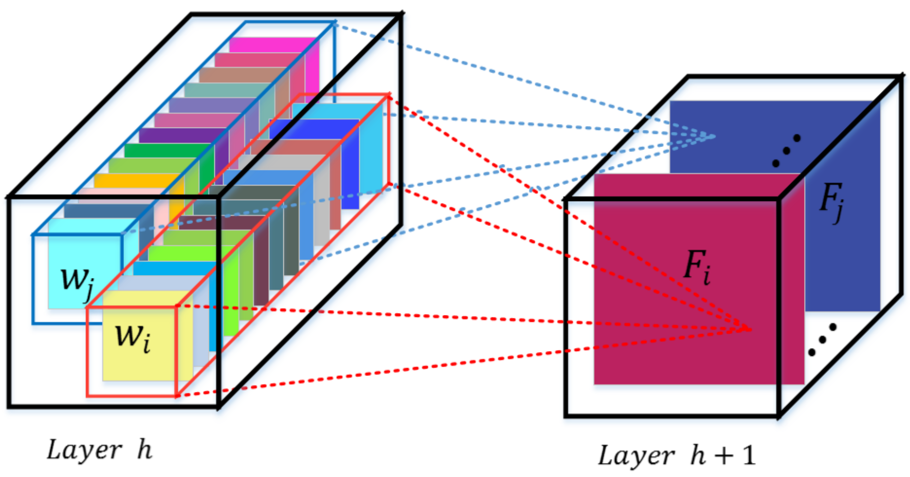
MOMENTA

UNIVERSITY OF OXFORD

Convolution

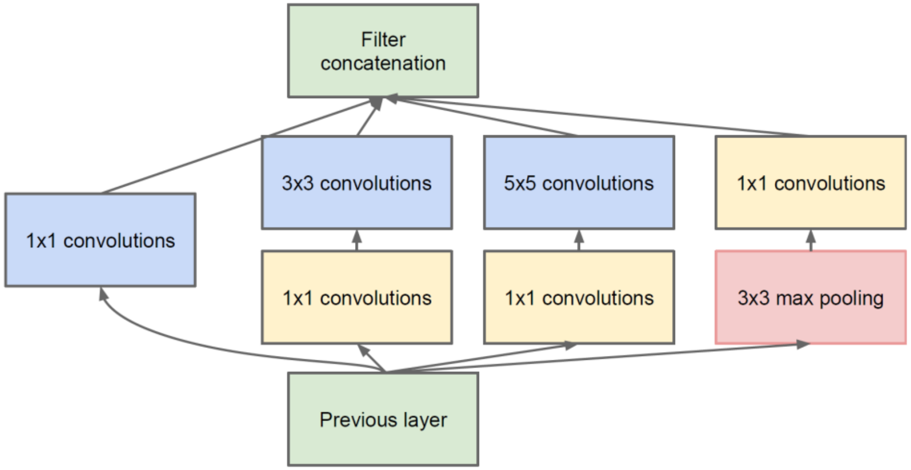
A convolutional filer is expected to be an informative combination

- Fusing **channel-wise** and **spatial** information
- Within local receptive fields



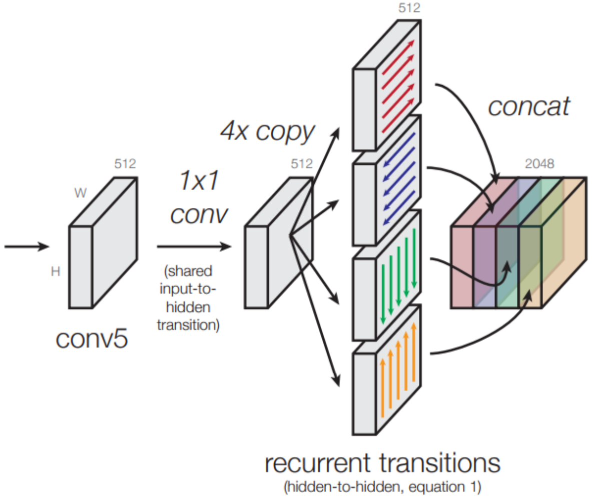
Exploration on Spatial Enhancement

Multi-scale embedding



Inception [9]

