

# 基于 DenseNet 的图像识别方法研究

高建瓴\*, 王竣生, 王 许

(贵州大学 大数据与信息工程学院, 贵州 贵阳 550025)

**摘 要:** 在深度学习中, 随着神经网络层数的加深, 训练网络变得越来越困难, 现有的浅层的网络无法明显提升网络的识别效果。针对在原有网络的情况下提升网络识别效果, 减少参数的问题, 本文提出一种改进的 DenseNet 网络模型, 借鉴 Inception V3 的思想, 利用非对称卷积将 DenseNet 网络中 Dense Block 层所包含的  $3 \times 3$  卷积分解成  $3 \times 1$  和  $1 \times 3$  的两个卷积, 简化网络结构模型。之后再对改进前与改进后的网络模型在数据集上进行训练, 实验结果表明, 与经典的 DenseNet 网络模型相比, 改进后的网络模型可以提高图像识别的准确性, 降低参数, 有效地缩短运行时间。

**关键词:** DenseNet 网络; 图像识别; 深度学习; 卷积神经网络

中图分类号: TP391.4

文献标识码: A

图像识别现今广泛应用于工业、医学、地理学以及人工智能, 其中图像识别是人工智能的一个重要应用领域。

1989 年, LE<sup>[1]</sup> 首次提出卷积神经网络, 在之后的几十年里卷积神经网络由于技术限制等因素使其发展速度较为缓慢。直到 2012 年提出 AlexNet<sup>[2]</sup> 网络, 卷积神经网络得到了急速发展, 在计算机视觉、计算机图形学、自然语言处理等领域上取得了较好的结果, 但 AlexNet 网络相比之前其他的卷积网络最显著的特点是参数规模变大, 网络层数更深。2014 年, Google 提出了 GoLeNet<sup>[3]</sup> 网络, 其主要创新 Inception 机制, 即对图像进行多尺度处理, 主要解决了网络参数增加之后出现过拟合, 以及计算量的增加问题。同年牛津大学视觉组 (Visual Geometry Group) 提出了 VGG<sup>[4]</sup> 网络, 相对于 AlexNet 网络, VGG 网络去掉了 LRN 层, 同时采取小尺寸的卷积核来模拟更大尺寸的卷积核, 从而减少了参数量。2015 年, HE 等提出了残差网络 (ResNet)<sup>[5]</sup>, 通过使用 Residual Unit 训练出 152 层深的神经网络, 该网络一经出世, 便在 ImageNet 中获得图像分类、检测和定位的冠军, 其中在图像分类的数据集上达到了 3.57% 的错误率, 同时参数量比 VGG 网络低。2017 年, GAO 等提出

了 DenseNet<sup>[6]</sup>, 该网络为全新的结构, 网络结构不复杂, 却非常有效, 在 CIFAR 指标上优于残差网络 (ResNet)。可以说 DenseNet 吸收了 ResNet 最精华的部分, 并做了更多创新的工作, 使得网络性能进一步提升。

卷积神经网络在图像识别中的应用仍然存在一些不足: 上述网络中卷积层包含参数过多会导致计算量过大; 随着网络层数的加深, 其网络结构变得更加复杂, 模型结构有待优化。本文针对存在的两个问题展开研究, 以期解决参数过大, 简化 DenseNet 网络结构模型。

## 1 DenseNet 网络

DenseNet 是一种具有密集连接的卷积神经网络, 该架构含有卷积层、池化层, 增长率为 4 的稠密连接模块, 其中增长率是为了保持通道特征维度适中, 以防过大。如图 1 所示, DenseNet 包含多个 Dense Block 模块, 其中 Dense Block 由 BN + ReLU + Conv( $1 \times 1$ ) + BN + ReLU + Conv( $3 \times 3$ ) 组成, Dense Block 之间的层称为 transition layers, 由 BN + Conv( $1 \times 1$ ) + Average Pooling( $2 \times 2$ ) 组成。由于输出的 feature map 维度是一个很大的数目, 所以在 transition layer 模块中加入了  $1 \times 1$  的卷积做降维, 从而提升计算效率。

收稿日期: 2019-03-25

基金项目: 贵州省科技厅基金项目资助 (黔科合 [2015]2045 号)

作者简介: 高建瓴 (1969-), 女, 副教授, 硕士, 研究方向: 图像处理, Email: 1145373761@qq.com.

\* 通讯作者: 高建瓴, Email: 1145373761@qq.com.

