Sparsely Aggregated Convolutional Networks

Ligeng Zhu[1] **Ruizhi Deng**[1] **Michael Maire**[2] **Zhiwei Deng**[1] **Greg Mori**[1] **Ping Tan**[1]

[1]Simon Fraser University [2]University of Chicago

{lykenz,ruizhid,zhiweid,mori,pingtan}@sfu.ca, mmaire@uchicago.edu

SFU SIMON FRASER UNIVERSITY ENGAGING THE WORLD

THE UNIVERSITY OF CHICAGO 1890

Overview

Contribution:

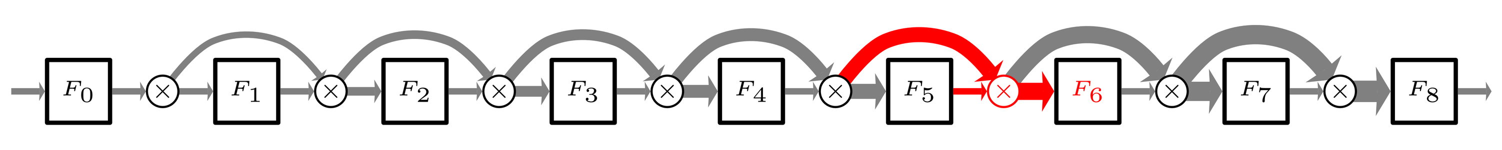
- New internal skip-link (aggregation) structure for CNNs
- Focus on *which* features to aggregate rather than on aggregation operator
- Sparsification of internal links applicable to both DenseNet [1] and ResNet [2]

Results:

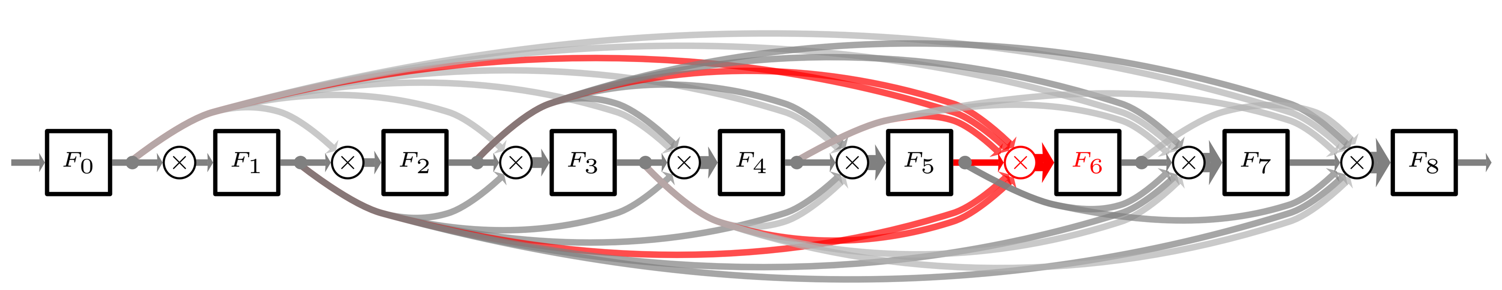
- SparseNet, a sparsified DenseNet:
  - matches accuracy of best DenseNet models, while
  - being more efficient (in both parameters and FLOPs)
- SparseNet scales more robustly to great depth (1000+ layers)

