- Densenet introduces connections from each layer to every other layer of the network.
    - It reduces vanishing gradients due to better gradient flow.
    - It increases feature reuse.
- Inputs of all previous layers are concatenated and fed into each layer.
- Resnets can become extremely deep as the preserve norm of gradient leading to stable back propagation : https://arxiv.org/pdf/1805.07477.pdf
- DenseNet layers are very narrow (e.g., 12 filters per layer), adding only a small set of feature-maps to the “collective knowledge” of the network and keep the remaining feature-maps unchanged—and the final classifier makes a decision based on all feature-maps in the network.
- Each layer has direct access to the gradients from the loss function and the original input signal, leading to an implicit deep supervision.
- Deeply-supervised nets (DSN) is a method that simultaneously minimizes classification error and improves the directness and transparency of the hidden layer learning process. Its internal layers are directly supervised by auxiliary classifiers, which can strengthen the gradients received by earlier layers.
- As loss function is directly connected to all layers by transition layers, it increases discriminative power of layers.
- It was observed that Dense connections have a regularisation effect.
- Highway networks use learned gating mechanisms to regulate information flow, inspired by Long Short-Term Memory (LSTM) recurrent neural networks.
- By using Stochastic Depth, the network is shortened during training, i.e. a subset of layers is randomly dropped and bypass them with the identity function.
- This shows that not all layers may be needed and highlights that there is a great amount of redundancy in deep (residual) networks.
- Pre-activation unit is on regularization that slightly increases training loss at convergence but with lower test error.
- Wide resnets proved that only increasing no. of filters of each layer is enough to increase accuracy if depth is sufficient.
- Concatenating feature-maps learned by different layers increases variation in the input of subsequent layers and improves efficiency.
- Information flow increases as each layer is directly connected to all subsequent layers.
- For downsampling, dense blocks are separated by transitional layers which have a BN layer, a 1x1 conv layer and 2x2 average pooling layer.
- If growth rate (no. of channels in output) is k, then each layer adds k feature maps to the ‘collective information’ of the network.
- As there are many inputs to each layer,bottlenecks are used to produce 4k feature map.
- For further compressing feature maps, bottlenecks are used at transition layers.
- At the last dense block, global average pooling is used before applying softmax.
- All layers spread their weights over many inputs within the same block. This indicates that features extracted by very early layers are, indeed, directly used by deep layers throughout the same dense block.

- The layers within the second and third dense block consistently assign the least weight to the outputs of the transition layer, indicating that the transition layer outputs many redundant features.In Densenet-BC, these outputs are compressed.
- Because of their compact internal representations and reduced feature redundancy, DenseNets may be good feature extractors for various computer vision tasks that build on convolutional features.
- It receives information from all previous layers, so it can manage with less no. of channels.