

# 基于二分支卷积单元的深度卷积神经网络

侯聪聪<sup>1</sup>, 何宇清<sup>1</sup>, 姜晓恒<sup>1</sup>, 潘静<sup>2</sup>

<sup>1</sup>天津大学电气自动化与信息工程学院, 天津 300072;

<sup>2</sup>天津职业技术师范大学电子工程学院, 天津 300222

**摘要** 深度卷积神经网络在图像分类任务中取得了极大的成功。现有的基于简化卷积的卷积神经网络结构能够减少网络参数,但会丢失部分特征信息,降低网络性能。为了提高图像分类正确率,提出一种二分支卷积单元。该卷积单元包含两种类型的滤波器,分别用于提取包含特征通道内与通道间信息的特征。以此卷积单元代替传统的滤波器,构建深度卷积神经网络,称为 CTsNet。将该网络应用于图像分类任务,在 CIFAR10、CIFAR100 数据库上进行验证实验。结果表明,二分支卷积单元能够有效提取包含不同信息的特征,增加特征的多样性,减少信息损失,基于二分支卷积单元的 CTsNet 结构能有效提升图像分类性能。

**关键词** 图像处理; 图像分类; 卷积神经网络; 二分支卷积单元; 级联二分支网络

**中图分类号** TP391.4      **文献标识码** A

**doi:** 10.3788/LOP55.021005

## Deep Convolutional Neural Network Based on Two-Stream Convolutional Unit

Hou Congcong<sup>1</sup>, He Yuqing<sup>1</sup>, Jiang Xiaoheng<sup>1</sup>, Pan Jing<sup>2</sup>

<sup>1</sup> School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

<sup>2</sup> School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

**Abstract** Deep convolutional neural networks are widely used in the image classification. Current convolutional neural networks architectures based on the simplified convolution can reduce the number of network parameters, but it will lose some of the important information, which decreases the performance of the networks. The two-stream convolutional unit is proposed, in order to improve the accuracy of image classification. The two-stream convolutional unit contains two convolutional filters, which extracts the features containing the information in and across the channels, respectively. Based on the proposed two-stream convolutional unit, a deep convolutional neural network called CTsNet is constructed. Experiments of image classification are conducted on the databases of CIFAR10 and CIFAR100. The experimental results demonstrate that the proposed two-stream convolutional unit can extract features containing the information in and across the channels separately, increase the diversity in features and reduce the information loss. The CTsNet based on the two-stream convolutional unit can improve the recognition performance effectively.

**Key words** image processing; image classification; convolutional neural networks; two-stream convolutional unit; cascaded two-stream network

**OCIS codes** 100.2960; 100.4996; 100.5010

## 1 引言

2006年,加拿大多伦多大学教授 Hinton 和他的学生 Salakhutdinov 在《Science》上发表的文章

中,第一次提出了深度学习的思想<sup>[1]</sup>。自此,越来越多的研究人员开始关注深度学习。近几年来,基于卷积神经网络思想的深度学习模型在不同计算机视觉任务中都取得了惊人的突破,例如图像分类<sup>[2]</sup>、物

**收稿日期:** 2017-08-02; **收到修改稿日期:** 2017-08-28

**基金项目:** 国家自然科学基金(61472274)

**作者简介:** 侯聪聪(1992—),女,硕士研究生,主要从事机器视觉、图像识别方面的研究。E-mail: houcc@tju.edu.cn

**导师简介:** 何宇清(1973—),男,博士,讲师,主要从事图像超分辨率重建方面的研究。E-mail: heyuqing@tju.edu.cn

体检测<sup>[3-4]</sup>、目标跟踪<sup>[5-6]</sup>、图像深度估计<sup>[7]</sup>等。

卷积神经网络(CNN)通常包含若干个卷积层、激活层和池化层。其中,卷积层主要用于提取多层次特征。通过不同的卷积层,可以学习到分级的特征<sup>[8]</sup>:浅层卷积层主要提取边缘、方向等信息;深层卷积层则用于提取深层次的语义结构特征。因此,卷积滤波器对于卷积神经网络的性能至关重要。