Approach

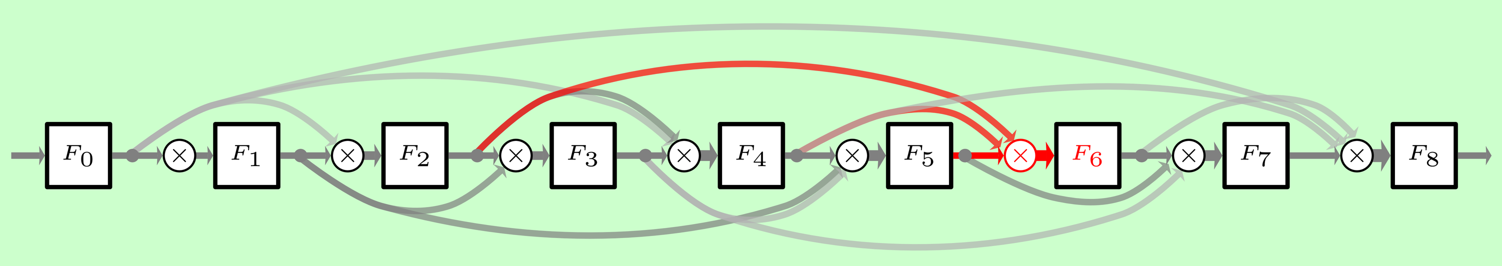
(a) Dense Aggregation (ResNet / DenseNet Topology)



(b) Dense Aggregation: Equivalent Exploded View of **(a)**



(c) Sparse Aggregation (Our Proposed Topology)



Intuition: Logarithmic Growth with Depth

- Avoid redundancy or wash-out from aggregating too many features
- Maintain relatively short gradient paths
- Reduce overall amount of skip-connection scaffolding

	Params	Path Length	#Features Aggregated
Plain	$O(N)$	$O(N)$	$O(1)$
ResNet	$O(N)$	$O(1)$	$O(\ell)$
DenseNet	$O(N^2)$	$O(1)$	$O(\ell)$
SparseNet[+] (sum)	$O(N)$	$O(\log(N))$	$O(\log \ell)$
SparseNet[$\oplus$] (concat)	$O(N \log N)$	$O(\log(N))$	$O(\log \ell)$

Table 1: Scaling properties for networks of depth N and for layers at depth ℓ.

