电光与控制
Electronics Optics & Control
ISSN 1671-637X,CN 41-1227/TN

《电光与控制》网络首发论文

题目：　基于双网络级联卷积神经网络的设计
作者：　潘兵，曾上游，杨远飞，周悦，冯燕燕
网络首发日期：　2018-11-19
引用格式：　潘兵，曾上游，杨远飞，周悦，冯燕燕．基于双网络级联卷积神经网络的设计[J/OL]．电光与控制.
http://kns.cnki.net/kcms/detail/41.1227.tn.20181115.1356.020.html

基于双网络级联卷积神经网络的设计

潘兵，曾上游，杨远飞，周悦，冯燕燕
（广西师范大学 电子工程学院，广西 桂林 541004）

摘　要：传统的卷积神经网络通常采用单一的网络结构进行特征提取，但是单一网络结构提取的特征不够充分，导致图片分类的精度不高。本文针对这一问题提出了采用两种网络同时进行特征提取，再将两种网络级联在一起，得到两种网络的融合特征，使提取的特征更具用辨别性。双网络级联是采用两条支路进行特征提取，一条支路为传统的 CNN，另一条支路为在传统的 CNN 基础上加上残差操作，在下一次特征图降维前通过级联操作将两条不同的网络支路结合在一起。本网络实验采用 101_food 和 caltech256 数据集中进行测试，将级联后的网络和两条支路网络进行对比，实验最后表现出来较好的结果。

关键词：卷积神经网络;图像识别;网络级联;特征图
中图分类号：TP391.4　　**文献标志码：**A

Design of Double Envelope Network Based on Convolution Neural Network

Pan Bing,Zeng Shangyou,Yang Yuanfei,Zhou Yue,Feng Yanyan
（College of Electronic Engineering Guangxi Normal University,Guilin Guangxi 541004 ）

Abstract: Conventional convolutional neural networks usually use a single network structure for feature extraction. However, the feature extraction of single network structure is not sufficient, and then it may induce the poor accuracy of the classification of the images. In this paper, aiming at this problem, two networks are proposed to extract features at the same time. Then the two networks are cascaded together to obtain the fusion features of the two networks, which make the extracted features more discriminative. The dual-network cascade uses two branches for feature extraction. One branch is the traditional CNN. The other branch is the traditional CNN plus residual operation. Before the next feature reduction, two different network branches are put together. We use the 101_food and caltech256 data set for testing in the network exeperiment. The cascaded network and two branches of the network are compared, and the experiment finally shows better results.

Keywords: convolutional neural network; image recognition; network cascade; feature map

0　引言

近年来，卷积神经网络(Convolutional Neural Networks, CNN)在飞速的发展。在诸多领域中都有或多或少的涉及，特别在图像处理方面引起了研究者的广泛关注。而在这几年的发展中，针对不同的问题提出来的不同网络，有很多经典的网络结构。为对这方面有兴趣的研究者，提供了很多的帮助。为了解决数据较少问题，2012 年 Krizhevsky 等人提出了一种名为 AlexNet[1]的网络，并提出在全连接层使用 Dropout 技术以防止过拟合。为了解决 AlexNet 的在卷积层次较浅的问题，在 2014 年的 ImageNet 大赛上，Andrew 教授所领导的研究组提出的了一种名为 VGGNet[2]的网络，其很好的继承了 AlexNet 衣钵，并提出了自己的观点——更深。

基金项目：国家自然科学基金（11465004）
作者简介：潘兵（1993-），男，江西上饶人，硕士，研究方向为深度学习。

Christian Szegedy 等人组建的 Google 团队提出的 GoogleNet[3]则除了在深度较之 AlexNet 有所增加以外，还提出了加宽网络，主要是使用一种 Inception 的结构取代传统网络中的一个卷积层。为了解决网络在深层次中出现的退化问题何凯明博士团队 2016 年提出了一种 ResNet[4]结构，以及在 2016 年的 Imagenet 上展露头角的 DenseNet[5]。然而，无论是比较早提出 AlexNet 还是最新的 DenseNet，在网络结构上都比较单一以至于在提取特征时不够充分。为了更好的表现出网络性能，本文提出采用双网络级联的方法，将两种单一的网路通过级联的方式重新组合在一起，通过这种方法能有效的提高网路的并行度从而提升网络性能。

