Advanced Machine Learning A.A. 2022/2023

Squeeze-and-Excitation Networks

Jie Hu, Enhua Wu, Gang Sun, Li Shen, Samuel Albanie

ILSVRC Competition
2017

Paper presentation made by
Giuseppina Gagliardi

POLITECNICO
DI TORINO

AMML

Publication

- First place at ILSVRC 2017 Classification competition
- 4 versions, the latest relased in 2019

Top 5-error reduced to 2.251% on the test set

Relative improvement of ~25%
compared to the previous year (2016)

Squeeze and Excitation (SE) Block

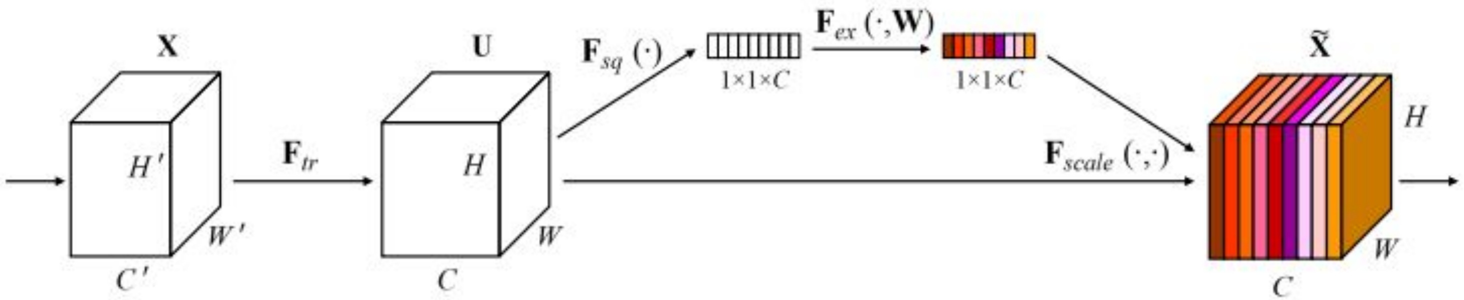
A new architectural unit for CNN's

- GOAL

Improving the quality of representations produced by a network

- How?