Results: CIFAR and ImageNet Classification

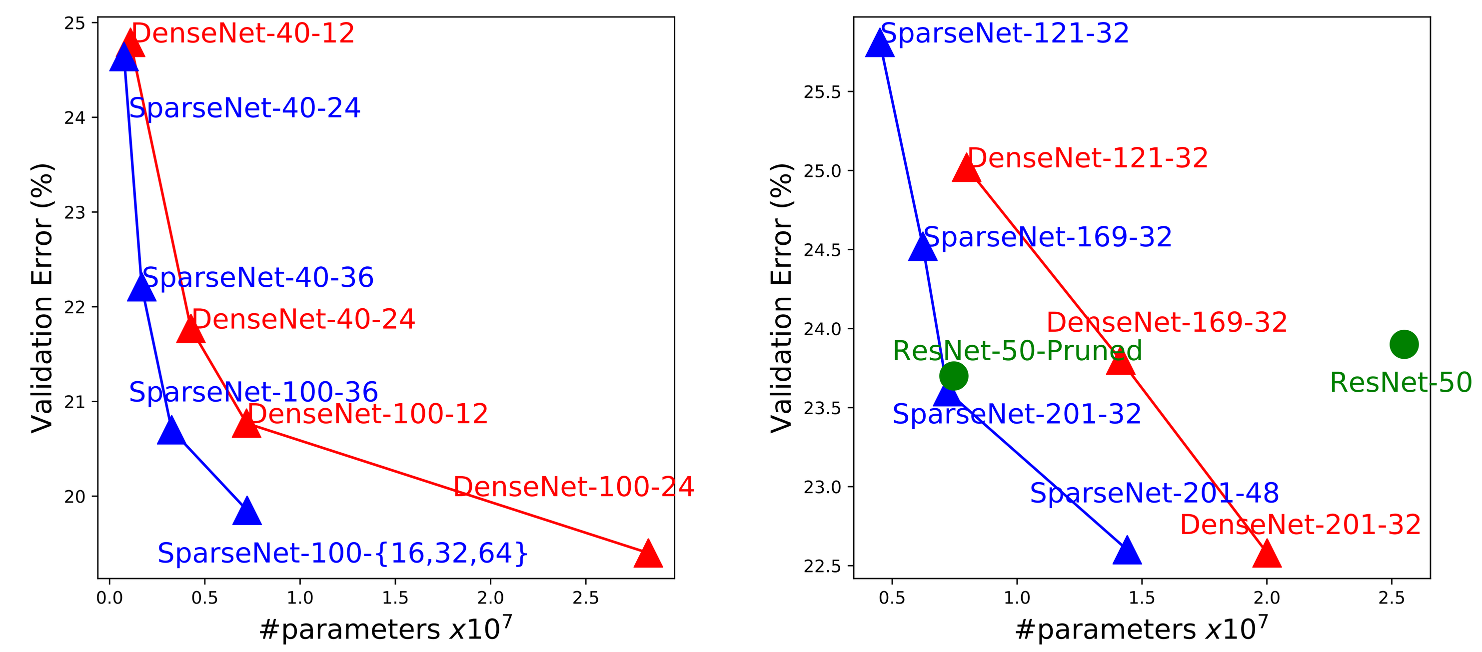


Figure 1: SparseNet[$\oplus$] , our sparse analogue of DenseNet, offers better accuracy for any parameter budget. ***Left:*** CIFAR-100 [3]. ***Right:*** ImageNet [4].

ImageNet Models	Error	Params	FLOPs	Time
DenseNet-121-32 [1]	25.0*	7.98M	5.7G	19.5ms
DenseNet-169-32 [1]	23.6*	14.15M	6.76G	32.0ms
DenseNet-201-32 [1]	**22.5***	20.01M	8.63G	42.6ms
SparseNet[$\oplus$]-121-32	25.8	4.51M	3.46G	13.5ms
SparseNet[$\oplus$]-169-32	24.5	6.23M	3.74G	18.8ms
SparseNet[$\oplus$]-201-32	23.6	7.22M	4.13G	22.0ms
SparseNet[$\oplus$]-201-48	**22.7**	14.91M	9.19G	43.1ms
ResNet-50	23.9	25.5M	8.20G	42.2ms
ResNet-50 Pruned [5]	**23.7**	7.47M	-	-

Table 2: The top-1 single-crop validation error, parameters, FLOPs, and inference time of each model on ImageNet.

Results: Robustness *vs.* ResNet at Great Depth

- Avoid wash-out from aggregating too many features
- SparseNet is robustly trainable when instantiated very deep
- Thin, deep SparseNets still squeeze performance from additional parameters

Model	Depth	Params	CIFAR 100+	Model	Depth	CIFAR 100+
ResNet	38	0.40M	28.43	SparseNet[+]	38	28.23
	56	0.59M	27.00		56	27.70
	110	1.15M	24.70		110	26.10
	200	2.07M	23.10		200	25.77
	1001	10.33M	**21.42**		1001	22.10
	2000	20.62M	22.76		2000	**21.01**

Table 3: SparseNet[+], our sparse analogue of ResNet, is robust to the performance regression observed on CIFAR-100 when ResNet is stretched too deep.

Results: Robustness *vs.* DenseNet at Great Depth

Model	Depth	Params	CIFAR 100+
DenseNet (k=12)	40	1.10M	24.79
	100	7.20M	20.97
	400	117M	N/A
DenseNet-BC (k=24)	250	25.6M	17.6
	400	216.3M	N/A
DenseNet-BC (k=4)	400	1.10M	32.94
SparseNet[$\oplus$]-BC (k=12)	100	0.40M	27.89
	400	1.70M	24.53

Table 4: Performance of SparseNet[$\oplus$] *vs.* DenseNet on CIFAR-100. N/A indicates the experiment setting is not applicable on available hardware.

