

Densely Connected Convolutional Networks

Xuwen Yang

June 10 2018

Abstract

Recent work has shown that convolutional networks can be substantially deeper, more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. In this paper, we embrace this observation and introduce the Dense Convolutional Network (DenseNet), which connects each layer to every other layer in a feed-forward fashion.

1. Introduction

Convolutional neural networks (CNNs) have become the dominant machine learning approach for visual object recognition. Although they were originally introduced over 20 years ago, improvements in computer hardware and network structure have enabled the training of truly deep CNNs only recently. The original LeNet5 consisted of 5 layers, VGG featured 19 [3], and only last year Highway Networks and Residual Networks (ResNets) have surpassed the 100-layer barrier.

As CNNs become increasingly deep, a new research problem emerges: as information about the input or gradient passes through many layers, it can vanish and wash out by the time it reaches the end (or beginning) of the network. Many recent publications address this or related problems. ResNets and Highway Networks bypass signal from one layer to the next via identity connections. Stochastic depth [2] shortens ResNets by randomly dropping layers during training to allow better information and gradient flow. FractalNets repeatedly combine several parallel layer sequences with different number of convolutional blocks to obtain a large nominal depth, while maintaining many short paths in the network. Although these different approaches vary in network topology and training procedure, they all share a key characteristic: they create short paths from early layers to later layers. 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 p-

reserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers. Figure 1 illustrates this layout schematically. Crucially, in contrast to ResNets, we never combine features through summation before they are passed into a layer.

2. Related Work

The exploration of network architectures has been a part of neural network research since their initial discovery. The recent resurgence in popularity of neural networks has also revived this research domain. The increasing number of layers in modern networks amplifies the differences between architectures and motivates the exploration of different connectivity patterns and the revisiting of old research ideas.

A cascade structure similar to our proposed dense network layout has already been studied in the neural networks literature in the 1980s. Their pioneering work focuses on fully connected multi-layer perceptrons trained in a layer-by-layer fashion. More recently, fully connected cascade networks to be trained with batch gradient descent were proposed. Although effective on small datasets, this approach only scales to networks with a few hundred parameters. In [1], utilizing multi-level features in CNNs through skip-connections has been found to be effective for various vision tasks. Parallel to our work, it derived a purely theoretical framework for networks with cross-layer connections similar to ours. Highway Networks were amongst the first architectures that provided a means to effectively train end-to-end networks with more than 100 layers. Using bypassing paths along with gating units, Highway Networks with hundreds of layers can be optimized without difficulty. The bypassing paths are presumed to be the key factor that eases the training of these very deep networks. This point is further supported by ResNets, in which pure identity mappings are used as bypassing paths. ResNets have achieved impressive, record-breaking performance on many challenging image recognition, localization, and detection tasks, such as ImageNet and COCO object detection. Recently, stochastic depth was proposed as a way to

successfully train a 1202-layer ResNet. Stochastic depth improves the training of deep residual networks by dropping layers randomly during training. This shows that not all layers may be needed and highlights that there is a great amount of redundancy in deep (residual) networks. Our paper was partly inspired by that observation. ResNets with pre-activation also facilitate the training of state-of-the-art networks with > 1000 layers.

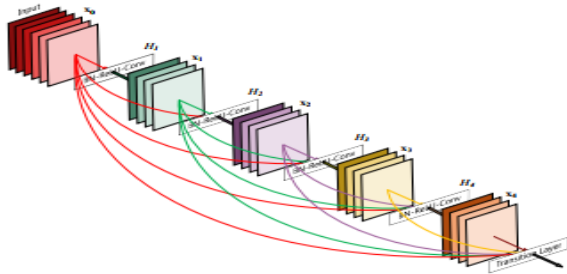


Figure 1. A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

References

- [1] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In *CVPR*, 2015. 1
- [2] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. Deep networks with stochastic depth. In *ECCV*, 2016. 1
- [3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. ImageNet large scale visual recognition challenge. *IJCV*, 2014. 1