2012年,Krizhevsky等<sup>[9]</sup>提出的 AlexNet 主要包含5个卷积层和3个全连接层,包含6000万个卷积参数,是当时最大型的深度卷积神经网络。2014年,Simonyan等<sup>[10]</sup>应用小尺度卷积核,将神经网络深度扩展到19层,所构建的VGG模型在当时取得了较好的效果。2015年,Szegedy等<sup>[11]</sup>构建了具有22个卷积层的GoogLeNet模型,该模型包含了多个Inception结构块,每个结构块均包含多个不同尺度的卷积滤波器,能有效地提取多尺度信息,进一步提升了图像分类的性能。2016年,He等<sup>[12]</sup>提出ResNet网络结构,通过引入残差学习的思想,构建了152层的深度卷积神经网络,经在Imagenet图像分类数据库上验证,达到了当时较好的性能。

通过加深网络的深度和宽度,以上方法在一定程度上提升了网络的性能,但是这些方法都没有将特征的通道内信息和通道间信息分开处理,因而限制了特征的多样性。NIN(Network in network)<sup>[13]</sup>网络使用较浅的多层感知机代替传统的卷积层,即加入尺寸为1的卷积滤波器,将特征信息进行跨通道融合,使得网络的表达能力大大提高。文献[14]提出通过级联2个特殊的卷积层以代替传统的卷积层,其中,第一个卷积层使得输入特征与输出特征一一对应,第二个卷积层采用尺寸为1的滤波器。这

样做能够减少网络参数,降低计算复杂度,但是,这2个特殊的卷积层在提取特征信息的过程中会丢失输入特征的部分信息,影响网络性能。

为解决上述问题,本文提出一种二分支卷积单元,以代替传统的卷积滤波器进行特征提取。该卷积单元由提取通道内和通道间信息的卷积滤波器并列组成。通过学习以上2种卷积滤波器,能够提取2种特征,一种包含通道内信息,一种包含通道间信息,从而有效减少信息损失,增加特征多样性。相较于传统的卷积滤波器,所提出的二分支卷积单元能够有效地减少网络参数,降低计算复杂度。应用该二分支卷积单元,并结合池化、激活等运算,构成一种新的深度卷积神经网络,称为CTsNet。在CIFAR10和CIFAR100图像分类数据库上进行验证实验,并与NIN<sup>[13]</sup>、ResNet<sup>[12]</sup>等经典方法进行对比,实验结果表明,CTsNet取得了较好的分类效果。

## 2 包含通道内与通道间信息的特征提取

特征信息主要分为特征通道内信息和特征通道间信息。由于特征图上每个位置点的值是通过卷积或非线性变换等操作得到的响应值,因此将每个位置点称为特征响应点。特征通道内信息是输入特征的单个特征通道内,同一邻域的特征响应点之间以及不同邻域之间存在的联系;特征通道间信息是指输入特征的不同特征通道上,各个对应特征响应点之间存在的联系。图1描述了包含通道内信息的特征和包含通道间信息的特征的提取过程。

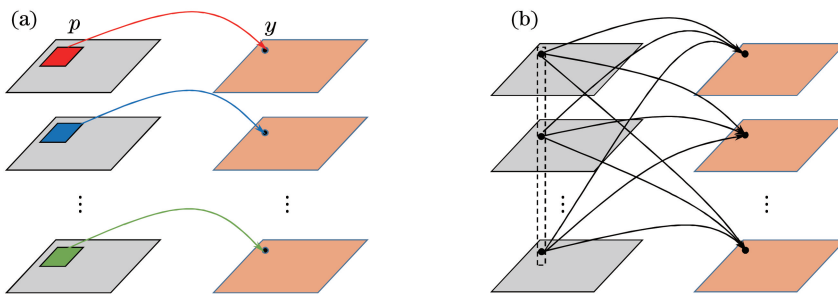


图1 包含特征通道内(a)与通道间(b)信息的特征提取过程

Fig. 1 Extraction process of features containing information in channels (a) and across the channels (b)

图1(a)描述了包含特征通道内信息的特征提取过程,在由 $M$ 个输入特征生成 $M$ 个输出特征的过程中,输出特征与输入特征一一对应。并且,输出特征中的某个响应点 $y$ ,只由相对应的输入特征中

某个邻域 $p$ 与滤波器卷积得到。图1(b)描述了包含特征通道间信息的特征提取过程,在由 $M$ 个输入特征生成 $N$ 个输出特征的过程中,输出特征的每个响应点由所有输入特征相同位置的响应点生成,而

