

# Densely Connected Convolutional Networks

Yufeng Jiang

May 29, 2018

## 1 Classification Results on CIFAR and SVHN

For example, our 250-layer model only has 15.3M parameters, but it consistently outperforms others models such as FractalNet and Wide ResNets that have more than 30M parameters. Figure 1 (right panel) 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.

**Overfitting.** One positive side-effect of the more efficient use of parameters is a tendency of DenseNets to be less prone to overfitting. On C10, the improvement denotes a 29% relative reduction in error from 7.33% to 5.19%. On C100, the reduction is about 30% from 28.20% to 19.64%. In this experiments, authors observed potential overfitting in a single setting: on C10, a  $4 \times$  growth of parameters produced by increasing  $\kappa = 12$  to  $\kappa = 24$  lead to a modest increase in error from 5.77% to 5.83%.

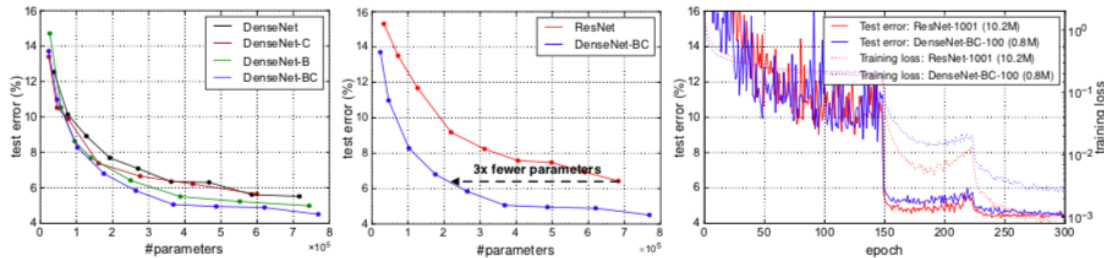


Figure 1: Left: Comparison of the parameter efficiency on C10+ between DenseNet variations. Middle: Comparison of the parameter efficiency between DenseNet-BC and (pre-activation) ResNets. DenseNet-BC requires about 1/3 of the parameters as ResNet to achieve comparable accuracy. Right: Training and testing curves of the 1001-layer pre-activation ResNet [1] with more than 10M parameters and a 100-layer DenseNet with only 0.8M parameters.

## 2 Classification Results on ImageNet

Authors evaluate DenseNet-BC with different depths and growth rates on the ImageNet classification task, and compare it with state-of-the-art ResNet architectures. To ensure a fair comparison between the two architectures, authors eliminate all other factors such as differences in data preprocessing and optimization settings by adopting the publicly available Torch implementation for ResNet by [2]<sup>1</sup>

Authors report the single-crop and 10-crop validation errors of DenseNets on ImageNet in Table 1. Figure 2 shows the single-crop top-1 validation errors of DenseNets and

ResNets as a function of the number of parameters (left) and FLOPs (right). The results presented in the figure reveal that DenseNets perform on par with the state-of-the-art ResNets. For example, a DenseNet-201 with 20M parameters model yields similar validation error as a 101-layer-ResNet with more than 40M parameters

It is worth noting that this experimental setup implies that authors use hyperparameter settings that are optimized for ResNets but not for DenseNets. It is conceivable that more extensive hyper-parameter searches may further improve the performance of DenseNet on ImageNet.

<sup>1</sup><https://github.com/facebook/fb.resnet.torch>

Model	top-1	top-5
DenseNet-121	25.02 / 23.61	7.71 / 6.66
DenseNet-169	23.80 / 22.08	6.85 / 5.92
DenseNet-201	22.58 / 21.46	6.34 / 5.54
DenseNet-264	22.15 / 20.80	6.12 / 5.29

Table 1: The top-1 and top-5 error rates on the ImageNet validation set, with single-crop / 10-crop testing.

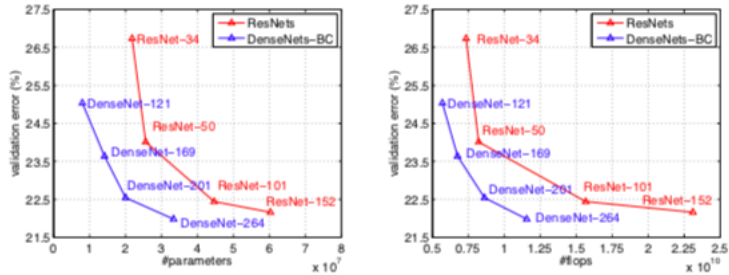


Figure 2: Comparison of the DenseNets and ResNets top-1 error rates (single-crop testing) on the ImageNet validation dataset as a function of learned parameters (left) and FLOPs during test-time (right).

### 3 Discussion

As a direct consequence of the input concatenation, the feature-maps learned by any of the DenseNet layers can be accessed by all subsequent layers. This encourages feature reuse throughout the network, and leads to more compact models.

**Model compactness.** The left two plots in Figure 1 show the result of an experiment that aims to compare the parameter efficiency of all variants of DenseNets (left) and also a comparable ResNet architecture (middle). In comparison with other popular network architectures, such as AlexNet [3] or VGG-net [4], ResNets with pre-activation use fewer parameters while typically achieving better results [1]. Hence, authors compare DenseNet ( $\kappa = 12$ ) against this architecture. The training setting for DenseNet is kept the same as in the previous section.

The graph shows that DenseNet-BC is the most parameter efficient variant of DenseNet. This result is in line with the

results on ImageNet presented in Figure 2. The right plot in Figure 1 shows that a DenseNet-BC with only 0.8M trainable parameters is able to achieve comparable accuracy as the 1001-layer ResNet [1] with 10.2M parameters.

### References

- [1] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks.” *In ECCV*, 2016. 1, 2
- [2] I. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, “Maxout networks.” *In ICML*, 2013. 1
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks.” *In NIPS*, 2012. 2
- [4] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, “Pedestrian detection with unsupervised multi-stage feature learning.” *In CVPR*, 2013. 2