Densely Connected Convolutional Networks

Yufeng Jiang

May 25, 2018

1 DenseNets

Pooling layers. The down-sampling layers change the size of feature-maps and are essential for convolutional networks. Authors divide the networks into multiple densely connected dense blocks to facilitate down-sampling in the architecture just as shown at Figure 1.

Growth rate. If each function H_ℓ produces κ feature-maps, it follows that the ℓ^{th} layer has $\kappa_0 + \kappa \times (\ell - 1)$ input feature-maps. κ_0 is the number of channels in the input layer. One explanation for this is that each layer has access to all the preceding feature-maps in its block. Each layer adds κ feature-maps of its own to this state.

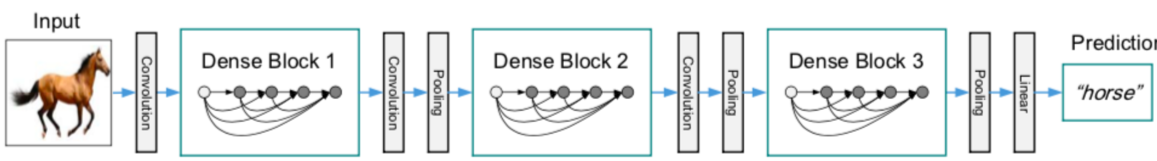

Figure 1: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 averge pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 averge pool, stride 2			
Dense Block (3)	28×28	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 averge pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		$2 \times$ 1000D fully-connected, softmax			

Table 1: DenseNet architectures for ImageNet. The growth rate for all the networks is $\kappa = 32$. Note that each conv layer shown in the table corresponds the sequence BN-ReLU-Conv.

Bottleneck layers. Although each layer only produces κ output feature-maps, it typically has many more inputs. It has been noted in [1,2] that a 1×1 convolution can be introduced as bottleneck layer before each 3×3 convolution to reduce the number of input feature-maps.

Compression To further improve model cpmpactness, authors reduce the number of feature-maps at transition layers. Authors let the following transition layer generate θ_m output feature-maps, where $0 < \theta \leq 1$ is referred to as the compression factor. In this experiments on ImageNet, author use a DenseNet-BC structure with 4 dense blocks on 224×224 input images. The exact network configureations are shown in Table1.

References

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

[2] R. K. Srivastava, K. Greff, and J. Schmidhuber., "Training very deep networks." *In NIPS*, 2015. 2