与单特征通道内的其他响应点无关。

传统的卷积滤波器的完整形式可以用  $W \times H \times M \times N$  表示,其中  $W$  表示卷积核的宽, $H$  表示卷积核的高, $M$  表示卷积核输入的特征数, $N$  表示卷积核输出的特征通道数。根据上述特征通道内信息与特征通道间信息的提取方式,特征通道内信息可以用  $W \times H \times 1 \times M$  滤波器提取,特征通道间信息可以用  $1 \times 1 \times M \times N$  滤波器提取。

图 2 描述了文献[14]提出的简化卷积的过程:第一卷积层的卷积核用  $W \times H \times 1 \times M$  表示,只能提取输入特征的通道内信息,而在该过程中丢失了输入特征的通道间信息;第二卷积层的卷积核用  $1 \times 1 \times M \times N$  表示,只能提取通道间信息,而丢失了输入特征的通道内信息。由于在卷积过程中丢失了部分信息,因此,特征多样性受限,进而影响深度神经网络的性能。

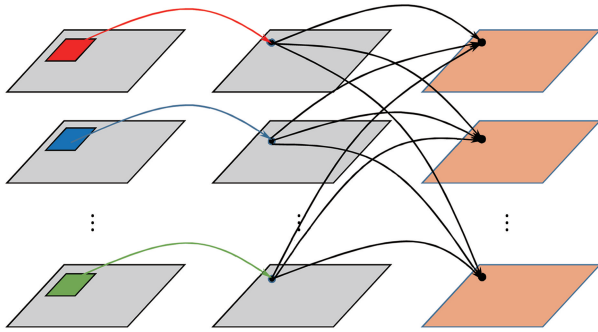


图 2 简化的卷积

Fig. 2 Simplified convolution

### 3 基于二分支卷积单元的神经网络

为解决上述问题,提出一种新的卷积单元——二分支卷积单元,用于同时提前包含特征通道内信息和特征通道间信息的特征。结合该卷积单元,构造深度卷积神经网络结构 CTsNet。

#### 3.1 二分支卷积单元

二分支卷积单元由两个滤波器并列组成,结构如图 3(a)所示:一个滤波器为  $W \times H \times 1 \times M$ ,用于提取输入特征的通道内信息;另一个滤波器为  $1 \times 1 \times M \times N$ ,用于提取输入特征的通道间信息。经过 2 种滤波器卷积后,将包含不同信息的输出特征进行合并,作为该卷积单元的最后输出特征。为了便于描述,采用的二分支卷积单元可用  $[(W \times H \times 1 \times M), (1 \times 1 \times M \times N)]$  表示。因此,提出的二分支卷积单元同时用滤波器  $W \times H \times 1 \times M$  和滤波器  $1 \times 1 \times M \times N$  对输入做卷

积运算。而图 3(b)描述的简化的卷积由滤波器  $W \times H \times 1 \times M$  和滤波器  $1 \times 1 \times M \times N$  串联组成,只用滤波器  $W \times H \times 1 \times M$  对输入做卷积运算。

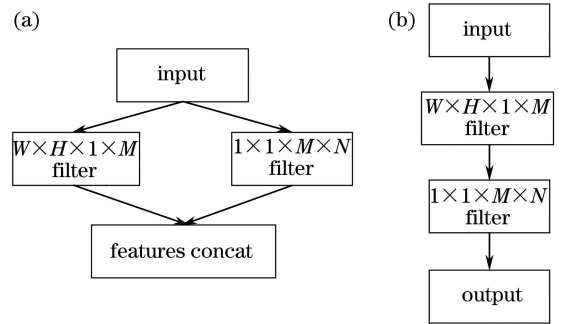


图 3 二分支卷积单元(a)与简化的卷积单元(b)对比

Fig. 3 Comparison of two-stream convolutional unit (a) and simplified convolutional unit (b)

二分支卷积单元中采用的特征合并过程如图 4 所示。具体地,假设输入特征维度为  $X \times Y \times M$ ,其中  $X$  表示输入特征的宽, $Y$  表示输入特征的高,采用二分支卷积单元  $[(W \times H \times 1 \times M), (1 \times 1 \times M \times N)]$ ,提取特征通道内信息的分支输出特征维度为  $X \times Y \times M$ ,提取特征通道间信息的分支输出特征维度为  $X \times Y \times N$ 。将两个分支的输出特征在特征通道维度上进行合并,经过特征合并后,该二分支卷积单元最终输出特征维度为  $X \times Y \times (M + N)$ 。