Contextual embedding



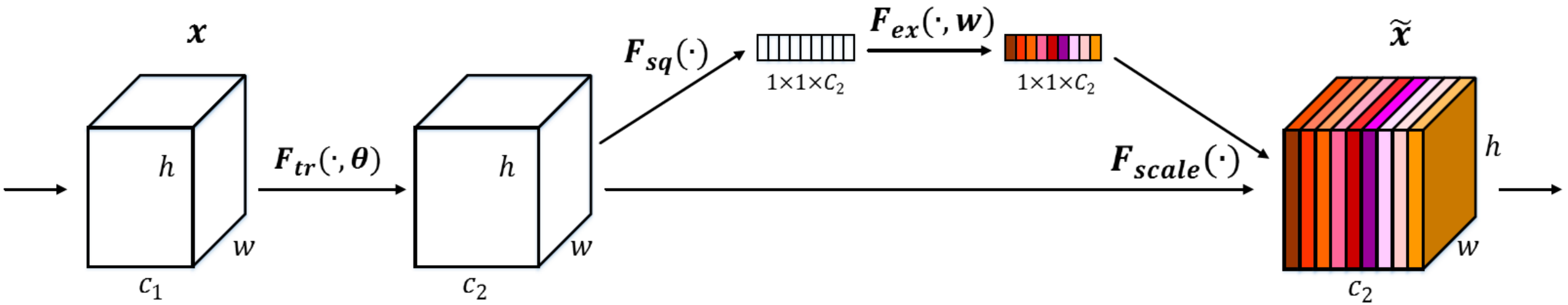
Inside-outside Network [13]

Squeeze-and-Excitation (SE) Networks

- If a network can be enhanced from the aspect of **channel relationship**?

- **Motivation:**
  - Explicitly model channel-interdependencies within modules
  - Feature recalibration
    - Selectively enhance useful features and suppress less useful ones

Squeeze-and-Excitation Module



Squeeze

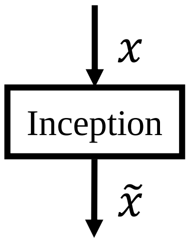
- Shrinking feature maps $\in \mathbb{R}^{w\times h\times c_2}$ through spatial dimensions $(w \times h)$
- Global distribution of channel-wise responses

Excitation

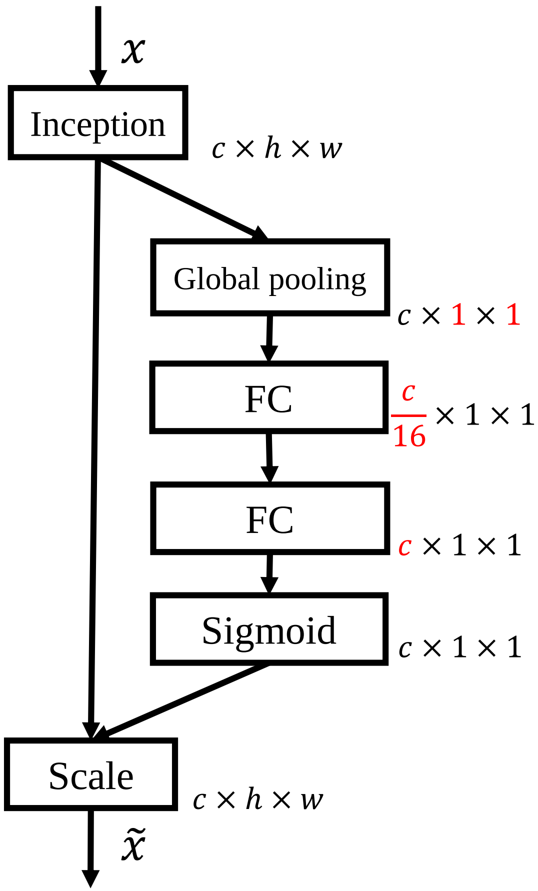
- Learning $W \in \mathbb{R}^{c_2\times c_2}$ to explicitly model channel-association
- Gating mechanism to produce channel-wise weights

Scale

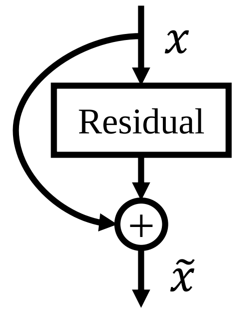
- Reweighting the feature maps $\in \mathbb{R}^{w\times h\times c_2}$