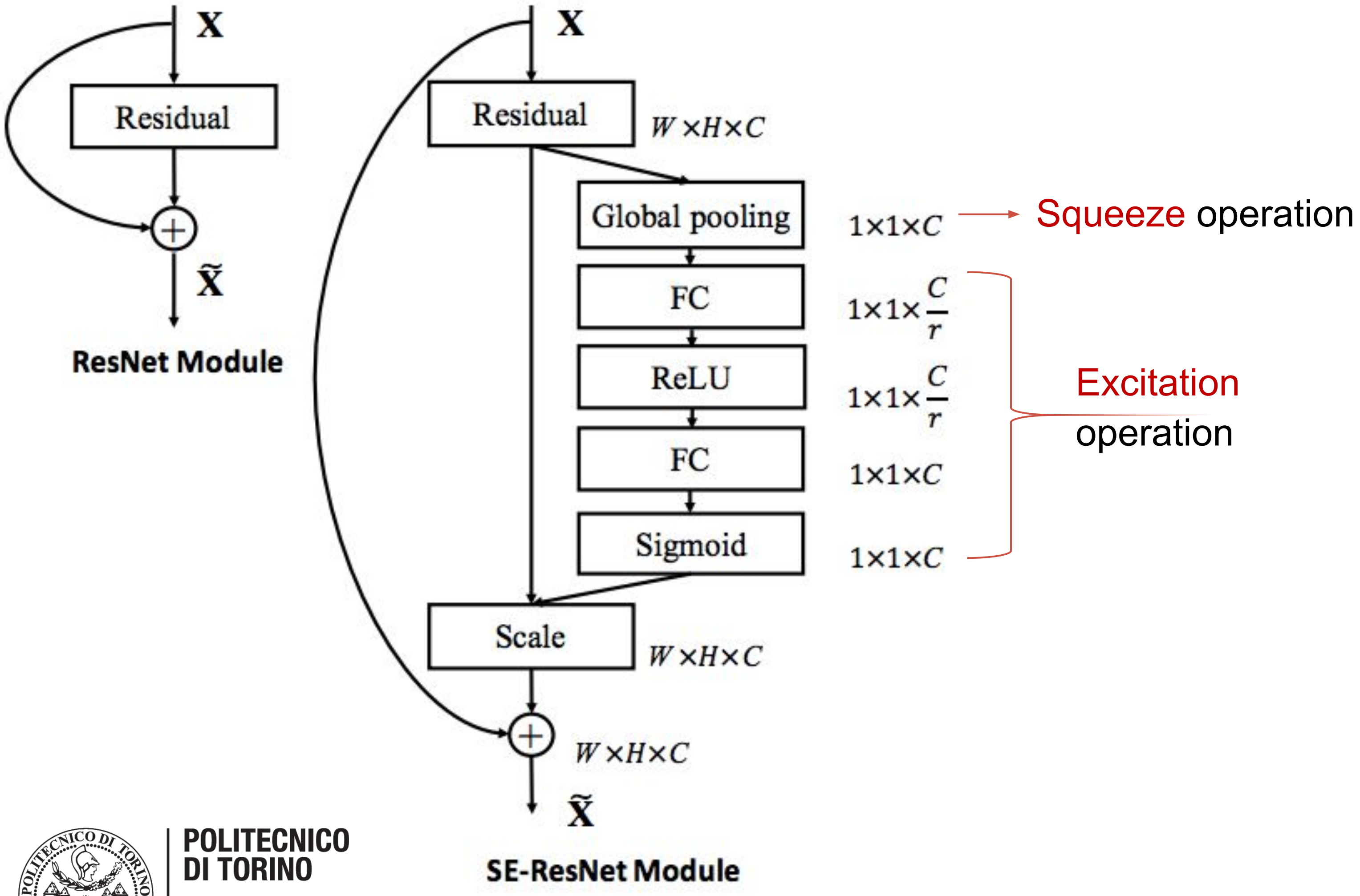
Modelling the interdependencies between the channels of network's convolutional features



Fsq -> Squeeze operation

Fex -> Excitation operation

Integration of SE blocks in other architectures



Additional parameters

$$\frac{2}{r} \sum_{s=1}^{S} N_s \cdot {C_s}^2$$

r = Reduction ratio

s = Number of stages

Cs = Dim of the output channels

Ns = Num of repeated blocks for stage s

Experiments

Single-crop error rates (%) on the ImageNet validation set and complexity comparisons.

	original		re-implementation			SENet		
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	GFLOPs	top-1 err.	top-5 err.	GFLOPs
ResNet-50 [13]	24.7	7.8	24.80	7.48	3.86	$23.29_{(1.51)}$	$6.62_{(0.86)}$	3.87
ResNet-101 [13]	23.6	7.1	23.17	6.52	7.58	$22.38_{(0.79)}$	$6.07_{(0.45)}$	7.60
ResNet-152 [13]	23.0	6.7	22.42	6.34	11.30	$21.57_{(0.85)}$	$5.73_{(0.61)}$	11.32
ResNeXt-50 [19]	22.2	-	22.11	5.90	4.24	$21.10_{(1.01)}$	$5.49_{(0.41)}$	4.25
ResNeXt-101 [19]	21.2	5.6	21.18	5.57	7.99	$20.70_{(0.48)}$	$5.01_{(0.56)}$	8.00
VGG-16 [11]	-	-	27.02	8.81	15.47	$25.22_{(1.80)}$	$7.70_{(1.11)}$	15.48
BN-Inception [6]	25.2	7.82	25.38	7.89	2.03	$24.23_{(1.15)}$	$7.14_{(0.75)}$	2.04
Inception-ResNet-v2 [21]	$19.9^{\dagger}$	$4.9^{\dagger}$	20.37	5.21	11.75	$19.80_{(0.57)}$	$4.79_{(0.42)}$	11.76

