

基于轻量级卷积神经网络的生活垃圾图像分类

易才键, 陈俊, 王师玮

(福州大学物理与信息工程学院, 福建 福州 350108)
✉ycjfu1998@163.com; 56851@qq.com; shiweiwangsw@163.com



摘要: 为解决人工分拣生活垃圾过程中存在的分类效率低、劳动强度大等问题, 提出一种轻量级生活垃圾图像分类模型DG-MobileNetV2。该模型以轻量级网络MobileNetV2为基础, 通过引入双尺度深度卷积模块、通道混洗技术、调整宽度因子等手段, 达到压缩模型参数量的同时提高分类准确率的目的。实验表明, DG-MobileNetV2的参数量仅为0.403 M(文中1 M=1,000,000), 在华为技术有限公司发布的生活垃圾分类数据集上的分类准确率达到90.58%, 具有良好的分类效果。后续该轻量级模型可以移植至移动端或嵌入式设备中, 为生活垃圾智能化的分类提供了新的思路和实现方法。

关键词: 卷积神经网络; 轻量级; 垃圾图像分类; 多尺度卷积; 通道混洗
中图分类号: TP391.4 **文献标识码:** A

Domestic Garbage Image Classification based on Lightweight Convolutional Neural Network

YI Caijian, CHEN Jun, WANG Shiwei

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China)
✉ycjfu1998@163.com; 56851@qq.com; shiweiwangsw@163.com

Abstract: Aiming at the problems of poor classification efficiency and high labor intensity in the process of manual sorting of domestic garbage, this paper proposes a lightweight domestic garbage image classification model DG-MobileNetV2 based on MobileNetV2. By introducing dual-scale depthwise convolution module, channel shuffle technology and adjusting the width factor, the parameter numbers of the model is compressed and the classification accuracy is improved. The experiment results shows that the parameter numbers of DG-MobileNetV2 is only 0.403 M(In the paper, 1 M=1,000,000), and its classification accuracy rate on domestic garbage classification dataset released by Huawei Technologies Co., Ltd. reaches 90.58%, which shows a good classification effect. Subsequently, the lightweight model can be transplanted to mobile terminals or embedded devices, providing new ideas and approaches for intelligent domestic garbage classification.

Keywords: convolutional neural network; lightweight; garbage image classification; multi-scale convolution; channel shuffle

1 引言(Introduction)

随着经济的高速发展, 城市规模逐渐扩大, 人们的生活垃圾产量也逐年增加。面对庞大的垃圾处理需求和“脏乱差”的工作环境, 人工分拣生活垃圾的方式存在垃圾分类效

率低、工人劳动强度大等问题^[1]。针对以上问题, 研究应用深度学习技术辅助实现智能化垃圾分类, 具有重要的现实意义和实用价值。

近年来, 深度学习技术在很多领域表现出色, 尤其在

图像分类方面取得了巨大进展^[2]。随着越来越多的学者对卷积神经网络开展研究,许多性能优异的大型网络被提出,如VGG^[3]、GoogleNet^[4]、ResNet^[5]等。但是,这些大型网络普遍存在参数量过大、运行速度慢等问题,无法应用在对实时性有一定要求的场景中。为解决上述问题,一些学者提出轻量型网络,如SqueezeNet^[6]、MobileNet^[7]、ShuffleNet^[8]等,这类轻量型网络能够满足实际生活和工业生产场景的需求,适合用于生活垃圾图像的分类。

华为技术有限公司发布了一组符合我国生活垃圾分类标准的数据集(以下简称为华为垃圾分类数据集),进一步推动了该领域的发展。在华为垃圾分类数据集上,董子源等^[9]提出GCNet(Garbage Classification Network, 垃圾分类网络),引入注意力机制提升模型的特征提取能力,取得96.73%的分类准确率。徐传运等^[10]以深度残差网络ResNet18为基础,引入多分支和特征融合结构,分类准确率可达97.53%。现有的垃圾图像分类算法虽然性能较好,但是因参量数过大而难以部署在实际的生活垃圾分类场景中。

针对上述问题,本文对轻量型网络MobileNetV2进行改进,提出了一种轻量级且鲁棒性强的生活垃圾图像分类模型DG-MobileNetV2,该模型可准确、高效地对生活垃圾图像进行分类。

2 MobileNetV2网络简介(Introduction to MobileNetV2 network)

MobileNetV2网络^[11]使用深度可分离卷积(Depthwise Separable Convolution)代替标准卷积,大大减少了网络的参数量。深度可分离卷积由深度卷积(Depthwise Convolution, DW)和点卷积(Pointwise Convolution, PW)组成。深度卷积是指每一个卷积核负责一个通道的计算,只提取图像特征,不进行通道的合并;点卷积是卷积核大小为 1×1 的标准卷积,用于重新组合通道信息,构建新的特征。标准卷积和深度可分离卷积的过程对比如图1所示,图1中 D_k 是卷积核尺寸, M 和 N 分别是输入通道数和输出通道数。

