

Densely Connected Convolutional Networks

Yufeng Jiang

May 27, 2018

1 Experiments

Authors empirically demonstrate DenseNet’s effectiveness on several benchmark datasets and compare with state-of-the-art architectures.

1.1 Datasets

CIFAR. The two CIFAR datasets consist of colored natural images with 32×32 pixels. The training and test sets contain 50000 and 10000 images respectively, and authors hold

out 5000 training images as a validation set. For preprocessing, authors normalize the data using the channel means and standard deviations. For the final run authors use all 50000 training images and report the final test error at the end of training.

SVHN. The Street View House Numbers (SVHN) dataset contains 32×32 colored digit images. There are 73257 images in the training set, 26032 images in the test set, and 521131 images for additional training. Authors select the model with the lowest validation error during training and report the test error.

Method	Depth	Params	C10	C10+	C100	C100+	SVHN
Network in Network [6]	-	-	10.41	8.81	35.68	-	2.35
All-CNN [7]	-	-	9.08	7.25	-	33.71	-
Deeply Supervised Net [5]	-	-	9.69	7.97	-	34.57	1.92
Highway Network [8]	-	-	-	7.72	-	32.39	-
FractalNet [4]	21	38.6M	10.18	5.22	35.34	23.30	2.01
with Dropout/Drop-path	21	38.6M	7.33	4.60	28.20	23.73	1.87
ResNet [2]	110	1.7M	-	6.61	-	-	-
ResNet (reported by [3])	110	1.7M	13.63	6.41	44.74	27.22	2.01
ResNet with Stochastic Depth [3]	110	1.7M	11.66	5.23	37.80	24.58	1.75
	1202	10.2M	-	4.91	-	-	-
Wide ResNet [9]	16	11.0M	-	4.81	-	22.07	-
	28	36.5M	-	4.17	-	20.50	-
	16	2.7M	-	-	-	-	1.64
ResNet (pre-activation) [1]	164	1.7M	11.26	5.46	35.58	24.33	-
	1001	10.2M	10.56	4.62	33.47	22.71	-
DenseNet ($\kappa = 12$)	40	1.0M	7.00	5.24	27.55	24.42	1.79
DenseNet ($\kappa = 12$)	100	7.0M	5.77	4.10	23.79	20.20	1.67
DenseNet ($\kappa = 24$)	100	27.2M	5.83	3.74	23.42	19.25	1.59
DenseNet-BC ($\kappa = 12$)	100	0.8M	5.92	4.51	24.15	22.27	1.76
DenseNet-BC ($\kappa = 24$)	250	15.3M	5.19	3.62	19.64	17.60	1.74
DenseNet-BC ($\kappa = 40$)	190	25.6M	-	3.46	-	17.18	-

Table 1: Error rates on CIFAR and SVHN datasets. κ denotes networks growth rate. Results that surpass all competing methods are **bold** and the overall best results are **blue**. + indicates standard data augmentation (translation and/or mirroring). All the results of DenseNets without data augmentation (C10, C100, SVHN) are obtained using Dropout. DenseNets achieve lower error rates while using fewer parameters than ResNet. Without data augmentation, DenseNet performs better by a large margin.

ImageNet. The ILSVRC 2012 classification dataset consists 1.2 million images for training, and 50000 for validation from 1000 classes. Authors adopt the same data augmentation scheme for training images as in [1, 2]. Following [2, 3], authors report classification errors on the validation set.

2 Training

All the networks are trained using stochastic gradient descent (SGD). Respectively, on CIFAR and SVHN authors train using batch size 64 for 300 and 40 epochs. To reduce the memory consumption on GPUs, please refer to authors technical report on the memory-efficient implementation of DenseNet.

Following [6], authors use a weight decay of 10^{-4} and a Nesterov momentum of 0.9 without dampening. Authors adopt the weight initialization introduced by [1]. The test errors were only evaluated once for each task and model setting.

3 Classification Results on CIFAR and SVHN

Authors train DenseNets with different depths, L , and growth rates, κ . The main results on CIFAR and SVHN are shown in Table 1. To highlight general trends, authors mark all results that outperform the existing state-of-the-art in **boldface** and the overall best result in **blue**.

Accuracy. Possibly the most noticeable trend may originate from the bottom row of Table 1, which shows that DenseNet-BC with $L = 190$ and $\kappa = 40$ outperforms the existing state-of-the-art consistently on all the CIFAR datasets. On SVHN, the DenseNet with $L = 100$ and $\kappa = 24$ also surpasses the current best result achieved by wide ResNet. This may be explained by that SVHN is a relatively easy task, and extremely deep models may overfit to the training set.

Parameter Efficiency. The results in Table 1 indicate that DenseNets utilize parameters more efficiently than alternative architectures. Figure 1 shows the training loss and test errors of these two networks on C10+. The 1001-layer deep ResNet converges to a lower training loss value but a similar test error.

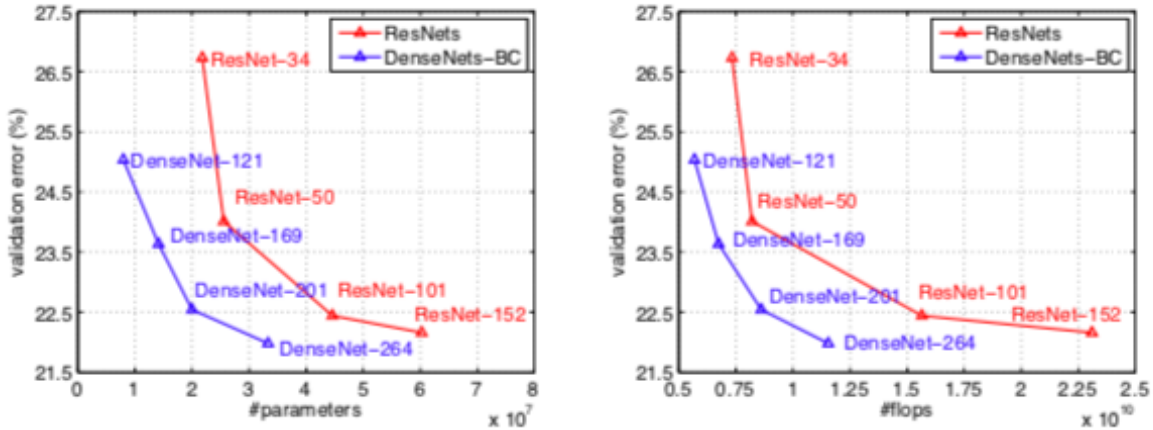


Figure 1: Comparison of the DenseNets and ResNets top-1 error rates (single-crop testing) on the ImageNet validation dataset as a function of learned parameters (left) and FLOPs during test-time (right).

References

- [1] K. He, X. Zhang, and S. Ren, “Identity mappings in deep residual networks.” In *ECCV*, 2016. 1, 2
- [2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition.” In *CVPR*, 2016. 1, 2
- [3] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, “Deep networks with stochastic depth.” In *ECCV*, 2016. 1, 2
- [4] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, and W. H. L. D. Jackel, “Backpropagation applied to handwritten zip code recognition.” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989. 1
- [5] C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, “Deeply supervised nets.” In *AISTATS*, 2015. 1
- [6] M. Lin, Q. Chen, and S. Yan, “Network in network.” In *ICLR*, 2014. 1, 2
- [7] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting.” *JMLR*, 2014. 1

- [8] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," *In NIPS*, 2015. [1](#)
- [9] S. Yang and D. Ramanan, "Multi-scale recognition with dagnns." *In ICCV*, 2015. [1](#)