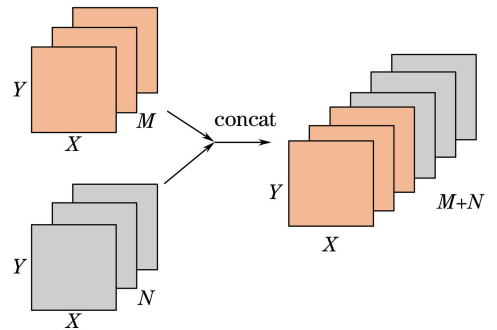


图 4 特征合并

Fig. 4 Features concat

为了更好地阐述二分支卷积单元的思想,用图 5 所示的例子来描述低维情况下的二分支卷积单元和简化的卷积。图中矩形框代表特征图,矩形框中的点代表响应点,空心点代表未画出的响应点。从图 5(a)可以看出,二分支卷积单元能够从输入中同时得到红色和蓝色两种特征,红色代表通道间信息,蓝色代表通道内信息。红色和蓝色特征合并得到输出特征,这样输出特征中包含了两种特征。图 5(b)中,简化的卷积只能从输入中直接得到蓝色特征。因此,二分支卷积单元能够获取更多输入特征的信息,增加特征的多样性,减少信息损失。

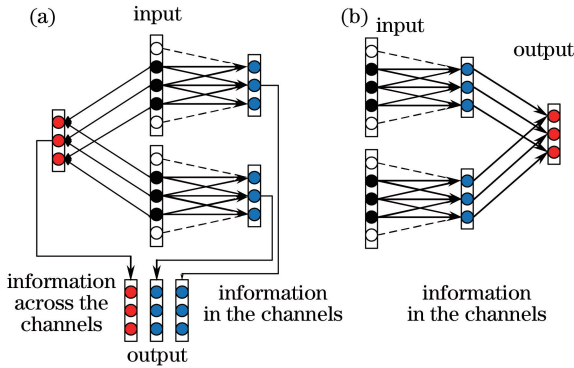


图5 二分支卷积单元(a)与简化的卷积单元(b)示意图

Fig. 5 Diagram of two-stream convolutional unit

(a) and simplified convolutional unit (b)

### 3.2 卷积单元间的连接

如上所述,假设第一个二分支卷积单元的输出特征维度为  $X \times Y \times (M+N)$ ,其输出特征通道数为  $(M+N)$ 。对于下一个二分支卷积单元,卷积滤波器尺寸应为  $[(W \times H \times 1 \times (M+N))]$ ,  $[1 \times 1 \times (M+N) \times N]$ ,其输出特征通道数为  $(M+2N)$ 。随着特征通道数增加,参数量和计算复杂度增加。

为避免参数量过度增长,在两个卷积单元之间,采用  $1 \times 1 \times (M+N) \times M$  的滤波器将其连接。经过该卷积滤波器,输出的特征维度为  $X \times Y \times M$ ,特征图通道数由  $(M+N)$  降低为  $M$ 。由此,二分支卷积单元的输出特征得到进一步融合,进而降低输出特征图维度,减少后续卷积单元的参数量,提升计算效率。采用图6所示的方式进行卷积单元间的连接。

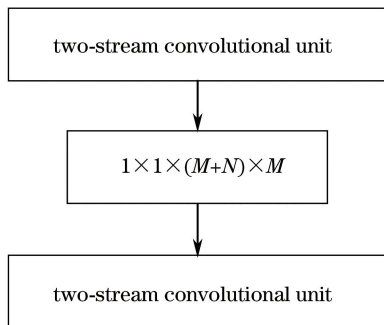


图6 卷积单元间的连接

Fig. 6 Connection between convolutional units

### 3.3 级联二分支网络

基于提出的二分支卷积单元,设计一种新的卷积神经网络结构,称为级联二分支网络(CTsNet)。该网络由2个卷积块组成,每个卷积块均由1个传统的卷积层和3个二分支卷积单元  $[(W \times H \times 1 \times M)]$ ,

