Dense Layer

Recent work has shown that convolutional networks can be deeper, more accurate, and more efficient with shorter connections between layers. Dense Convolutional Networks (DenseNets) build on this by connecting each layer to every other layer in a feed-forward way. Unlike traditional networks with L layers and Lconnections, DenseNets have L(L+1)/2 connections. This approach alleviates the vanishing-gradient problem, enhances feature propagation, encourages feature reuse, and reduces the number of parameters.

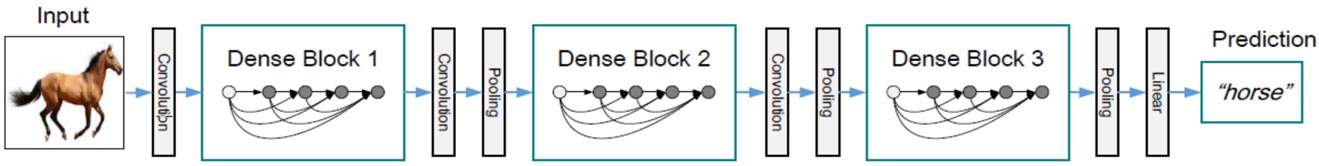

Figure 2: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling.

Dense Connectivity

Dense connectivity. To further improve the information flow between layers we propose a different connectivity pattern: we introduce direct connections from any layer to all subsequent layers. Figure 1 illustrates the layout of the resulting DenseNet schematically. Consequently, the ℓ^{th} layer receives the feature-maps of all preceding layers, $\mathbf{x}_0, \ldots, \mathbf{x}_{\ell-1}$, as input:

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}]), \tag{2}$$

where $[\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\ell-1}]$ refers to the concatenation of the feature-maps produced in layers $0, \ldots, \ell-1$. Because of its dense connectivity we refer to this network architecture as *Dense Convolutional Network (DenseNet)*. For ease of implementation, we concatenate the multiple inputs of $H_\ell(\cdot)$ in eq. (2) into a single tensor.

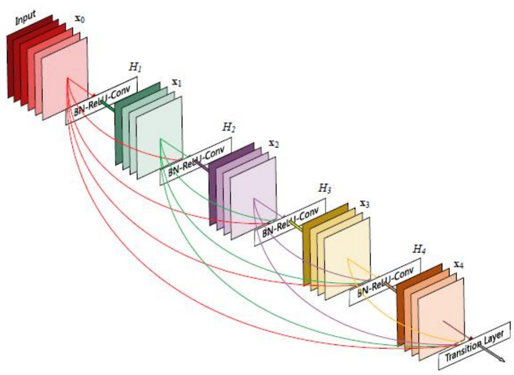

Figure 1: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

Growth Rate

In DenseNets, each function Hℓ produces k feature maps, so the ℓth layer has k0+k(ℓ−1) input feature maps, where k0 is the number of channels in the input layer. A key difference is that DenseNet can have very narrow layers, such as k=12. We call this hyperparameter k the growth rate of the network.

The growth rate regulates how much new information each layer contributes to the global state. The global state, once written, can be accessed from everywhere within the network