在不考虑偏置参数的情况下,以 3×3 尺寸的卷积核为例,深度可分离卷积的计算量约标准卷积的 $1/9$ 倍,计算量大幅降低。

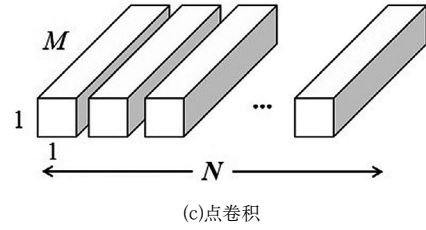
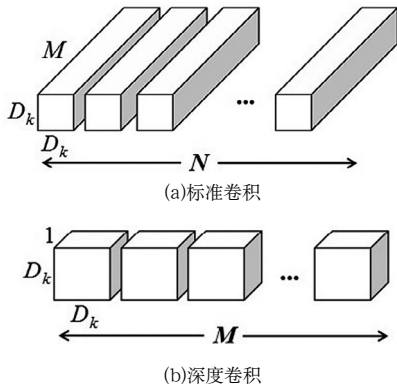


图1 标准卷积与深度可分离卷积对比

Fig.1 Comparison between standard convolution and depthwise separable convolution

除了保留深度可分离卷积操作,MobileNetV2在MobileNetV1的基础上还提出逆残差结构(Inverted Residual)和线性瓶颈结构(Linear Bottleneck)。线性瓶颈结构如图2所示,首先使用PW进行升维操作,其次经过DW提取特征,最后利用PW降低维度,图中 t 为维度缩放的倍数。线性瓶颈结构中的“先升维后降维”卷积操作即为逆残差结构,能够在DW前引入尽可能多的图像信息,更有利于特征提取。根据步长 $stride$ 的不同,线性瓶颈结构分为两种情况:步长 $stride=1$ 时,引入残差连接结构; $stride=2$ 时,没有残差连接结构。

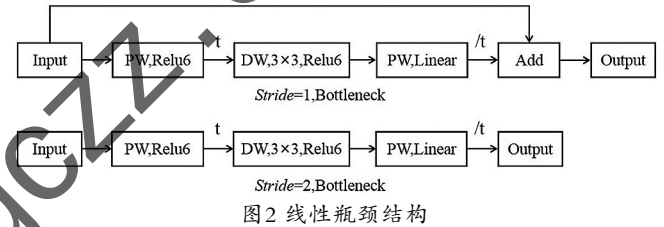


图2 线性瓶颈结构

Fig.2 Structure of linear bottleneck

3 生活垃圾图像分类模型(Domestic garbage image classification model)

3.1 双尺度深度卷积模块

在卷积神经网络中,特征提取主要依赖卷积操作,为提高模型的特征提取能力,本文提出双尺度深度卷积(Dual-Scale Depthwise Convolution, DSDC)模块。多尺度卷积的思想源自于GoogleNet的Inception结构,本文通过并联 3×3 和 5×5 两种尺度的深度卷积核替换原来的 3×3 卷积核。DSDC模块的结构如图3所示,输入特征图分别进行 3×3 和 5×5 的深度卷积,之后将计算后的特征图拼接起来作为整个模块的输出。

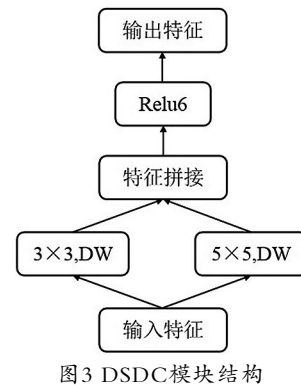


图3 DSDC模块结构

Fig.3 Dual-scale depthwise convolution structure

3.2 分组卷积与通道混洗技术

在深度可分离卷积过程中,深度卷积仅关注图像的空间信息,通道信息的融合依靠点卷积操作,然而点卷积操作往往会消耗大量的计算资源。因此,本文利用分组卷积(Group Convolution, GConv)结合通道混洗(Channel Shuffle, CS)技术代替网络中的部分点卷积操作,不仅压缩了模型的参数量,还保证了通道间信息的流通。

通道混洗技术以分组卷积为前提,首先将输入通道数 C 分为 G 组,令每个卷积核仅负责 C/G 个通道的卷积计算,然后将结果拼接得到输出特征,通过分组卷积能够将计算量压缩至标准卷积的 $1/G$ 。但是,分组卷积的弊端也很明显,不同组之间相互独立,通道之间的信息没有交流。通道混洗技术将分组卷积后的输出特征进行打乱与重组,在不增加任何参数量的前提下,使不同组的通道信息充分融合。图4(a)是普通的分组卷积操作,通道信息无法交流;图4(b)利用通道混洗技术保证了通道信息的交流与融合。