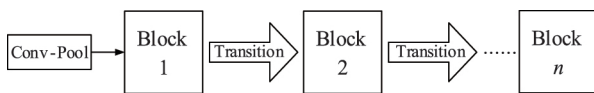


图1 DenseNet 的网络结构示意图

Fig.1 The Structural sketch of DenseNet

在 DenseNet 网络中,任意两层之间都有着密不可分的联系,即网络中任意一个下一层的输入都是前面所有层输出结果的叠加,而该层所得到的结果也将和之前的输出作为输入传输给下一层,这样依次向上传输。将之前所有层的特征输出作为当前层的输入,即  $x_0, x_1, \dots, x_{l-1}$  为第 1 层到第  $l-1$  层的特征图,通过级联的连接方式经过第  $l$  层,最后用复合函数  $H_l(\cdot)$  得到输出  $x_l$ ,即

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (1)$$

$H_l(\cdot)$  定义为依次经过  $3 \times 3$  的卷积 (Conv2d)、池化 (pool)、批量归一化 (BN) 以及非线性激活函数 (Relu) 四种运算的复合函数。

这种密集连接不会带来冗余,每层网络计算量的减少以及特征的重复利用使得 DenseNet 比其他网络效率更高。该方式能进一步提高模型的容量和参数利用率,在准确率上也比传统的残差网络 (ResNet) 要高。

一般的  $L$  层卷积神经网络中,总共有  $L$  层连接,而 DenseNet 有  $\frac{L(L+1)}{2}$  层连接 (DenseNet 采用级联的连接方式)。

DenseNet 网络结构没有按照以往用加深网络或者加宽网络的方式入手,而是对特征序列进行调整;另一方面,该网络每一层都有非常浅的特征图或非常少的卷积核,因此简化了网络模型参数,从而在某些领域中达到了较好的效果。总体来说 DenseNet 有如下优点:

- (1) 一定程度上解决梯度消失;
- (2) 加强特征传播;
- (3) 更有效地利用特征;
- (4) 减少参数数量。

## 2 Inception V3 网络

Inception V3 网络引入了 Factorization into small convolutions 的思想,思想的主旨就是把一个较大的二维卷积拆分成两个较小的一维卷积,一方面节约了大量参数加速运算,同时减轻了过拟合,除此之外该网络还增加了一层非线性扩展模型表达能力。例如将  $7 \times 7$  的卷积拆成  $1 \times 7$  和  $7 \times 1$  两个卷积的并联或者串联,比拆成 3 个  $3 \times 3$  的卷积更节

约参数,卷积效果更好,而且可以处理更细、更全面的特征空间,有效地提高了特征多样性。

Inception V3 还优化了 Inception Module 的结构,现有的 Inception Module 有  $35 \times 35$ 、 $17 \times 17$  和  $8 \times 8$  三种结构,像这 3 种结构的 Inception Module 一般都只能放在卷积层之后,该卷积层为普通的卷积层。Inception V3 网络的核心组件采用了非对称卷积 ( $N \times N$  分解成  $N \times 1$  和  $1 \times N$  的叠加),该方法能减少很多计算量,如图 2 所示。

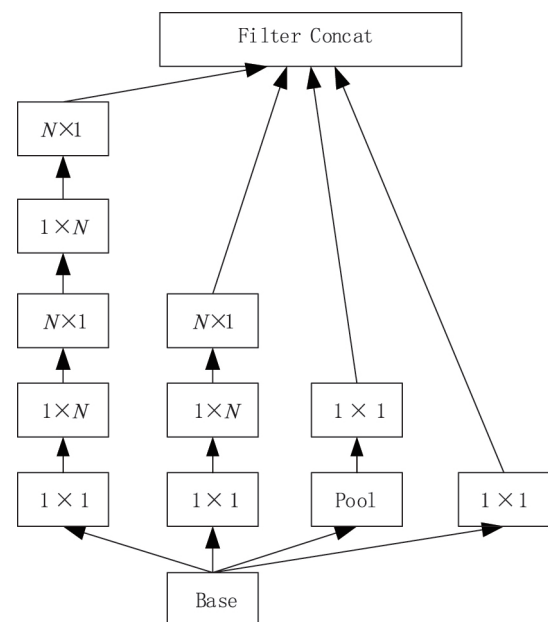
图2  $N \times N$  卷积结构分解为  $1 \times N$  和  $N \times 1$  的卷积结构相叠加Fig.2  $N \times N$  convolution structure is decomposed into the superposition of  $1 \times N$  and  $N \times 1$  convolution structure

