<u>Densely Connected Convolutional Networks - Summary</u>

1. Abstract-

- The network has L(L+1)/2 direct connections.
- Input of a layer is a feature map of all preceding layers.
- Alleviate the vanishing-gradient problem.
- Encourage feature reuse and substantially reduce the number of parameters.

2. Introduction -

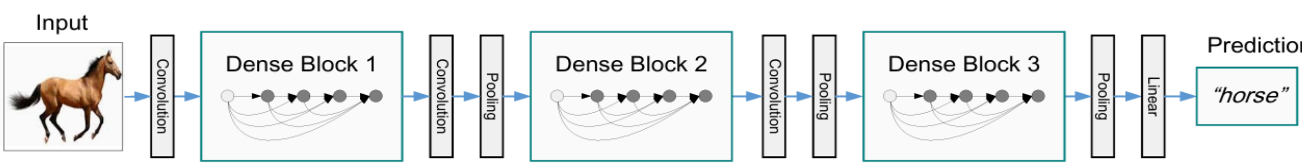
- As CNN becomes increasingly deep, the gradient can "wash out" when it reaches the end.
- ResNet ,Highway Net, and Fractal Net handle this by "creating short paths from early layers to later layers"
- Dense Net connects all layers directly with each other.
- It combines features by concatenating them. Hence, the lth layer has l inputs, consisting of the feature maps of all preceding convolutional blocks. Its feature maps are passed on to all subsequent L layers. This introduces L(L+1) 2 connections in an L-layer network.
- Densenet Function - x= Hl([x0,x1,x2… xl]), [x0,x1,x2… xl] refers to the concatenation of the feature maps produced in layers.
- Hl is a composite function of BN -> ReLU -> 3*3 Conv 2d.
- The feature maps of previous layers allow it to build on already-learned features rather than learning redundant ones.
- DenseNet doesn't require large feature maps, as information flows more effectively through the network, minimizing parameter usage.
- Growth Rate - If each function H produces k feature maps, it follows that the l th layer has k0+k*(l-1) input feature maps, with k = 12 only, k is known as growth rate.
- Each layer has direct access to the loss function and original input.
- Bottle Neck Layers - Introduces 1*1 conv before every 3*3 conv, which reduces the number of features.
  BN-ReLU-Conv(1 1)-BN-ReLU-Conv(3 3) , known as DenseNet B.
- Compression Factor
  If a dense block contains m feature maps, we let the following transition layer

generate m output feature maps,where 0< Theta <=1 is referred to as the compression factor.
We refer the DenseNet with Theta<1 as DenseNet-C, and we set Theta=0.5 in our experiment.

- Dense Net BC uses both Bottle Neck layers and Compression Factor.

3. Implementation Details

- It has 3 dense blocks with equal number of layers
- Before entering the dense block a convolution of 16 output channels is implemented.
- Transition Layer - 1*1 conv -> 2*2 averagepool 2d
- At the end, the Gloabal Avg pool is followed by the soft max.



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	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times 1 \text{ conv} \\ 3\times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			