$(1 \times 1 \times M \times N)]$  级联组成,每个卷积单元之间由采用  $1 \times 1 \times (M+N) \times M$  卷积滤波器的卷积融合层进行连接。为减少网络参数,降低计算复杂度并应用于CIFAR数据库,将用于提取通道内信息的滤波器  $W \times H \times 1 \times M$  设置为  $3 \times 3 \times 1 \times 192$ ,将用于提取通道间信息的滤波器  $1 \times 1 \times M \times N$  设置为  $1 \times 1 \times 192 \times 192$ ,即采用  $[(3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192)]$  的二分支卷积单元提取特征,采用  $1 \times 1 \times (192 \times 2) \times 192$  的卷积滤波器进行卷积单元连接,实现特征融合和特征降维。所实现的CTsNet网络结构配置实例,如图7所示。

为了加快训练时的收敛速度,在每个滤波器后加上批归一化(BN)<sup>[16]</sup>层;同时为了防止梯度消失,在BN层后添加线性矫正单元(ReLU)<sup>[9]</sup>。为保持网络的平移不变性,同时降低特征图维度,在每个卷积块之后分别引入最大池化层,以保证网络端对端训练的顺利进行。同时,CTsNet引入了引入了随机丢弃层(dropout)防止过拟合的产生——dropout加上更合适<sup>[17]</sup>防止过拟合的产生,dropout的比率为0.5。在网络的最后加上全局平均池化层,用来代替全连接层,减少网络参数。因为数据库有10类图片需要分类,所以网络的最后一个滤波器输出特征数为10,当数据库有100类图片时,需要将输出特征数改为100。

在CTsNet结构中,利用二分支卷积单元,把2种滤波器放在一层,以减少参数,并同时提取输入特征的通道内信息和通道间信息,再由下一层的滤波器进行融合,以有效地处理信息,增加特征的多样性,提升网络性能。

## 4 实验及结果分析

为了验证提出的CTsNet网络结构的有效性,在标准的数据库CIFAR10<sup>[15]</sup>及CIFAR100<sup>[15]</sup>上进行验证实验,用分类错误率作为评价网络性能的标准,并与其他几个经典网络结构进行对比,包括Maxout<sup>[18]</sup>、ALL-CNN<sup>[19]</sup>、NIN<sup>[13]</sup>、CSNet<sup>[20]</sup>、ResNet<sup>[12]</sup>、FitNet<sup>[21]</sup>等。

### 4.1 实验基础

CIFAR10数据集包含10类的彩色图像,每类有6000张,共60000张,其中包含50000张训练图像、10000张测试图像。该数据集的10类物体包括飞机、汽车、船、卡车、马、鸟、狗、猫、鹿、青蛙。实验中,训练阶段分别在有数据增强和无数据增强2种情况下进行,数据增强通过在每张图像的每边填充4 pixel后进行随机的裁剪和翻转来实现,并把经过