图 2 中假设最左侧为  $5 \times 5$  的卷积,可表示为 2 个  $3 \times 3$  卷积,利用非对称卷积的思想又可以表示为首先执行一个  $3 \times 1$  的卷积再执行一个  $1 \times 3$  的卷积,这种方法在成本上要比单个  $3 \times 3$  的卷积降低 33%。

## 3 改进的 DenseNet 网络

受到 Inception V3 网络结构的启发,结合非对称卷积的特点,本文提出一种改进的 DenseNet 网络,其网络结构如表 1 所示。网络首先在输入图像上做一次  $3 \times 3$  步长为 2 的卷积,然后做  $3 \times 3$  步长为 2 的最大池化操作,接着进入到第一个 Dense Block,一共用到 3 个 Dense Block。在 Dense Block 中,该模块的主要作用在于一直反复不停地做密集连接形式的运算,每一层都将利用密集连接运算使得上一层的特征图与本层所产生的特征图合并,最终 Dense Block 的最后一层将会得到模块中所有特征图的并集。

表 1 改进的 DenseNet-40 网络结构图  
Tab.1 Improved DenseNet-40 Structure Diagram

Layers	Output Size	DenseNet-40
Convolution	32×32×16	3×3 conv
BN	32×32×16	BN( Batch Normalization)
Dense Block(1)	32×32×160	$\left[ \begin{array}{c} 1 \times 3 \text{ conv} \\ 3 \times 1 \text{ conv} \end{array} \right] \times 12$
Transition Layer(1)	32×32×160 16×16×160	1×1 conv 2×2 average pool ,stride 2
BN	16×16×304	BN( Batch Normalization)
Dense Block(2)	16×16×304	$\left[ \begin{array}{c} 1 \times 3 \text{ conv} \\ 3 \times 1 \text{ conv} \end{array} \right] \times 12$
Transition Layer(2)	16×16×448 8×8×448	1×1 conv 2×2 average pool ,stride 2
BN	8×8×448	BN( Batch Normalization)
Dense Block(3)	8×8×448	$\left[ \begin{array}{c} 1 \times 3 \text{ conv} \\ 3 \times 1 \text{ conv} \end{array} \right] \times 12$
Classification Layer	1×1×448 10	8×8 global average pool 10D fully-connected ,softmax

第一个 Dense Block 中包含了 12 个  $1 \times 3$ 、 $3 \times 1$  的卷积操作,第 12 个子结构的输入是前面 11 层的输出结果,每层输出的通道数是 12 (growth rate),其中为了避免网络增长过于迅速,并优化参数效率,拟定增长率为一个较小的整数,本文增长率设置为 12。原始 DenseNet 网络中,在每个  $3 \times 3$  卷积层之后都会接上一个  $1 \times 1$  卷积层,以减少输入到卷积层中特征图的数量,提高网络运行速率,随之计算效率也相应提高,经对比发现这个设计对 DenseNet 网络非常有效,所以将一维卷积层保留到本文的网络中。

每两个 Dense Block 中间都包含了 Transition Layer。在 Transition Layer 中,参数 reduction (取值为  $0 \sim 1$ ) 代表输出缩小为原来的多少倍,本文 reduction 设置为 0.5。Dense Block 模块紧接着经过之后的 Transition Layer,即一个  $1 \times 1$  卷积、 $2 \times 2$  的平均池化以及批量归一化(BN),下一个模块结构和之前的相类似,唯一不一样在于输出的特征图尺寸越来越小,网络的深度在逐渐加深。网络的最后一层为分离 10 类的 Softmax 层,采用全连接结构。

网络结构代码由 Keras 框架实现。代码实现用  $1 \times 3$  和  $3 \times 1$  的卷积来代替所有 Dense Block 中  $3 \times 3$  的卷积,达到节约参数加速运算,以及减轻过拟合的目的。

## 4 实验结果与分析

为验证改进后 DenseNet 网络的性能,本文在 CIFAR-10 数据集上测试识别率和损失函数,并与经典的 DenseNet 网络进行比较分析。

### 4.1 数据集

本次实验的数据集为 CIFAR-10 数据集,该数据集包含 10 个类(飞机,汽车,鸟,猫,鹿,狗,青蛙,马,船,卡车)共 60 000 张  $32 \times 32$  的彩色图像,每个类别分别有 6 000 张不同图像。本文将 50 000 张图片作为训练图像,10 000 张作为测试图像来进行实验。

