

基于并行和切片的深度卷积网络设计研究

杨远飞,曾上游,冯燕燕,周 悦,潘 兵
(广西师范大学 电子工程学院,广西 桂林 541004)

摘 要: 提出了基于并行和切片的卷积神经网络模型,并行网络通过两条通道分别对上一层特征图进行特征提取,但是同时网络参数也会倍增;切片网络是把上层输出特征图先切成两份,再分别对每份特征图进行卷积操作,最后再把两部分的卷积输出进行融合. 和同规模的模型相比,这两种模型能够提取更为本质的图像特征,但是切片网络不会增加网络的参数. 在实验阶段用这两种网络在 cifar10 和 cifar100 数据集上进行测试,取得了良好的效果.

关键词: 卷积神经网络;分类精度;特征图;网络参数

中图分类号: TP391. 4

文献标识码: A

文章编号: 1000—7180(2018)03—0120—05

Design and Research of Deep Convolution Network Based on Paralleland Slice

YANG Yuan-fei , ZENG Shang-you , FENG Yan-yan , ZHOU Yue, PAN Bing

(College of Electronic Engineering, Guangxi Normal University, Guilin 541004, China)

Abstract: We have proposed the convolution neural network model based on parallel and slice, network of parallel extract feature of the top feature map is carried out through two channels ,however, the parameters of the network will be doubled;network of slice cut the top output maps into two parts at first, convolute each map, and then fuse the convolution output of the two parts finally. Compared with the same scale model, those two models are able to extract more essential image features , but network of slice will not be added the network parameters. Those networks are tested on the cifar10 and cifar100 datasets in the experimental stage, and good results are obtained.

Key words: convolution neural network; classification accuracy; feature maps; network parameters

1 引言

目前,深度学习^[1]在图像处理、语音识别、视频检测和自然语言处理方面取得了巨大的进展^[2-3]. 卷积神经网络(CNN)是一种自带卷积结构的深度学习模型,它用局部感受野和权值共享的方式有效减少了网络的复杂度,使网络更容易训练,在形式上更加类似于生物神经网络. 卷积神经网络的主要优点在于可以直接将二维图像作为网络的输入,通过网络的多个隐含层逐层自动对图像信息提取特征并找到好的表征方法,避免了传统方法人工定义特征的复杂过程,并在很大程度上提高网络的分类精度.

卷积神经网络虽然在图像分类问题上具有许多

优势,但是仍存在一些问题. 网络对数据集的要求较高,若数据集较少,网络很可能会出现过拟合问题;传统卷积神经网络训练时采用 Sigmoid 激活函数可能出现梯度消失或梯度溢出^[4-5]等问题;网络层次较深时,网络虽然使用权值共享机制,但参数复杂度仍然很高,可能会导致网络难以收敛、训练时间过长或过拟合,因此,在防止网络过拟合的前提下如何减少网络参数复杂度并能有效提取图像的本质特征具有重要意义.

为了解决数据集较少的问题, Krizhevsky 等人设计了 AlexNet^[6] 网络,提出用数据扩增方法来增加数据集的数量,并在全连接层使用了 Dropout^[7] 技术,可以有效防止过拟合;为了减少网络参数复杂

度,Min Lin等提出的Network in Network^[8]首次提出用 1×1 的卷积核,并用全局池化的方式代替全连接层,除此之外,为了解决网络优化困难的问题,He K等提出了ResNet^[9]网络。

2 卷积神经网络的工作原理

卷积神经网络的训练可以分为两部分,前向计算与误差反向传播^[10]。前向计算是通过卷积层、池化层和全连接层的一系列变换得到特征向量,最后通过Softmax分类器进行分类;反向传播是用来修正误差,用的算法是梯度下降法,用误差来逐层更新网络的权值等参数。