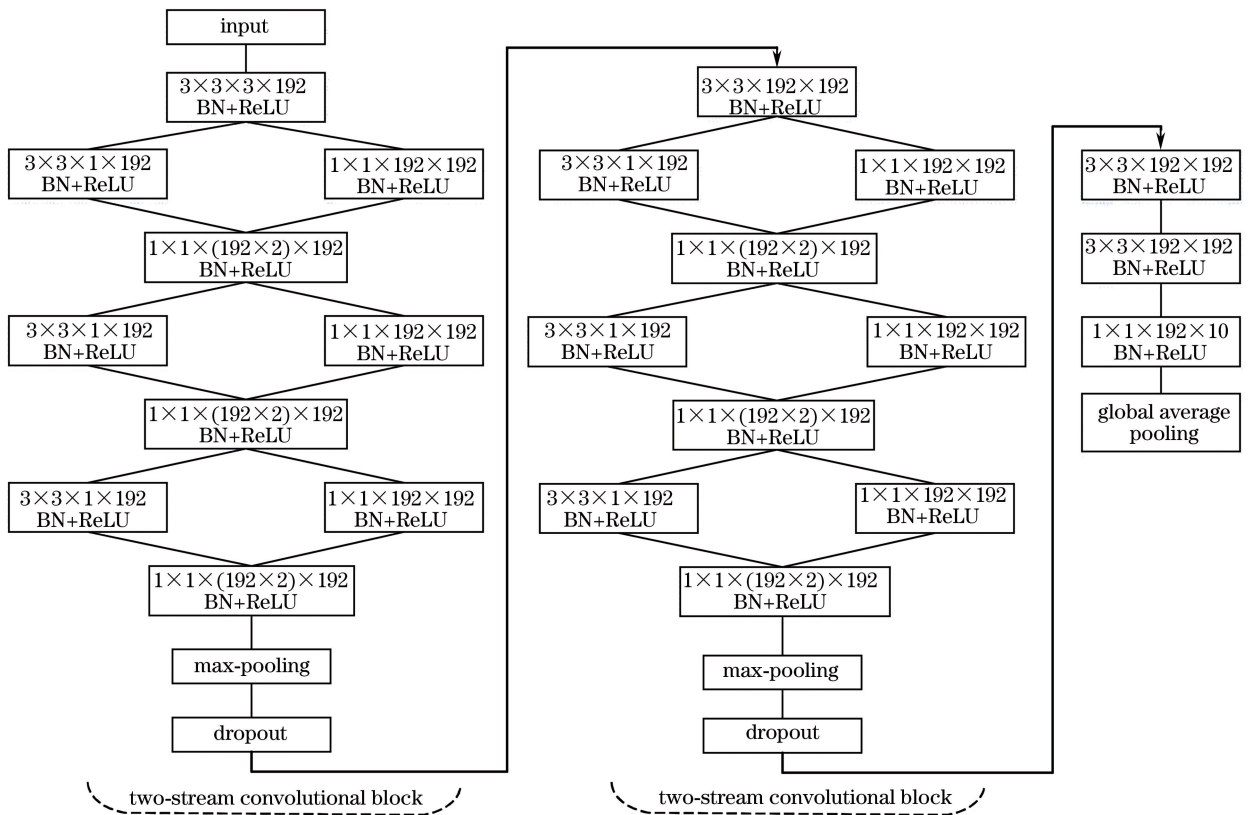


图7 应用于 CIFAR 数据库的 CTsNet 结构

Fig. 7 Architecture of CTsNet for CIFAR database

数据增强的 CIFAR10 数据集称为 CIFAR10+。在实验的测试阶段,只测试无增强的图像。

CIFAR100 数据集与 CIFAR10 数据库相似,包含 50000 张训练图像、10000 张测试图像。该数据库包含 100 类图像,因此每类有 500 张训练图像、100 张测试图像。在该数据库上不采用数据增强。

实验基于 MatConvNet<sup>[22]</sup> 深度学习工具箱进行改进实现。在本实验中,训练循环次数及学习率按照文献[2]中提出的方法设置。循环次数设置为 230 次,前 80 个循环的学习率为 0.5,第 81 到 180 个循环的学习率由 0.5 下降到 0.005,下降的步长为 0.005,第 181 到 230 个循环的学习率由 0.005 下降到 0.00005,下降的步长为 0.0001。

## 4.2 实验结果分析

### 4.2.1 二分支卷积与简化卷积实验对比

表 1 列出了 CTsNet 的参数配置,以及该网络结构中采用的所有滤波器。序号为 1、5、9、10、11 的滤波器是传统的滤波器,序号为 2、3、4、6、7、8 的单元中第一行是二分支卷积单元,第二行是用于二

支卷积单元连接的滤波器。作为对比,基于 CTsNet 的基本框架,实现了文献[14]中所提出的简化卷积的方法 LSNet,并列出该网络结构的参数配置。序号为 1、5、9、10、11 的滤波器是传统的滤波器,序号为 2、3、4、6、7、8 的单元中是简化的卷积滤波器。

采用上述 2 个网络结构,在 CIFAR10+ 数据库上进行实验。表 2 对比了 CTsNet 与本文实现的 LSNet<sup>[14]</sup> 的分类性能。CTsNet 网络结构具有 1.67 M 的参数,分类错误率为 6.33%;LSNet 网络结构具有 1.95 M 的参数,分类错误率为 6.86%。结果表明,CTsNet 在参数量少于 LSNet 参数量 0.28 M 的情况下,其分类性能较 LSNet 提高 0.53%。

由此可以看出,CTsNet 能在保持较少网络参数的情况下取得更好的效果。这是因为每个二分支卷积单元都能比简化的卷积从前一层获取更多的信息,较多的信息更利于图像分类,因此 CTsNet 取得了较好的分类效果。