### 4.2 实验设置

实验基于 Keras 框架。训练使用计算机: Intel Core i7 处理器,内存为 8 GB。模型的初始学习率设为 0.000 1,批量的大小为 32 (batch size),采用的优化器是 Adam 算法,迭代的次数为 40 次 (epoch),其中,Adam 方法能够在训练的过程中自适应地调整各个参数的学习率,非常适合于高维度和大数据集的非凸优化<sup>[10]</sup>。

### 4.3 实验结果及分析

实验中使用 3 个 Dense Block,每个 Dense Block 内部有两个卷积层,其卷积层分别为  $1 \times 3$  和  $3 \times 1$  的卷积层,同时因为 Dense Block 后接  $1 \times 1$  的卷积层,参数也得到精简。实验结果见表 2。

表 2 实验结果

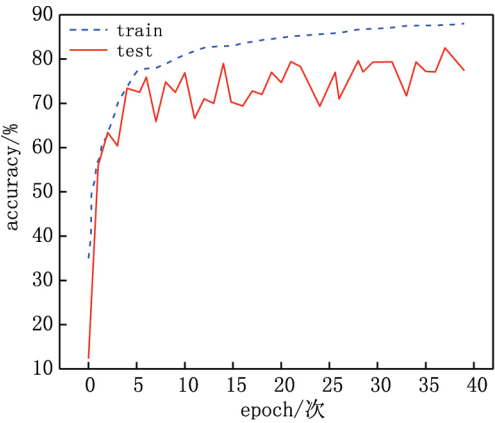
Tab.2 Experimental data sheet

识别方法	识别率/%	参数量	损失函数
经典 DenseNet-40	80.62	1 003 546	0.84
改进 DenseNet-40	86.16	433 306	0.71

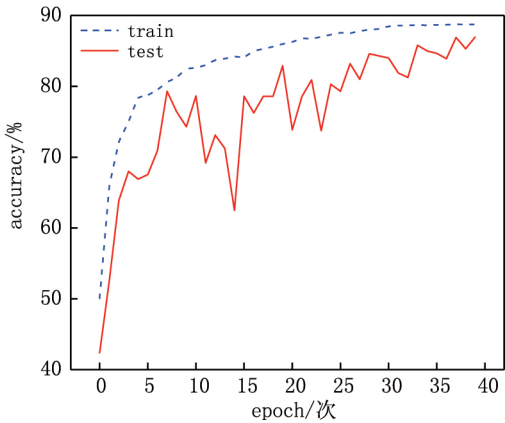
从表 2 可知 在实验数据不变的条件下 较为经典的 DenseNet-40 网络模型的最高识别率只有 80.62% 改进后的 DenseNet-40 网络模型的最高识别率为 86.16%。实验结果表明改进后的 DenseNet-40 网络模型识别率高于经典的 DenseNet-40 网络模型, 而且网络收敛速度也比经典的快。另一方面, 参数量从原来的 1 003 546 降低到现在的 433 306 极大

地缩短了程序运行时间。

本次实验结果在总卷积层数保持不变的情况下 通过对 Dense Block 内部的结构进行调整 观测识别率变化程度。经典的 DenseNet-40 网络模型与改进后的 DenseNet-40 网络模型对训练集与测试集的识别精度如图 3 所示。由图可知: 经过 40 次网络迭代 改进后的 DenseNet-40 网络模型与经典的 DenseNet-40 网络模型都逐渐趋于收敛、平稳, 改进后的网络模型较经典的收敛更加迅速; 当迭代第 40 次时 本文提出的网络模型识别准确率较原网络高出 9.60%; 体现改进的 DenseNet-40 网络模型在收敛速度及识别率方面的优越性。