卷积层的输出可由式(1)得到,其中 $\mathbf{x}_i^{l-1}$ 是上层的输出特征图, $\mathbf{M}_j$ 为上层特征图的子集, $\mathbf{k}_{i,j}^{l-1}$ 为卷积核的权值, $\mathbf{b}_j^l$ 为对应的偏置项, $f(\cdot)$ 为激活函数。

$$\mathbf{x}_j^l = f\left(\sum_{i \in \mathbf{M}_j} \mathbf{x}_i^{l-1} * \mathbf{k}_{i,j}^{l-1} + \mathbf{b}_j^l\right) \quad (1)$$

池化层的参数可由式(2)得到,其中 β_j^l 和 $\mathbf{b}_j^l$ 是权值和偏置参数, $\text{down}(\cdot)$ 和 $f(\cdot)$ 为下采样函数和激活函数。

$$\mathbf{x}_j^l = f(\beta_j^l \text{down}(\mathbf{x}_j^{l-1}) + \mathbf{b}_j^l) \quad (2)$$

全连接层的输出可由式(3)得到,其中 ω_i 和 $\mathbf{b}_i$ 是权值和偏置参数, $f(\cdot)$ 为激活函数。

$$\mathbf{x}_j^l = f(\omega_i \mathbf{x}_{i-1} + \mathbf{b}_i) \quad (3)$$

网络的平均误差代价函数可由式(4)得到,权值的更新可由式(5)得到,其中, N 为样本个数, C 为输入的种类, ∂ 为学习率。在实际操作中,为了防止过拟合,会在代价函数中加入 L_2 正则化项。

$$E = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^C (u_k^n - x_k^n)^2 \quad (4)$$

$$\omega_i = \omega_{i-1} + \partial \frac{\partial E}{\partial \omega_{i-1}} \quad (5)$$

网络通过多次前向运算与误差反向传播的不断交替来调整权重、偏置等参数,以此来极小化误差,直至网络达到收敛要求为止。

3 基于并行和切片的卷积神经网络模型

3.1 模型原理

传统的卷积神经网络是通过卷积核对上层输出进行卷积操作得到卷积层的输出特征图,但是卷积操作只有一条通道,可能提取的图像特征不充分;本文提出的并行卷积神经网络是通过两条通道分别对上层特征图进行卷积操作,最后再把两条通道的输出特征图融合在一起;本文提出的切片卷积网络是

把上层输出分成两部分,再用同等数量的卷积核分别对两部分输出进行卷积操作,最后把得到的两部分特征图进行融合。

式(6)和式(7)分别给出了并行卷积和切片卷积的卷积层输出,其中, $\mathbf{M}$ 为上层特征图的子集, $\mathbf{M}_1$ 为上层特征图切片后第一部分的子集, $f_1(\cdot)$ 为激活函数, $\mathbf{x}_{j,1}^l$ 为第一个通道的卷积输出, $f_2(\cdot)$ 为两个通道的特征图融合方式, $\mathbf{x}_j^l$ 为融合之后输出。

$$\begin{cases} \mathbf{x}_{j,1}^l = f_1\left(\sum_{i \in \mathbf{M}} \mathbf{x}_i^{l-1} * \mathbf{k}_{i,j}^{l-1} + \mathbf{b}_j^l\right) \\ \mathbf{x}_{j,2}^l = f_1\left(\sum_{i \in \mathbf{M}} \mathbf{x}_i^{l-1} * \mathbf{k}_{i,j}^{l-2} + \mathbf{b}_j^l\right) \\ \mathbf{x}_j^l = f_2(\mathbf{x}_{j,1}^l, \mathbf{x}_{j,2}^l) \end{cases} \quad (6)$$

$$\begin{cases} \mathbf{x}_{j,1}^l = f_1\left(\sum_{i \in \mathbf{M}_1} \mathbf{x}_i^{l-1} * \mathbf{k}_{i,j}^{l-1} + \mathbf{b}_j^l\right) \\ \mathbf{x}_{j,2}^l = f_1\left(\sum_{i \in \mathbf{M}_2} \mathbf{x}_i^{l-1} * \mathbf{k}_{i,j}^{l-2} + \mathbf{b}_j^l\right) \\ \mathbf{x}_j^l = f_2(\mathbf{x}_{j,1}^l, \mathbf{x}_{j,2}^l) \end{cases} \quad (7)$$

