**Paper Summary**

Haocong Wang

**Paper Title:** Densely Connected Convolutional Networks

**Authors:** Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger

**Introduction**

Convolutional neural networks (CNNs) were introduced 30 years ago, but the truly deep CNNs only started to be finished training in recent years due to the development of computer hardware and network structures. However, along with the rising of the number of layers in CNNs, the information passing through the networks may vanish when they reach the end or the beginning of the networks. Many research groups proposed different methods to solve this problem, which share a same key characteristic: they all create short paths between layers.

In this paper, the authors propose a new architecture of CNNs to ensure maximum information flow between layers in the network by connecting all layers directly with each other, which leads to connections in a L-layer network. According to the dense connectivity pattern of the proposed network, it is named as Dense Convolutional Network (DenseNet).

**Approach and Methodology**

Suppose that is a single image passed through a neural network, which has layers. Each layer implements a non-linear transformation , where indexes the layer. As mentioned in the introduction, the layer receives the feature-maps of all preceding layers, , as inputs: . is defined as a composite function including three consecutive operations: batch normalization (BN), a rectified linear unit (ReLU) and a convolution (Conv).

图示, 示意图

描述已自动生成

Fig 1: An example of DenseNet with three dense blocks. Between two dense blocks, there is a transition layer including convolution and pooling, which can change feature-map sizes. [1]

As shown in Fig 1, authors add a layer called transition layer between two dense blocks, which will finish convolution and pooling. The transition layer consists of a batch normalization layer and a convolutional layer followed by a average pooling layer.

Suppose each function produces feature-maps. is denoted as the growth rate of the network, which can be small in the DenseNet. The reason is that each layer has access to all the preceding feature-maps in its block and, therefore, to the network’s “collective knowledge”. On the other hand, despite the small amount of the output feature-maps, each layer has much more inputs. To reduce the number of input feature-maps, authors add a bottleneck layer including a convolutional layer. One parameter, , can be added in the transition layer to reduce the number of feature-maps. DenseNet with a bottleneck and as marked as DenseNet-BC.

**Performance Evaluation**

Several benchmark datasets, CIFAR, Street View House Numbers (SVHN) and ImageNet, are used for evaluating the performance of DenseNet, which will be compared with other state-of-the-art architectures. Here I take the results on CIFAR and SVHN as examples.

表格

描述已自动生成

Fig 2: Error rates on CIFAR and SVHN. C10 and C100 represent images drawn from 10 and 100 classes respectively. “+” marks the datasets using standard data augmentation scheme. Results that surpass all competing methods are bold and the overall best results are blue. DenseNet achieves lower error rates compared to ResNet with fewer parameters. [1]

All the networks are trained with stochastic gradient descent (SGD). The batch size is 64. For CIFAR and SVHN, there are 300 and 40 epochs respectively. At beginning, the learning rate is set to be 0.1, and is divided by 10 at 50% and 75% of the total number of the training epochs.

The accuracy of DenseNet is much lower than existing architectures on all datasets with fewer parameters. The smaller number of parameters shows high parameter efficiency. When and the depth of the networks, marked as , increase, DenseNet performs better, which shows the capacity of bigger and deeper models and the ability to avoid overfitting.

**Related Work**

A cascade network that is similar to DenseNet was proposed decades ago [2]. In 2010, Wilamowski et al. trained a fully connected cascade network with batch gradient descent [3]. However, this approach only scales to networks with a few hundred parameters.

In ResNet [4], pure identity mappings are used as bypassing paths, which help achieve impressive, record-breaking performance on many challenging image recognition, localization, and detection tasks. In 2016, Huang et al. [5] utilized stochastic depth to successfully train a 1202-layer ResNet by dropping layers randomly during training, which also inspires this paper partly. Targ er al. [6] propose a variant of ResNets with wide generalized residual blocks, which makes the network deeper by increasing the network width.

Compared to all the above architectures, DenseNets are simpler and more efficient. DenseNets exploit the potential of the network through feature reuse, leading to condensed models that are easy to train and highly parameter-efficient.

**References**

[1] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[2] Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In Advances in neural information processing systems (pp. 524-532).

[3] Wilamowski, B. M., & Yu, H. (2010). Neural network learning without backpropagation. IEEE Transactions on Neural Networks, 21(11), 1793-1803.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[5] Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. Q. (2016, October). Deep networks with stochastic depth. In European conference on computer vision (pp. 646-661). Springer, Cham.

[6] Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029.