	original		re-implementation				SENet			
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	MFLOPs	Params	top-1 err.	top-5 err.	MFLOPs	Params
MobileNet [64]	29.4	-	28.4	9.4	569	4.2M	$25.3_{(3.1)}$	$7.7_{(1.7)}$	572	4.7M
ShuffleNet [65]	32.6	-	32.6	12.5	140	1.8M	$31.0_{(1.6)}$	$11.1_{(1.4)}$	142	2.4M

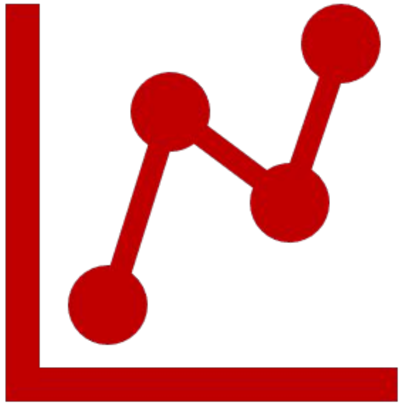
Accuracy improved by a large margin at a minimal increase in computational cost

POLITECNICO
DI TORINO

Other datasets used to demonstrate the benefits of SE blocks

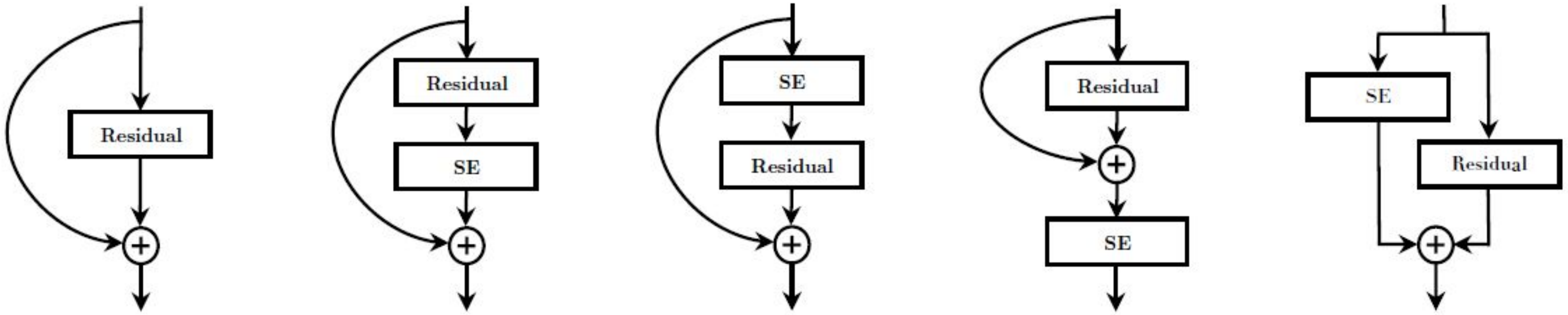
- CIFAR-10
- CIFAR-100
- Places365-Challenge dataset
- COCO dataset

In every comparison the SENets outperform the basic architectures!



Ablation study

To evaluate the influence of the location of the SE block when integrating it into existing architectures



(a) Residual block (b) Standard SE block (c) SE-PRE block (d) SE-POST block (e) SE-Identity block

Effect of different SE block integration strategies with ResNet-50 on ImageNet (error rates %).