(a) 经典的 DenseNet 网络



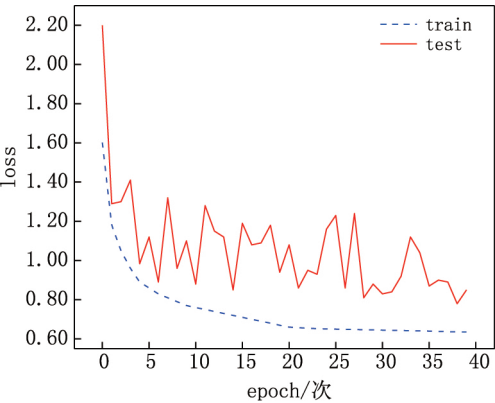
(b) 改进后的 DenseNet 网络

图 3 识别率对比图

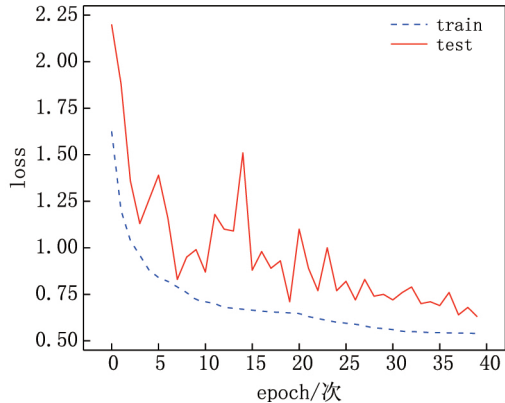
Fig.3 Recognition rate contrast map

经典的 DenseNet-40 网络模型与改进后的 DenseNet-40 网络模型的训练与测试过程中损失函数曲线如图 4 所示。可以看出 随着网络迭代次数的增加 损失函数值也趋于平稳, 当迭代 40 次时,

改进后的 DenseNet-40 网络模型损失函数值较经典的 DenseNet-40 网络模型损失函数值降低了 0.22 与预测值更加接近, 网络模型拟合度也相应的提高。



(a) 经典的 DenseNet 网络



(b) 改进后的 DenseNet 网络

图 4 损失函数对比图

Fig.4 Loss function contrast diagram

## 5 结论

本文利用 Inception V3 的核心思想之一(非对称卷积)对 DenseNet 网络模型中 Dense Block 模块进行重新构建优化,对改进后的 DenseNet 网络模型进行训练与测试,经实验结果分析,改进的 DenseNet 网络结构效果满足实验之前预期目标:其减少了随网络层数的增加而出现参数过多的问题,即在迭代次数不断增加的条件下,收敛速度、识别率都有明显的提高,结构得到了优化,网络模型的损失函数值也在不断降低,有较好的泛化能力。但是本次实验也有些许缺点,在网络设置的深度上并没有达到很深的层次,下一步将把该改进后的网络应运到更深层的结构上,而且对不同的数据集进行验证,观察实验效果。

### 参考文献:

- [1] LE C, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [2] ALEX K, ILYA S, GEOFFREY E H. ImageNet Classification with Deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 6(6): 84-90.
- [3] SZEGEDY C, LIU W, YANG Q J, et al. Going Deeper with Convolutions [C]//Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2014: 7-12.
- [4] IMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014, 75: 23-36.
- [5] HE K M, ZHANG X G, REN S Q. Deep Residual Learning for Image Recognition [C]//Computer Vision and Pattern Recognition, USA: IEEE, 2016: 770-778.
- [6] GAO H, ZHANG L. Densely Connected Convolutional Networks [C]//Computer Vision and Pattern Recognition, USA: IEEE, 2017: 1-28.
- [7] 张艺玮, 赵一嘉, 王馨悦, 等. 结合密集神经网络与长短时记忆模型的中文识别 [J]. 计算机系统应用, 2018, 27(11): 35-41.
- [8] XU H, SU F. A robust hierarchical detection method for scene text based on convolutional neural networks. Multimedia and Expo [C]//International Conference on Electrical Engineering, USA: IEEE, 2015: 1-6.
- [9] 常亮, 邓小明, 周明全. 图像理解中的卷积神经网络 [J]. 自动化学报, 2016, 42(9): 1300-1312.
- [10] HE K, ZHANG X, REN S, et al. Identity Mappings in Deep Residual Networks [J]. European Conference on Computer Vision, Springer, 2016, 32(6): 630-645.
- [11] 彭淑敏, 王军宁. 基于神经网络的图像识别方法 [J]. 电子科技杂志, 2005, 32(1): 39-42.

(责任编辑: 周晓南)

## Research on Image Recognition Method Based on DenseNet

GAO Jianling\*, WANG Junsheng, WANG Xu

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

**Abstract:** With the deepening of the layers of neural network in deep learning, the training network becomes more and more difficult, and the existing shallower network can not significantly improve the recognition effect of the network. In order to improve the network recognition effect and reduce the parameters in the case of the original network, an improved DenseNet network model was proposed. Based on the idea of Inception V3, the  $3 \times 3$  convolution contained in Dense Block layer in DenseNet network was decomposed into  $3 \times 1$  and  $1 \times 3$  convolutions by using asymmetric convolution, which simplifies the network structure model. The network was trained before and after the improvement on the dataset. The experimental results show that compared with the classical DenseNet network model, the improved network model can improve the accuracy of image recognition, reduce the parameters and shorten the running time effectively.

**Key words:** DenseNet network; image recognition; deep learning; convolutional neural network