1 卷积神经网络概述

1.1 卷积神经网络

卷积神经网络最初始模仿于生物神经网络结构，通过一个个的神经元链接感受野将信息经过一系列的加工传递到下一个神经元。卷积神经网络就是通过一层层的结构将上一层的信息传递到下一层。现公认的第一个能工程实现的卷积神经网络为 BP 网络，可以追溯到 1986 年 BP[6]算法的题出，之后在 1989 年 LeCun 等人将其用到多层神经网络中，直到 1998 年 LeCun 正式提出 LeNet5[7]模型，神经网络的雏形逐渐成型。然而在接下来的十年的时间里，卷积神经网络的发展处于暂停状态，直到 2006 年 Hinton 等人[8]在《Science》上提出了深度学习的概念。从此，卷积神经网络再次迎来新的发展期，并取得长足的发展。传统的卷积神经络主要结构有输入层、卷积层、池化层及全连接层。

1.2 卷积神经网络的一般结构

1.2.1 卷积层

卷积层是由多个特征图组成，每一个特征图是由多个神经元组成，而每一个神经元通过卷积核与上一层的特征图进行卷积运算得出。卷积核为一个权值矩阵[9]，涵盖网络需要学习的内容，它包括权值 W（weight）和 b 偏置 （bias）。此处的卷积运算不同于信号处理中一维的卷积运算，而是二维平面上两个二维数据对应位置上的数据相乘后的总和而成。假设以及分别表示为第 j 层和 j-1 层所对应的神经元的输出值。y^j 可用公式表示为：

$$y^j = f(wx^{j-1} + b) \tag{1}$$

其中，W 为权重值，b 为偏置值。由公式(1)可以看出，前一层卷积的输出值为下一层的输入值。

在卷积神经网络的结构中，卷积的层次越深，网络的学习能力就越强特征图得到信息就越全。但是，随着网络层次结构的加深网络的计算量将会随之增加，也就导致网络变得更复杂，这样很容易会出现过拟合的现象。在一般的 CNN 结构中，提取的特征都是逐级递进的，由简单的的颜色、边缘特征逐渐变为复杂的纹理特征，最后的网络结构将提取的关键特征，以能精确的辨别特征图的属性。

1.2.2 池化层

池化层也称为取样层，顾名思义，就是对每特征图进行下采样的作用。池化层一般跟在卷积层之后，也是由多个特征图组成。池化层在网络结构中有对特征图进行下采样的同时对特征图进行尺度缩小的作用，但是它的每一个特征面唯一对应于其上一层的一个特征面，不会改变特征面的个数。在搭建网络过程中，之所以会使用池化层是因为在网络结构中如果一直采用卷积操作，会使得整体网络中的计算量过大而使得整个网络变得极其复杂。

池化层旨在通过降低特征图的分辨率以获得具有空间不变的特性。池化层起到二次提取特征的作用，它的每个神经元对局部接受域进行池化操作。池化的方法有多种，通常用到的池化操作有最大池化[10]（max-pooling）即选取图像区域的最大值作为该区域池化后的值和平均池化（mean-pooling）即计算图像区域的平均值作为该区域池化后的值。如下图 1 所示为简单池化过程：

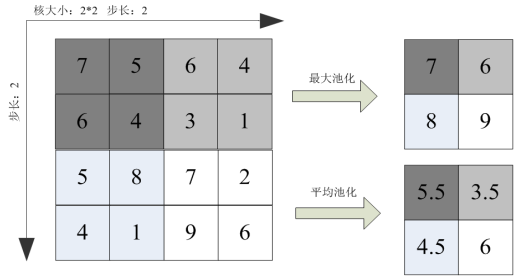


图 1　池化过程

Fig.1 Pooling

1.2.3 全连接层

在一个卷积神经网络中经过多次卷积层和池化层后会紧接着会跟一个或多个全连接层，这是因为由卷积层和池化层交替连接而成的卷积神经网络尽管提取了一部分的特征，与此同时也存在一些信息的丢失的情况，从而使得网络性能没有达到理想的要求。因此，加入全连接层后以实现收集更多的特征来满足特征提取的要求。