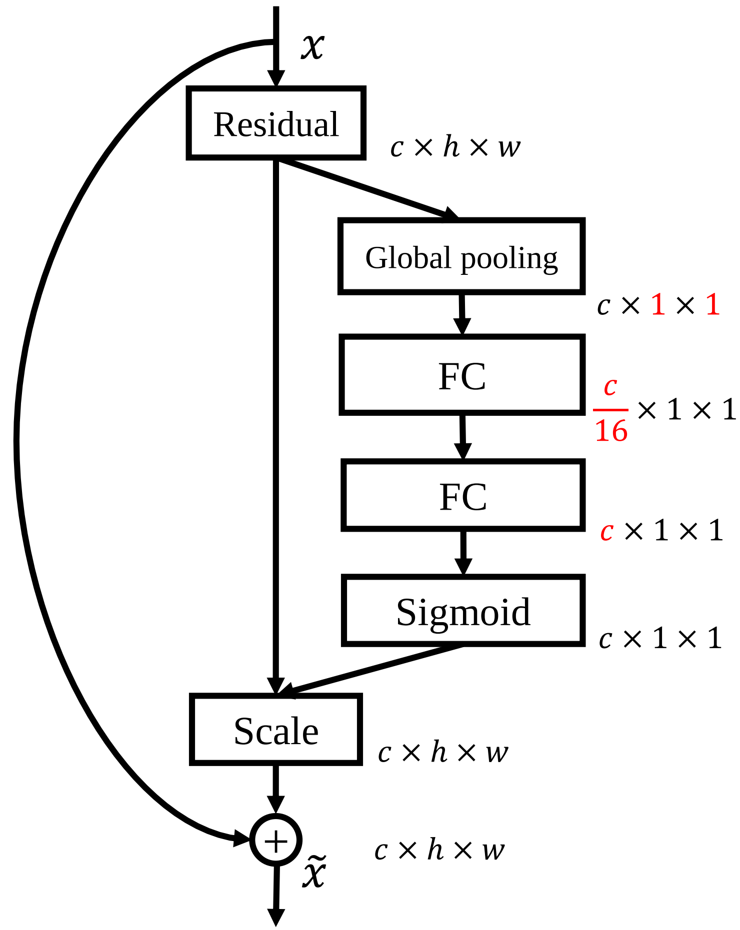
Inception Module



SE-Inception Module



ResNet Module



SE-ResNet Module

Model and Computational Complexity

SE-ResNet-50 vs. ResNet-50

- Parameters: 2%~10% additional parameters
- Computation cost: <1% additional computation (theoretical)
- GPU inference time: 10% additional time
- CPU inference time: <2% additional time

Training – Momenta ROCS

- Data augmentation
  - ✓ Mirror flip, Random size crop [9], Rotation, Color Jitter
- Mini-batch data sampling
  - ✓ Balance-data strategy [7]
- Training hyper-parameters
  - ✓ 4 or 8 GPU severs (8 NVIDIA Titan X per server)
  - ✓ Batch-size: 1024 / 2048 (32 per GPU)
  - ✓ Initial learning rate : 0.6 (decrease each 30 epochs)
  - ✓ Synchronous SGD with momentum 0.9

Experiments on ImageNet-1k dataset

- Empirical investigations on:
  - Benefits against Deeper Networks
  - Incorporation with modern architectures
- ILSVRC 2017 Classification Task

Benefits against Network Depth

	Original		Our re-implementation		SE-module	
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	top-1 err.	top-5 err.
ResNet-50 [1]	24.7	7.8	24.80	7.48	**23.29**(**1.51**)	**6.62**(**0.86**)
ResNet-101 [1]	23.6	7.1	23.17	6.52	**22.38**(**0.79**)	**6.07**(**0.45**)
ResNet-152 [1]	23.0	6.7	22.42	6.34	**21.57**(**0.85**)	**5.73**(**0.61**)

Table 1. Error rates (%) of single-crop results on the ImageNet-1k validation set.