3.2 模型结构

图1~3分别给出了传统网络、并行网络和切片网络的卷积结构图。

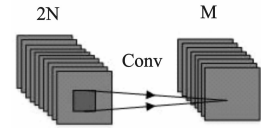


图1 传统网络卷积结构

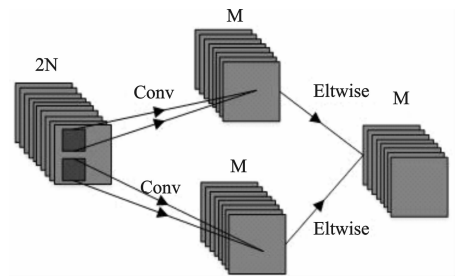


图2 并行网络卷积结构

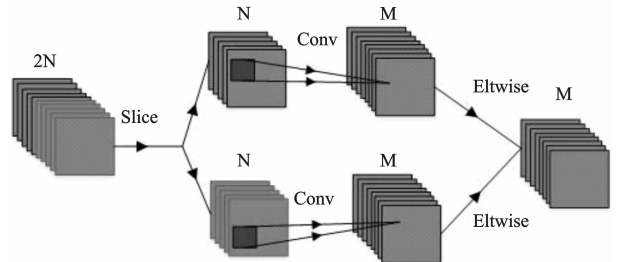


图3 切片网络卷积结构

网络中卷积层的参数个数可由式(8) 得出,其中, M_1 和 M_2 分别为该卷积层输入和输出的特征图数量, K 是卷积核的尺寸.

$$P = M_1 * K * M_2$$

(8)

为了简化计算,式(8) 未考虑偏置参数. 当输入输出的特征图数量相同并在假设卷积核尺寸相同的前提下,传统卷积网络、切片卷积网络和并行卷积网络的卷积层参数数量分别为 $2N * K * M$, $N * K * M * 2$ 和 $2N * K * M * 2$. 但在本文的实验部分,为了使并行网络和切片网络的两条通道提取不同的特征,每个通道使用大小不同的卷积核($3 * 3$ 和 $1 * 1$),而传统网络卷积层采用大小为 $3 * 3$ 和 $1 * 1$ 交替的卷积核. 和传统网络相比,切片网络不会增加卷积层参数,但是两条通道的特征图融合可能会得到更本质的特征;和并行网络相比,切片网络减少了一半的卷积层参数.

3.3 模型设置

并行网络和切片网络的整体结构都是通过多个卷积模块(图 2,图 3)级联组成的,两个通道在融合时采用相加的方式,并行网络的整体结构和具体参数设置如表 1 所示.

表 1 并行网络的整体结构和参数设置

操作类型	卷积核大小/个数/步长/pad	深度	参数数量
卷积	5 * 5/32/1/2	1	1 * 2 400
	3 * 3/32/1/1		6 * 9 216
并行卷积模块	1 * 1/32/1/0	6	6 * 1 024
	3 * 3/64/2/1		1 * 18 432
并行卷积模块(降维)	1 * 1/64/2/0	1	1 * 2 048
	3 * 3/64/1/1		4 * 36 864
并行卷积模块	1 * 1/64/1/0	4	4 * 4 096
	3 * 3/128/2/1		1 * 73 728
并行卷积模块(降维)	1 * 1/128/2/0	1	1 * 8 192
	3 * 3/128/1/1		4 * 147 456
并行卷积模块	1 * 1/128/1/0	4	4 * 16 384
	global pooling/AVE		无
池化		1	
全连接	10/100	1	1 280/12 800

