# Densely Connected Convolutional Networks

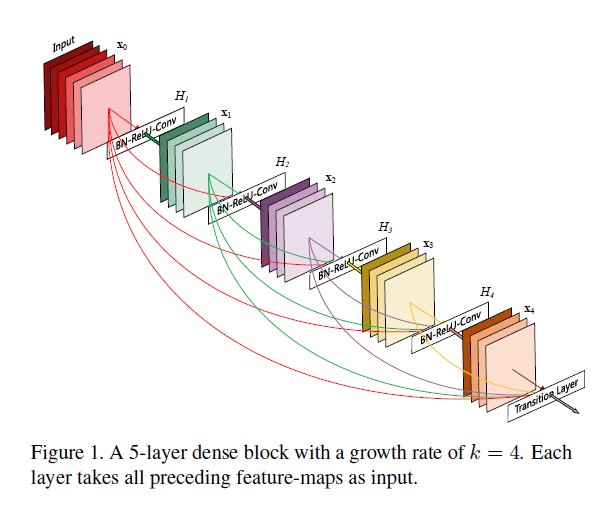
阅读参考<https://www.jianshu.com/p/08a2a1f2a814>

稠密的卷积神经网络（DenseNet），它将每一层以一种前馈的方式连接到其它每一层。传统的具有L层的卷积网络在每一层和它的后续层之间有L个连接，我们的网络拥有L(L+1)/2个直连接。对于每一层，使用前面所有层的特征图作为输入，它自己的特征图作为所有后续层的输入。

ResNets [11] and Highway Networks [33] bypass signal from one layer to the next via identity connections.

ResNet通过身份连接（identity connections）将信号从一层传到下一层。

In this paper, we propose an architecture that distills this insight into a simple connectivity pattern: to ensure maximum information flow between layers in the network, we connect all layers (with matching feature-map sizes) directly with each other. To preserve the feed-forward nature,each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers.  
  
值得注意的是：相比于ResNet，我们没有通过将特征求和的组合方式传到一个图层中，相反，我们使用了连接的组合方式。因此，第i层有i个输入，这些输入是由第i层之前的所有卷积块的特征图组成，它自己的特征图传向了其后L-i个后续层。这就引入了L层网络的L(L+1)/2个连接，而不是传统的L个连接。因为密集的连接模式，我们将这种方法叫做**密集卷积网络（DesneNet)**。



几个优点：

1. 、Our proposed DenseNet architecture explicitly differentiates between information that is added to the network and information that is preserved.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 featuremaps unchanged—and the final classifier makes a decision based on all feature-maps in the network

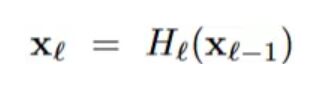
densnet 网络结构参数少，每个block里面的filter也比较少，而我们在使用alexnet，通常filter都是上百的，而这里的filter 12、24、16 等，所以非常narrow

1. one big advantage of DenseNets is their improved flow of information and gradients throughout the network, which makes them easy to train.每层都和后面的层相连，有利于信息和梯度在整个网络中的传递。

3.we also observe that dense connections have a regularizing effect, which reduces overfitting on tasks with smaller training set sizes.同时densenet网络也有正则化的作用，在小数据集上训练也能减少过拟合的风险

## DesNets

传统前馈卷积网络



在[5]ResNet中

IMG_256

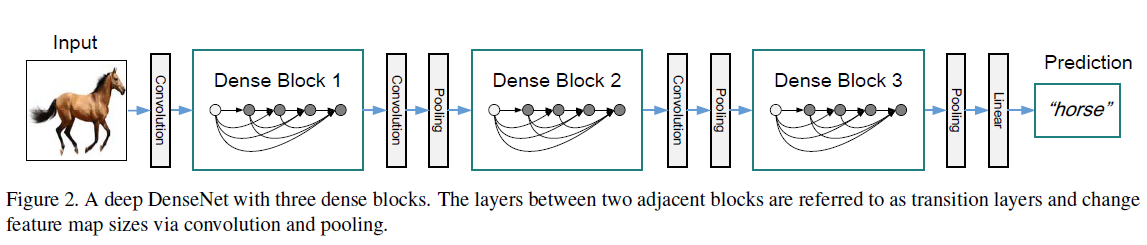
### Dense connectivity.

从任何层向其所有后续层的直连接

IMG_256

**Composite function.** Motivated by [12], we define H`() as a composite function of three consecutive operations: batch normalization (BN) [14], followed by a rectified linear unit (ReLU) [6] and a 3\*3 convolution (Conv).

我们将H\_l(·)定义为一个复合函数，它由三部分连续的操作组成：Batch Normalization、ReLU和一个3x3的卷积操作。



**Pooling layers**

如图2所示。我们将块之间的层称为过渡层，它们进行卷积和池化操作。

**Growth rate，**

如果每个H函数产生k个特征图，那么第l层的输入特征图总数为

IMG_256

**Bottleneck layers**

尽管每一层仅产生k个特征图，但它通常情况下拥有更多的输入。在每个3x3卷积前引入1x1卷积作为瓶颈层可以减少输入特征图的数量，从而可以提高计算效率

**Compression**

为了进一步提高模型的紧凑性，我们可以在过渡层减少特征图的数量。如果一个密集块包含m个特征图，我们让其后的过渡层产生一个IMG_256

**Implenentation Details**

除了ImgageNet以外的所有数据集，我们实验中使用的DenseNet都有三个密集块，每个块的层数都是相等的。在进入第一个紧密块之前，在输入图像上进行一个16通道的卷积。对于卷积核大小为3x3的卷积层，输入的每一侧都被填充一个像素以保持固定的特征图的大小。我们在1x1的卷积后跟一个2x2的平均池化层作为过渡层连接两个相邻的密集块。在最后一个密集块之后，执行一个全局平均池化，然后附加一个softmax分类器。三个紧密块的特征图尺寸分别为32x32、16x16和8x8。

实验结果

DenseNet-BC 要达到和 ResNet 一样的效果，只需要其三分之一的参数量即可。