表1 网络参数配置  
Table 1 Configurations of network parameters

Layer	LSNet <sup>[14]</sup>	CTsNet
1	$3 \times 3 \times 3 \times 192$	$3 \times 3 \times 3 \times 192$
2	$\begin{bmatrix} 3 \times 3 \times 1 \times 192 \\ 1 \times 1 \times 192 \times 192 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
3	$\begin{bmatrix} 3 \times 3 \times 1 \times 192 \\ 1 \times 1 \times 192 \times 192 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
4	$\begin{bmatrix} 3 \times 3 \times 1 \times 192 \\ 1 \times 1 \times 192 \times 192 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
5	$3 \times 3 \times 192 \times 256$	$3 \times 3 \times 192 \times 192$
6	$\begin{bmatrix} 3 \times 3 \times 1 \times 256 \\ 1 \times 1 \times 256 \times 256 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
7	$\begin{bmatrix} 3 \times 3 \times 1 \times 256 \\ 1 \times 1 \times 256 \times 256 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
8	$\begin{bmatrix} 3 \times 3 \times 1 \times 256 \\ 1 \times 1 \times 256 \times 256 \end{bmatrix}$	$\begin{bmatrix} (3 \times 3 \times 1 \times 192), (1 \times 1 \times 192 \times 192) \\ 1 \times 1 \times (192 \times 2) \times 192 \end{bmatrix}$
9	$3 \times 3 \times 256 \times 256$	$3 \times 3 \times 192 \times 192$
10	$3 \times 3 \times 256 \times 256$	$3 \times 3 \times 192 \times 192$
11	$1 \times 1 \times 256 \times 10$	$1 \times 1 \times 192 \times 10$

表2 CTsNet与LSNet的参数数量和分类错误率对比  
Table 2 Comparison on parameters and classification error rate of CTsNet and LSNet

Method	Error / %	Parameter / M
LSNet <sup>[14]</sup>	6.86	1.95
CTsNet	6.33	1.67

#### 4.2.2 CTsNet与经典方法实验对比

表3列出了CTsNet与其他几个经典网络结构在数据库CIFAR10、CIFAR100及增强后的CIFAR10+上的效果对比。CTsNet在CIFAR10上的分类测试错误率为7.95%，在CIFAR10+上的分类测试错误率为6.33%，与经典方法相比，

表3 不同网络在CIFAR10、CIFAR100上的分类错误率

Table 3 Classification error rate of CIFAR10 and CIFAR100 with various networks %

Method	CIFAR10	CIFAR10+	CIFAR100
Maxout <sup>[18]</sup>	11.68	9.38	38.57
NIN <sup>[13]</sup>	10.41	8.81	35.68
NIN+LA <sup>[23]</sup>	9.59	7.51	34.40
FitNet <sup>[21]</sup>	—	8.39	35.04
DSN <sup>[24]</sup>	9.75	8.22	34.57
Highway <sup>[25]</sup>	—	7.54	32.24
ALL-CNN <sup>[19]</sup>	9.08	7.25	33.71
RCNN-160 <sup>[26]</sup>	8.69	7.10	31.75
ResNet-110 <sup>[12]</sup>	—	6.43	—
CSNet-M <sup>[20]</sup>	8.15	6.38	30.24
CTsNet	7.95	6.33	30.12

CTsNet效果最好。CTsNet在CIFAR100上的分类测试错误率为30.12%，其分类性能超过NIN方法5.56%，超过Highway方法2.12%，超过RCNN-160方法1.53%。从表3可以看出，在有数据增强和无数数据增强的情况下，与经典方法对比，CTsNet均取得了最好的分类效果，充分证明了该方法的有效性。

## 5 结 论

提出了一种新的卷积单元——二分支卷积单元，该卷积单元由分别提取特征通道内信息与特征通道间信息的滤波器并列组成，代替传统的卷积层进行特征提取。以二分支卷积单元为基础，设计一种深度神经网络，称为CTsNet。该网络充分利用二分支卷积单元提取分别包含特征通道内信息与特征通道间信息的特征，能有效增加卷积层的特征多样性，减少信息损失，同时保持较少的网络参数量。在数据库CIFAR10及CIFAR100上的实验结果表明，该网络结构能够有效降低图像分类错误率，提升图像分类性能。