Design	top-1 err.	top-5 err.
SE	22.28	6.03
SE-PRE	22.23	6.00
SE-POST	22.78	6.35
SE-Identity	22.20	6.15

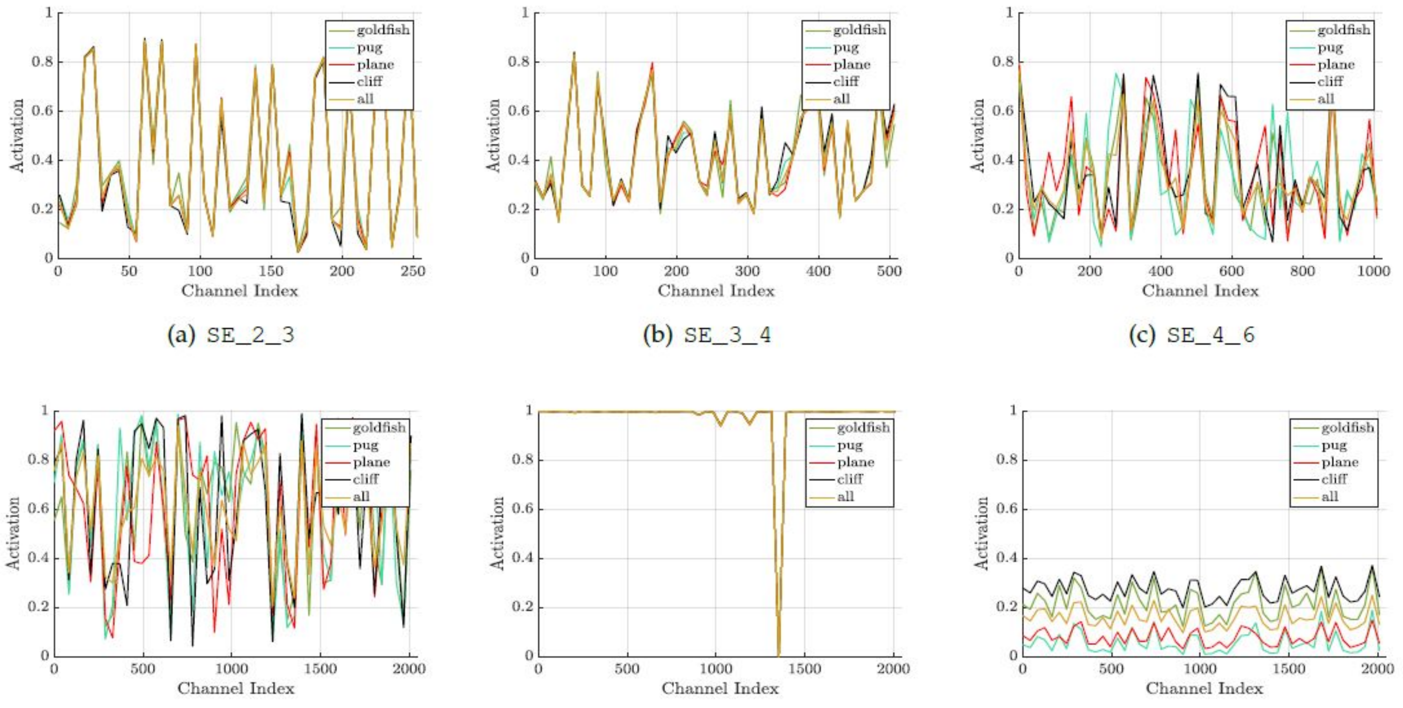
POLITECNICO DI TORINO

Effects of Squeeze and Excitation

Effect of Squeeze operator on ImageNet (error rates %).

	top-1 err.	top-5 err.	GFLOPs	Params
ResNet-50	23.30	6.55	3.86	25.6M
NoSqueeze	22.93	6.39	4.27	28.1M
SE	**22.28**	**6.03**	3.87	28.1M

Activations induced by the Excitation operator at different depths in the SE-ResNet-50 on ImageNet.



(a) SE_2_3 (b) SE_3_4 (c) SE_4_6

(d) SE_5_1 (e) SE_5_2 (f) SE_5_3

POLITECNICO
DI TORINO

Conclusions

Channel interdependencies improved at almost no computational cost.

Possible future uses of the SE block for other tasks, such as network pruning for model compression.