Benefits against Network Depth

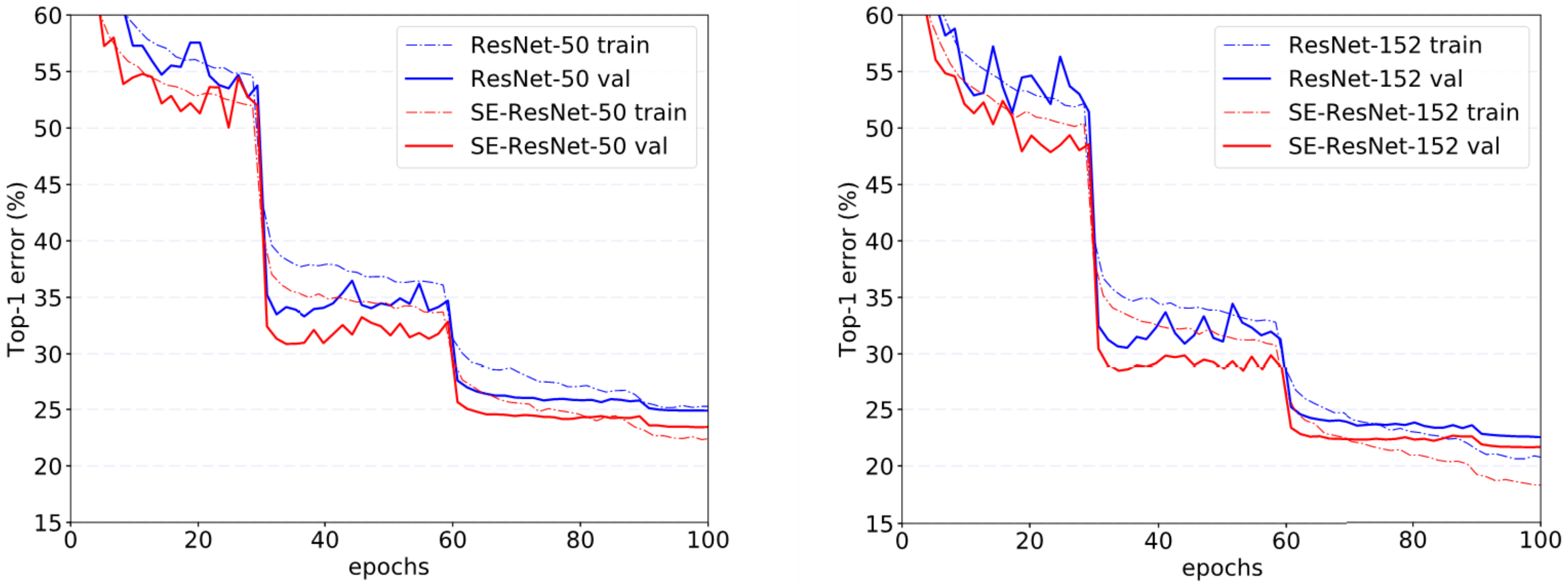


Figure 1. Training curves on ImageNet-1K validation set. (**Left**): ResNet-50 and SE-ResNet-50; (**Right**): ResNet-152 and SE-ResNet-152.

Incorporation with Modern Architectures

	Original		Our re-implementation		SE-module	
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	top-1 err.	top-5 err.
ResNeXt-50 [7]	22.2	-	22.11	5.90	**21.10**$_{(1.01)}$	**5.49**$_{(0.41)}$
ResNeXt-101 [7]	21.2	5.6	21.18	5.57	**20.70**$_{(0.48)}$	**5.01**$_{(0.56)}$
BN-Inception [4]	25.2	7.82	25.38	7.89	**24.23**$_{(1.15)}$	**7.14**$_{(0.75)}$
Inception-ResNet-v2 [5]	19.9†	4.9†	20.37	5.21	**19.80**$_{(0.57)}$	**4.79**$_{(0.42)}$

Table 2. Error rates (%) of single-crop results on the ImageNet-1k validation set. Error rate followed by † means that the image size for center crop is not clear and it evaluates on the non-blacklisted subset of validation set [5], which may lead to slightly better results.