2　基于双网络特征级联的卷积神经网络模型

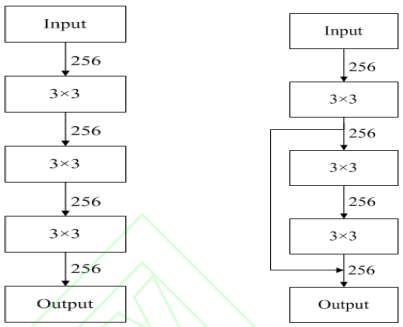
2.1 模型原理

传统的卷积神经网络是通过不同的卷积层和池化层级交替堆叠而成的通道，通过对卷积核大小及不同池化的设计来实现所需要的效果。而在一系列的操作过程中，一条通道、一种网络对特征图进行卷积，可能提取的图像特征不够充分。本文提出结合两条不同网络对相同的特征图进行卷积操作，最后再通过级联操作使两条网络的的输出结果结合在一起。本文的具体操作是将降维后的的特征图并行分为两条支路，一条采用 3*3 的卷积核进行传统的卷积操作；另一条在前一条的基础上采用残差操作；在下一次图片进行降维前通过级联操作将两条支路上的特征图结合在一起，形成一个新的特征图。

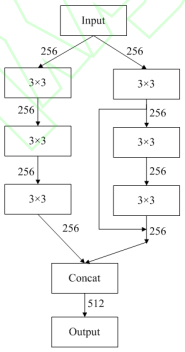
残差网络结构的提出主要是为了解决在原始的 CNN 随着深度的增加而伴随的网络退化的问题。主要原理是在初始卷积层结构的外部使用一个短接（shourtcut）操作构成一个一个基本的残差模块，通过逐级累加残差模块可以成功的缓解网络随深度而增加的退化问题，从而提升整体网络的性能。

2.2 模型结构

图 2 中（a）表示支路一卷积网络图，（b）表示支路二卷积网络图，（c）表示双网级联卷积结构图。



（a）支路一卷积网络结构　　（b）支路二卷积网络结构



（c）双网级联卷积结构

图 2　各网络结构图

Fig.2　Network structure diagram

可以很清楚的看出，本文所采用的双网络融合是在上一次降维之前将支路一网络即传统的神经网络和支路二网络即含有残差模块的卷积神经网络通过级联（Concat）操作将两个特征整合在一起，然后整体输出。

$$P = I \times K \times O \tag{2}$$

模块间的参数可以用公式 2 表示，其中 I 和 O 分别表示输入输出特征图的个数，K 表示卷积核的面积大小。如上图 2 所示，模块间的参数为 2×3×256×3×3×256＝3538944。在单个模块中双网络级联结构的参数近似等于两条支路参数的总和。

2.3 模型设置

双网络级联结构的整体是由多个模块组成，每个模块都是有两条支路图 2(a)、图 2(b)级联而成。支路一采用的是传统的 CNN 结构，即在采用平均池化降维之前一直采用 3*3 的卷积核进行卷积；支

路二采用的操作是在传统 CNN 结构的基础上加上残差操作。双网络级联结构的各层级设置如表 1 所示：

表 1 双网络级联各层级参数配置

Table 1 Dual-layer cascaded configuration parameters

各层结构		卷积核大小/步长/pad	输出维度
Input		---	227*227*3
Conv0/BN		11*11/4/0	55*55*128
AvePool0		2*2/2/1	28*28*128
并行分流	Conv1/BN（1–3 层）	3*3/1/1	28*28*128
	Conv2/BN（1–3 层）带有残差	3*3/1/1	28*28*128
级联（concat）		---	28*28*256
AvePool1		2*2/2/1	15*15*256
并行分流	Conv1/BN（4–6 层）	3*3/1/1	15*15*256
	Conv2/BN（4–6 层）带有残差	3*3/1/1	15*15*256
级联（concat）		---	15*15*512
AvePool2		2*2/2/1	8*8*512
并行卷积	Conv1/BN（7–9 层）	3*3/1/1	8*8*512
	Conv2/BN（7–9 层）带有残差	3*3/1/1	8*8*512
级联（concat）		---	8*8*1024
平均池化 Pool3		2*2/2/1	5*5*1024
全连接		---	101/256