网络中参数数量是依据式(8)计算得出的,除此之外,为了防止过拟合,网络中每个卷积层之后加入了批量归一化层(Batch-Normalization, BN^[11]),并用全局池化的方式减少全连接层的参数,式(9)给出了 BN 层的数据归一化过程:

$$\left\{ \begin{array}{l} \mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \\ \sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \\ x_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ y_i \leftarrow \gamma x_i + \beta \end{array} \right.$$

(9)

式中, $\sum_{i=1}^m x_i$ 是需要归一化的数据(卷积层的输出), μ_B 和 σ_B^2 是均值和方差, y_i 是归一化之后的输出. 加入 BN 层的好处是,逐层归一化,避免了梯度消失和梯度溢出,减少了网络对权值等初始化的依赖;能够加速网络收敛;能够作为一种正则化手段,减少了网络对于 Dropout 的需求.

最后,切片网络的整体框架设置也如表 1 所示,只是把表 1 中除降维之外的其他并行卷积模块(图 2)换成切片卷积模块(图 3),各个卷积层的卷积核个数,大小,步长以及 pad 保持不变,这样,切片网络和并行网络在整体设置和深度上保持相同,使实验结果更具有对比性.

4 实验

4.1 实验设置

实验所用的数据集是 cifar10 和 cifar100 的公开数据集. cifar10 数据集包含了 60 000 张 $32 * 32$ 的彩色图像,一共可分为 10 类,其中,有 50 000 个训练图像和 10 000 个测试图像. cifar100 数据集也包含了 60 000 张 $32 * 32$ 的彩色图像,分为 100 小类,20 大类. 每小类包括 500 个训练图像和 100 个测试图像. 相比之下, cifar100 数据集种类明显多于 cifar10 数据集,而且每个种类数量更少,在分类时更加困难.

实验主要对比传统卷积网络、并行卷积网络和切片卷积网络在两个数据集上的性能优劣. 为了使实验结果具有对比性,三种网络的整体结构和深度保持相同,也即传统网络的整体结构也如表 1 所示;只是把表 1 中的并行卷积模块换成卷积操作(图 2). 实际操作中,传统网络的卷积层采用 $3 * 3$ 和 $1 * 1$ 交替的卷积核,改变维度时,采用 $3 * 3$ 的卷积核($s=2, p=1$).

在对图像进行训练之前,对输入图像进行去均值操作. 网络在训练时采用数据扩增技术,对每幅图像分别从左上角、左下角、右上角、右下角和中间裁

剪 28×28 的像素大小,之后再进行水平翻转,这样可以把训练集扩展为原来的 10 倍,并增加一定的泛化能力;网络在测试时,则只对输入图像进行水平翻转. 实验中所有的网络结构都使用 caffe 框架^[12]来部署,计算机的配置为 i7-6700K 四核 CPU、Ubuntu14.04 操作系统、32 GB 内存以及 NVIDIA-GTX 1070 的 GPU.

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,但可能收敛不到全局最小点;若学习率较小,网络训练过慢,需要很长时间才能收敛. 以下给出了在两种数据集上较好的学习率大小以及学习率变化的方式. 在 cifar10 上训练时,学习率大小设置为 0.001,学习率变化方式为 poly,其中 gamma 为 0.9, power 为 1.5;在 cifar100 上训练时,学习率大小设置为 0.008,学习率变化方式为 multistep,其中 gamma 为 0.2, stepvalue 设置为 24 000 和 48 000.

4.2 实验结果及分析

表 2 和表 3 给出了各模型的参数数量(不包括偏置)及在 cifar10 和 cifar100 上的分类精度;图 4 和图 5 给出了各模型在两个数据集上的准确率曲线.