## 参 考 文 献

- [1] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [2] Pang Y, Sun M, Jiang X, et al. Convolution in convolution for network in network [J]. IEEE

- Transactions on Neural Networks and Learning Systems, 2017.
- [3] Cao J, Pang Y, Li X. Pedestrian detection inspired by appearance constancy and shape symmetry [J]. IEEE Transactions on Image Processing, 2015, 26(10): 2477-2486.
- [4] Ye G L, Sun S Y, Gao K J, *et al.* Nighttime pedestrian detection based on faster region convolution neural network [J]. Laser & Optoelectronics Progress, 2017, 54(8): 081003.  
叶国林, 孙韶媛, 高凯珺, 等. 基于加速区域卷积神经网络的夜间行人检测研究[J]. 激光与光电子学进展, 2017, 54(8): 081003.
- [5] Mao N, Yang D D, Yang F C, *et al.* Adaptive object tracking based on hierarchical convolution features [J]. Laser & Optoelectronics Progress, 2016, 53(12): 121502.  
毛宁, 杨德东, 杨福才, 等. 基于分层卷积特征的自适应目标跟踪[J]. 激光与光电子学进展, 2016, 53(12): 121502.
- [6] Gao L, Wang J F, Fan Y, *et al.* Robust visual tracking based on convolutional neural networks and conformal predictor[J]. Acta Optica Sinica, 2017, 37(8): 0815003.  
高琳, 王俊峰, 范勇, 等. 基于卷积神经网络与一致性预测器的稳健视觉跟踪[J]. 光学学报, 2017, 37(8): 0815003.
- [7] Xu L, Zhao H T, Sun S Y, *et al.* Monocular infrared image depth estimation based on deep convolutional neural networks [J]. Acta Optica Sinica, 2016, 36(7): 0715002.  
许路, 赵海涛, 孙韶媛, 等. 基于深层卷积神经网络的单目红外图像深度估计[J]. 光学学报, 2016, 36(7): 0715002.
- [8] Chang C H. Deep and shallow architecture of multilayer neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(10): 2477-2486.
- [9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of Advances in Neural Information Processing Systems(NIPS), 2012: 1106-1114.
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [C]. ICLR, 2015: 1-14.
- [11] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C]// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition(CVPR), 2015: 1-9.
- [12] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition[C]// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [13] Lin M, Chen Q, Yan S. Network in network [J]. arXiv preprint arXiv: 1312.4400, 2013.
- [14] Sun M, Song Z, Jiang X, *et al.* Learning pooling for convolutional neural network [J]. Neurocomputing, 2017, 224: 96-104.
- [15] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [D]. Toronto: University of Toronto, 2009.
- [16] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]. International Conference on Machine Learning, 2015: 448-456.
- [17] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [18] Goodfellow I J, Warde-Farley D, Mirza M, *et al.* Maxout networks [C]. International Conference on Machine Learning, 2013: 13719-1327.
- [19] Springenberg J T, Dosovitskiy A, Brox T, *et al.* Striving for simplicity: the all convolutional net[EB/OL]. (2014-12-21) [2017-08-02]. <https://arxiv.org/abs/1412.6806>.
- [20] Jiang X, Pang Y, Sun M, *et al.* Cascaded subpatch networks for effective CNNs [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 99: 1-10.
- [21] Romero A, Ballas N, Kahou S E, *et al.* Fitnets: hints for thin deep nets[C]. ICLR, 2015: 1-13.
- [22] Vedaldi A, Lenc K. MatConvNet: convolutional neural networks for matlab [C]// Proceedings of ACM Conference on Multimedia Conference, 2015: 689-692.
- [23] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [C]. Advances in Neural Information Processing Systems(NIPS), 2015: 2377-2385.
- [24] Liang M, Hu X. Recurrent convolutional neural network for object recognition[C]. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition(CVPR), 2015: 3367-3375.
- [25] Agostinelli F, Hoffman M, Sadowski P, *et al.* Learning activation functions to improve deep neural networks[C]. ICLR, 2015: 1-9.
- [26] Lee C Y, Xie S, Gallagher P, *et al.* Deeply-supervised nets [C]. Artificial Intelligence and Statistics, 2015: 562-570.