本文所有网路结构中卷积层后均接入一层 BN(Batch Normalization)[11]层，即批量归一化层，加入 BN 层的主要优点有以下几点：第一，网络在选择较大学习率时可以不受梯度弥散[12]的影响，可以直接加快网络的收敛速度；第二，一定程度上有防止过拟合的作用，网络可以减少对 Dropout[13]参数的需求；第三，很完美的取代局部响应归一化层；第四，可以彻底打乱训练数据。但是在加入 BN 层后会增加网络的计算量，使网络的训练时间加长。BN 层的计算公式可由公式（3）给出：

$$\begin{cases} \mu_B \longleftarrow \dfrac{1}{m}\sum_{o=1}^{m} x_o \\ \sigma_B^2 \longleftarrow \dfrac{1}{m}\sum_{o=1}^{m}(x_o-\mu_B)^2 \\ \hat{x}_o \longleftarrow \dfrac{x_o-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}} \\ y_o \longleftarrow \gamma\hat{x}_o+\beta \end{cases} \tag{3}$$

公式（3）中，$\sum_{o=1}^{m} x_o$ 表示为卷积层的输出也即是所学归一化的数据，μ_b 和 σ_B^2 分别表示均值和方差，y_0 是归一化之后的输出。

3 实验

3.1 实验设置

数据集：实验数据所采用的数据集是公开数据集 101_food 和 caltech256。101_food 数据集包括 101 类食品总共含有 101000 张图片，每类食品含有 1000 张。实验中训练图片和测试图片的比例为 3:1，有 75750 张训练图片和 25250 张测试图片；caltech256 数据集中包含了 257 个类别的图片，但是其中有一类是背景类，在本实验的操 作中，去除了背景类，按照训练集和数据集以 4:1 的比例随机分开，最终得到 23919 张训练图片和 5862 张测试图片。目前所做的图像识别方面的研究中，这两个数据集在公开的数据集中使用相当广泛的。

准备操作：在本实验中，先对我们的数据集做了一系列的预处理，主要包括尺度归一化、去均值以及图像的裁剪与扩增。具体操作体现为，先把两个数据集中的图片大小都定义成 256*256，并在训练之前去均值，最后将图片按照 Alexnet 网络的的剪裁尺寸将所有的图片按左上角、右上角、左下角、右下角以及中间，随机剪裁成尺度为 227*227 大小，并在此基础上做水平翻转。

实验主要对比第一条支路网络即传统卷积网络、第二条网络即采用残差结构的传统网络以及我改进后的双网级联网络，在两个数据集上的性能优劣。为了使本实验结果更具有说服力，三种网络的整体结构的层次深度保持一致。

实验环境：整个实验的过程所有网络结构都是基于 caffe 框架[14]布置的的。本实验所用的计算机配置为 i7-6700K 四核 CPU、Ubuntu14.04 操作系统、32GB 内存以及 NVIDIA-GTX 1070 的 GPU。

参数设置：在卷积神经网络中，学习率大小的选取对网络训练至关重要，学习率较大，网络虽然收敛较快，但有可能跨过了全局最小点；但是学习率较小的情况下，网络训练速度比较慢，从而需要较长时间才能达到收敛。下面列出本实验在两个数据集中所设定的学习率的参数以及变化值，此时设置的参数值有较好的效果；在 101_food 数据集上训练时，学习率大小设置为 0.005，学习率变化方式为 multistep，gamma 为 0.1，stepvalue 设置为 40000、80000、120000，最大迭代次数为 150000；在 caltech256 上训练时，学习率大小设置为 0.01，学习率变化方式为 multistep，其中 gamma 为 0.1，stepvalue 设置为 24000 和 48000，最大迭代次数为 60000。

3.2 实验结果及分析

表二、表三分别表示各支路以及级联网络在 101_food 和 caltech256 上的分类精度、实验所花时间以及实验存储 caffemodel 的大小；图 3 和图 4 分别给出了各模型在两个数据集上的准确率曲线。

表 2 不同网络模型在 101_food 上的性能

Table 2 Performance of different network models on 101_food

网络模型	准确率（%）	训练时间（min）	Caffemode 大小(MB)
TraCNN	66.31	293	36.7
ResCNN	66.51	302	36.7
MyCNN	68.02	562	85

表 3 不同网络模型在 caltech256 上的性能

Table 3 Performance of different network models on caltech256

网络模型	准确率（%）	训练时间（min）	Caffemode 大小(MB)
TraCNN	59.10	115	44.6
ResCNN	60.01	118	44.6
MyCNN	61.62	138	100.8