表 2 不同 CNN 模型在 cifar10 上的性能

网络模型	参数数量	准确率/ %	训练时间/ min
传统 CNN	536 160	87.62	77
并行 CNN	986 720	89.56	176
切片 CNN	546 400	89.29	168

表 3 不同 CNN 在 cifar100 上的性能

网络模型	参数数量	准确率/ %	训练时间/ min
传统 CNN	547 680	65.76	74
并行 CNN	998 240	67.56	169
切片 CNN	557 920	67.11	160

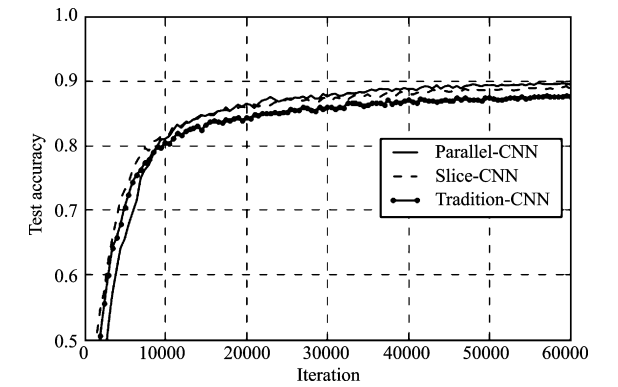


图 4 不同 CNN 模型在 cifar10 上的准确率曲线

从表 2 和表 3 中可以看出,并行网络和切片网络在 cifar10 和 cifar100 数据集上性能都很优异,它

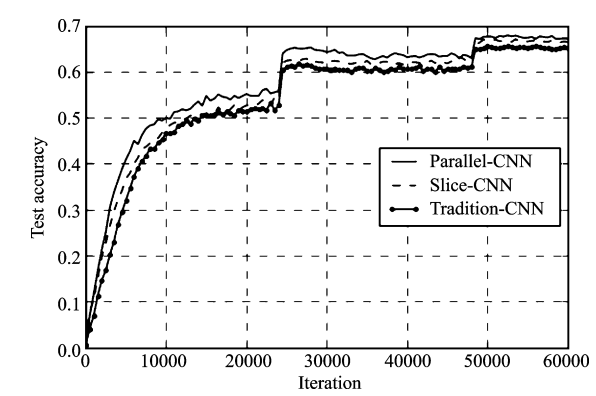


图 5 不同 CNN 模型在 cifar100 上的准确率曲线

们的分类精度均高于传统网络,但是并行网络参数数量最多并且训练时间最长;切片网络的分类精度略低于并行网络,但是其参数数量和并行网络相比几乎少了一半,网络的参数复杂度较低. 其训练时间比传统网络长是因为虽然通过切片模块不会增加参数数量,但是切片的每个通道在融合之前都要经过 BN 层计算,这样就比传统卷积操作多了一倍的 BN 层计算量,而且切片模块两条通道在融合时也会占用部分时间来计算融合后输出.

同时可以看出相同的网络在 cifar100 上训练的时间更短,从参数数量上来讲,相同的网络在 cifar100 上的参数多于 cifar10(全连接层神经元个数). 这种现象主要是因为两种数据集上选用的学习率变化方式不同. 学习率在 cifar100 上训练时采用的是 multistep,学习率每隔 24 000 次才会变化一次;而在 cifar10 上训练时,学习率采用的是 poly,每次迭代学习率都会发生变化.

我们可以得到以下结论:

- (1)并行网络和切片网络的多通道特征提取对整体识别率是有贡献的;
- (2)切片网络和并行网络相比,在保证精度的同时可以有效减少参数数量;
- (3)切片网络 and 传统网络相比,会增加训练时间复杂度,但不会增加网络参数复杂度,而且还能提高网络性能.

5 结束语

本文通过对传统卷积神经网络的分析提出了基于并行和切片的卷积神经网络模型,并对两种模型的结构进行了讨论,最后在 cifar10 和 cifar100 数据集上进行实验,结果表明并行网络在增加参数数量的前提下可以提高网络性能,而切片网络在具备少量参数的情况下性能同样优越.

参考文献:

[1] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504.