Incorporation with Modern Architectures

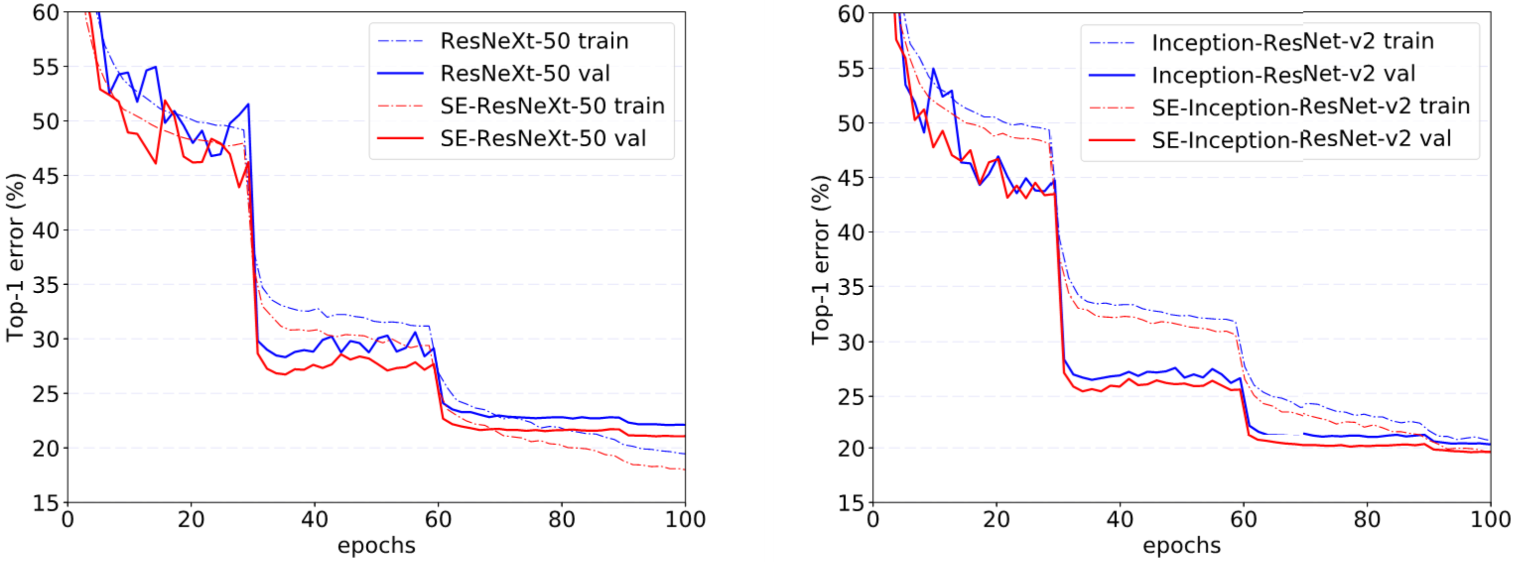


Figure 2. Training curves on ImageNet-1K validation set. (**Left**): ResNeXt-50 and SE-ResNeXt-50; (**Right**): Inception-ResNet-v2 and SE-Inception-ResNet-v2.

Comparison with State-of-the-art

	224 × 224		320 × 320 / 299 × 299	
	top-1 err.	top-5 err.	top-1 err.	top-5 err.
ResNet-152 [1]	23.0	6.7	21.3	5.5
ResNet-200 [3]	21.7	5.8	20.1	4.8
Inception-v3 [10]	-	-	21.2	5.6
Inception-v4 [8]	-	-	20.0	5.0
Inception-ResNet-v2 [8]	-	-	19.9	4.9
ResNeXt-101 [11] (64 × 4d)	20.4	5.3	19.1	4.4
DenseNet-161 [4] (k = 48)	22.2	-	-	-
Very Deep PolyNet [12]	-	-	18.71	4.25
SENet	**18.68**	**4.47**	**17.28**	**3.79**

Table 4. Single-crop error of state-of-the-art CNNs on ImageNet-1k validation set. The size of test crop is 224×224 and 320×320 (299×299 for Inception models) as in [3]. The **SENet** is our well-structured model whose error rates are remarkably lower than previous models.

SENet is a SE-ResNeXt-152 (64 × 4d)

ILSVRC 2017 Classification Task

Team	Top-5 error (%)
WMW	**2.251**
Trimps-Soushen	2.481
NUS-Qihoo-DPNs	2.740
BDAT	2.962
ILSVRC 2016 Winner	2.991

References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2015.
[2] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015.
[3] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *ECCV*, 2016.
[4] G. huang, Z. Liu, K. Weinberge, and L. Maaten. Densely connected convolutional networks. In *CVPR*, 2017.
[5] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.
[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. In *IJCV*, 2015.
[7] L. Shen, Z. Lin, and Q. Huang. Relay backpropagation for effective learning of deep convolutional neural networks. In *ECCV*, 2016.
[8] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. In *arXiv preprint arXiv:1602.07261*, 2016.
[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015.
[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016.
[11] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *CVPR*, 2016.
[12] X. Zhang, Z. Li, C. Chen, and D. Lin. Polynet: A pursuit of structural diversity in very deep networks. In *CVPR*, 2017.
[13] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks. In *CVPR*, 2016.

Thank you