Analysis: Layer-to-Layer Connection Strengths

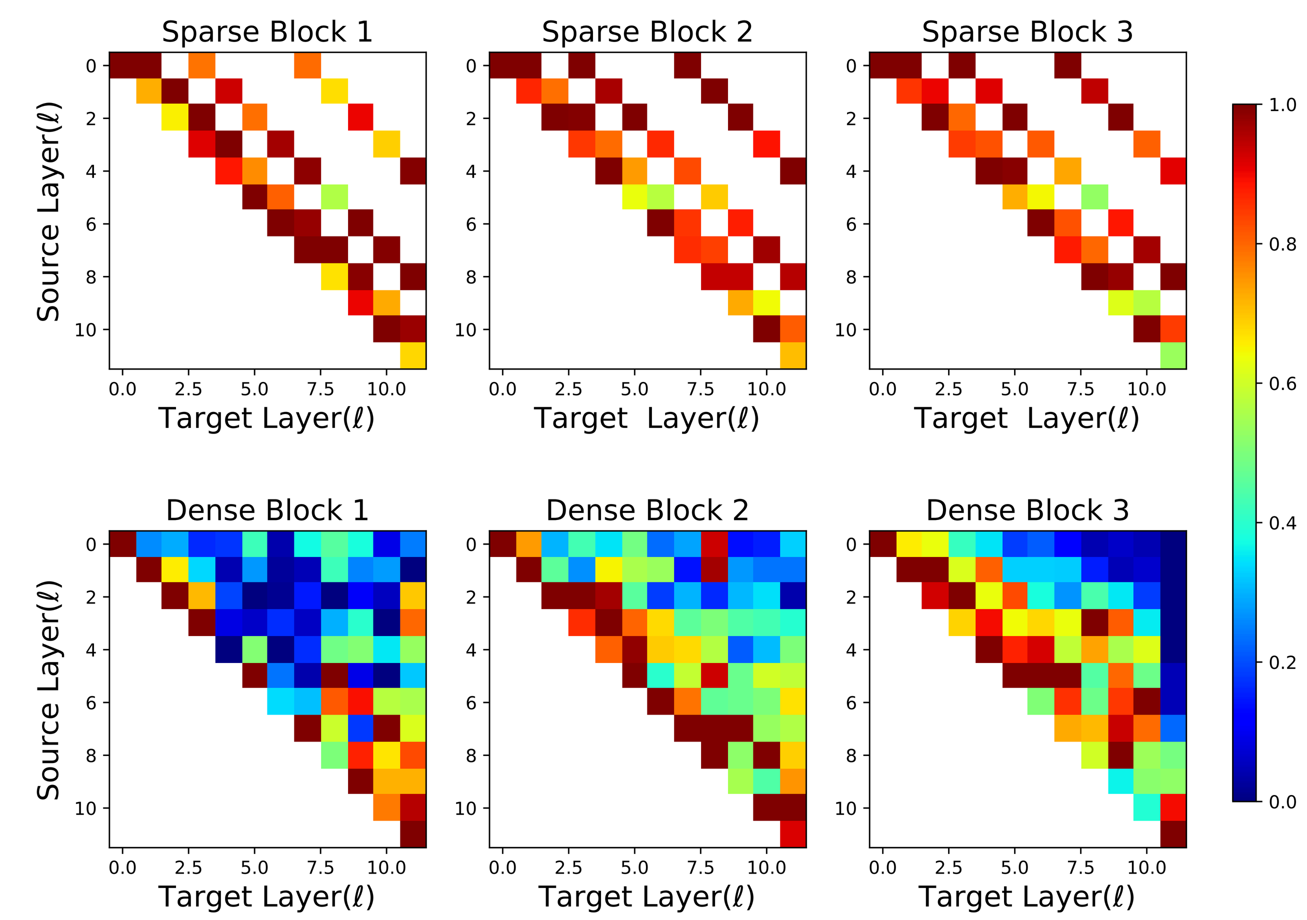


Figure 2: SparseNet[$\oplus$] layers highly utilize (red) feature representations from all incoming links, whereas DenseNet layers learn some weak connections (blue). This suggests DenseNet is over-budgeted (inefficient) in terms of skip-links.

References

[1] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *CVPR*, 2017.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, 2016.

[3] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," tech. rep., University of Toronto, 2009.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *CVPR*, 2009.

[5] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," *ICLR*, 2016.