[2] Chen H, Qi X, Yu L, et al. DCAN: Deep contour-aware networks for object instance segmentation from histology images. [J]. Medical Image Analysis, 2017, 36:135-146.

[3] 张立民, 刘凯. 基于深度玻尔兹曼机的文本特征提取研究[J]. 微电子学与计算机, 2015, 32(2):142-147.

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

[5] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9:249-256.

[6] 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.

[7] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.

[8] Lin M, Chen Q, Yan S. Network In Network[C]//

ICLR. IEEE, 2014.

[9] 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.

[10] Rumelhart D E, Hinton G E, Williams R J. Learning internal representation by back-propagation of errors [J]. Nature, 1986, 323(323):533-536.

[11] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Computer Science, 2015.

[12] 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.

作者简介:

杨远飞 男, (1992-), 硕士. 研究方向为深度学习. E-mail: yyf287852761@163. com.

曾上游 男, (1974-), 博士, 教授. 研究方向为非线性动力学、计算神经科学.

冯燕燕 女, (1993-), 硕士. 研究方向为深度学习.

周悦 男, (1993-), 硕士. 研究方向为深度学习.

潘兵 男, (1993-), 硕士. 研究方向为深度学习.

(上接第 119 页)

5 结束语

本文提出基于 FFD 的大规模高维数据集中局部异常数据挖掘算法, 将 FFD 应用于进行数据挖掘是一次大胆尝试, 是对 RFD 应用弊端的思考与改进. 本文在进行 FFD 挖掘算法设计时, 着重考虑了挖掘效率与挖掘任务完成比率的均衡挖掘性能, 达成了以往所提方法无法比拟的挖掘性能, 并轻松通过了传输、数据本地化以及挖掘三项功能测试, 表明本文方法具有很强的可用性, 达到了期望效果, 而且具有挖掘效率高、挖掘任务完成量大优点.

参考文献:

[1] 李永红, 张继福, 苟亚玲. 相关子空间中的局部离群数据挖掘算法研究[J]. 小型微型计算机系统, 2015, 36(3):460-465.

[2] 农晓锋. 高维数据集中局部离散文本数据挖掘方法研究[J]. 现代电子技术, 2017, 40(19):89-105.

[3] 董本清, 彭健钧. 复杂网络数据流中的异常数据挖掘算法仿真[J]. 计算机仿真, 2016, 33(1):434-437.

[4] 焦润海, 张谦, 陈超. 基于 Spark 改进的最大频繁项集挖掘算法[J]. 计算机工程与设计, 2017, 38(7):

1839-1843.

[5] 梅毅, 熊婷, 罗少彬. 复杂属性环境下 NoSQL 分布式大数据挖掘方法研究[J]. 科学技术与工程, 2017, 17(9):239-243.

[6] 张继福, 李永红, 秦啸, 等. 基于 MapReduce 与相关子空间的局部离群数据挖掘算法[J]. 软件学报, 2015, 26(5):1079-1095.

[7] 陈宏. 基于关联规则挖掘算法的用电负荷能效研究[J]. 电子设计工程, 2017, 25(4):79-82, 86.

[8] 刘继华, 强彦. 云计算下的一种数据挖掘算法的研究[J]. 科技通报, 2016, 32(12):133-137.

[9] 杨晓波. 基于投影模式支持集的数据挖掘算法研究[J]. 计算机应用与软件, 2017, 34(7):273-276.

[10] 刘晓丹. 分布式环境中保护隐私数据挖掘算法研究[J]. 自动化与仪器仪表, 2016, 32(10):155-156.

作者简介:

张凯斐 男, (1981-), 硕士研究生, 讲师. 研究方向为数据挖掘等. E-mail: zhangkaifei035@163. com.

刘继华 女, (1975-), 硕士研究生, 副教授. 研究方向为软件工程、云计算.

张菊芳 女, (1977-), 硕士研究生, 副教授. 研究方向为非线性泛函分析.