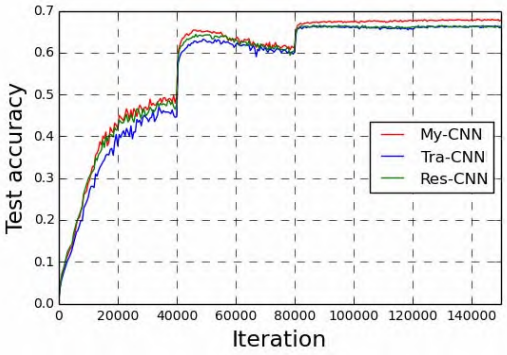


图 3 不同 CNN 模型在 101_food 上的准确率曲线

Fig.3 The accuracy curve of different CNN models on 101_food

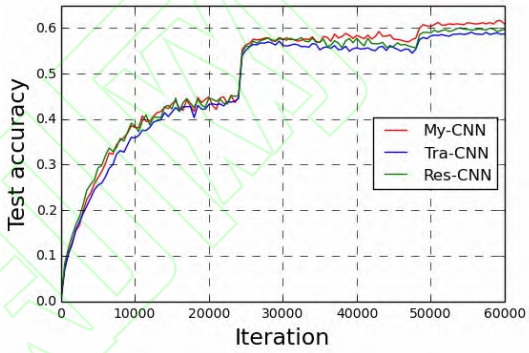


图 4 不同 CNN 模型在 caltech256 上的准确率曲线

Fig.4 The accuracy curve of different CNN models on caltech256

从表 2 和表 3 中可以看出，传统网络和残差网络在两个数据集中都有较好的分类精度。传统的支路一网络整体的参数较小，训练的时间也相对较少；加了残差结构的支路二整体参数与支路一的参数大致相同，在多加一步短接过增加了整个网络的计算量，导致训练所耗时间稍微加长，精度与传统 CNN 相比有所提高。而级联网络由于结合两条支路的特征所得出的分类精度比两条支路中的任意一条都要高，但是也因为网络加宽而参数加大（相当于两条支路参数相加关系），整个实验耗时也有所加长。随着计算机硬件的提升，网络改进带来的参数增加而导致训练时间加长的问题，很轻易得以解决，所以在准确率提高一定的前提下，训练时间的改变对整个网络增益没有大的影响。

4 结束语

本文提出了一种结合两种网络特征级联的新型结构，并将这个级联后网络与两条不同的支路网络进行对比讨论，以及在两个公开数据集 101_food 和 caltech256 上进行实验验证。理论上融合网络有加宽整体网络结构的宽度，提高了网络的复杂度，

效果会更好。实验结果也表明级联网络比两条支路网络中的任何一条网络的精确效果都好，有一点表现不足的是由于网络的加宽以及网络复杂度的加大，网络的参数几乎等于两条支路参数之和，间接导致网络在训练的时间上增加了很多。在接下来的工作中，有两个方向：第一，将两种网络拓展成三个网络或是多个网络以及更换不同种的网络结构，然后逐个进行对比，等到一个阈值，观察最终几条网络级联能得到最好效果；第二，在两条网络中实现同时提过精度和适当减少参数的设计，这样在更大规模数据集上也能在低性能计算机上运行。

参考文献

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[2] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv:1409.1556, 2014.

[3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015:1-9.

[4] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.

[5] Gao Huang, Zhang Liu, Laurens van der Maaten. Densely Connected Convolutional Networks[J].arXiv:1608.06993v4, 2017

[6] David E.Rumelhart, Geoffrey Hinton, Ronald J.Williams. Learning representations by back-propagating errors. Nature[J],1986,323(6088):533-536

[7] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[8] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504.

[9] Gao Li-Gang, Chen Pai-Yu, Yu Shi-Meng. Demonstration of convolution kernel operation on resistive cross-point array. IEEE Electron Device Letters[J], 2016,37(7):870-873

[10] Matthew D. Zeiler, Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks[J]. arXiv:1301.3557v1,2013

[11] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J].arXiv:1502.03167, 2015.

[12] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions[M]. World Scientific Publishing Co. Inc. 1998.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research ,2014,15(1):1929-1958.

[14] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]// ACM International Conference on Multimedia. ACM, 2014:675-678.