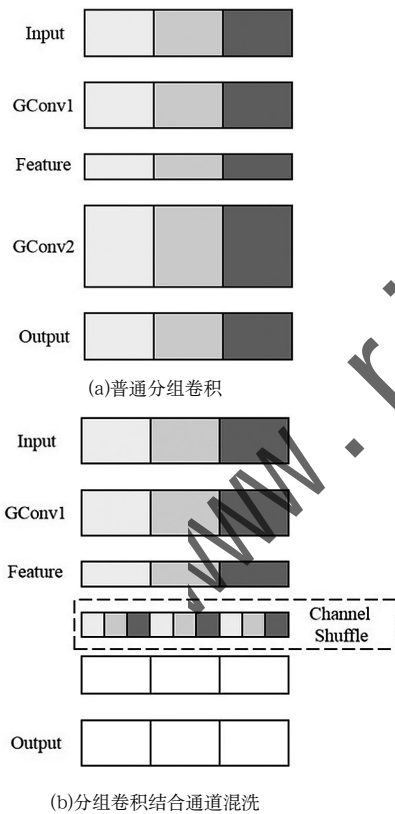


图4 分组卷积与通道混洗技术

Fig.4 Group convolution and channel shuffle technology

3.3 DG-MobileNetV2整体结构

DG-MobileNetV2以MobileNetV2为基础,构建新的线性瓶颈模块,调整网络宽度因子,达到压缩模型参数量、提高分类准确率的效果。改进后的线性瓶颈模块(New-Bottleneck)如图5所示,其中 t 为维度缩放的倍数。图像特征提取方面,利用DSDC模块替换原DW操作;通道特征融合方面,利用GConv结合CS技术替换负责降维的PW。

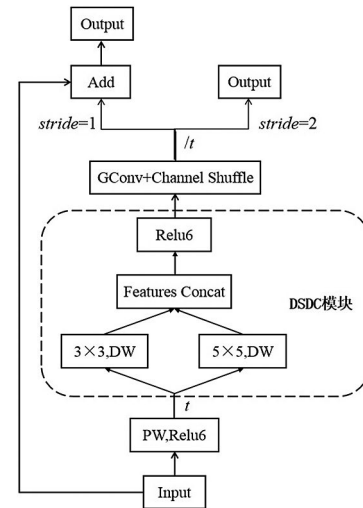


图5 改进后的线性瓶颈模块

Fig.5 Improved linear bottleneck

网络宽度因子 α 是MobileNetV2的一个超参数,例如 $\alpha=0.50$ 代表Bottleneck的输出通道数变为原来的一半。通过调整宽度因子,能够在牺牲一些分类准确率的前提下,再次压缩模型的参数量。在DG-MobileNetV2中, α 设置为0.50。

表1是DG-MobileNetV2网络的整体结构,其中 t 为New-Bottleneck中PW的升降维倍数; c 为输出通道数,此时 $\alpha=0.50$; n 为New-Bottleneck的重复次数; s 为步长。

表1 DG-MobileNetV2网络整体结构

Tab.1 Overall structure of DG-MobileNetV2 network

输入	操作	t	c	n	s
$224 \times 224 \times 3$	标准卷积	—	16	1	2
$112 \times 112 \times 16$	New-Bottleneck	1	8	1	1
$112 \times 112 \times 8$	New-Bottleneck	6	12	2	2
$56 \times 56 \times 12$	New-Bottleneck	6	16	3	2
$28 \times 28 \times 16$	New-Bottleneck	6	32	4	2
$14 \times 14 \times 32$	New-Bottleneck	6	48	3	1
$14 \times 14 \times 48$	New-Bottleneck	6	80	3	2
$7 \times 7 \times 80$	New-Bottleneck	6	160	1	1
$7 \times 7 \times 160$	PW	—	640	1	1
$7 \times 7 \times 640$	平均池化	—	—	1	—
$1 \times 1 \times 640$	PW	—	4	—	—

4 实验与结果分析(Experiments and result analysis)

4.1 实验环境

本文实验在Windows 10操作系统下完成,基于Pytorch深度学习框架。硬件环境为16 GB内存的Intel Core i7-9000K CPU处理器、11 GB显存的NVIDIA GeForce RTX 2080 Ti显卡。

4.2 数据来源与数据增强

本文采用华为垃圾分类数据集,该数据集集中含有40个小类、4个大类,共14,683张生活垃圾图像和对应的分类标

签。其中，4个大类按照我国的生活垃圾分类标准，分为可回收垃圾、厨余垃圾、有害垃圾和其他垃圾。在本文实验中，按照4:1的比例划分数据集，80%的数据作为训练集，20%的数据作为测试集。

为增强模型的泛化能力，防止训练过拟合现象，对训练集做数据增强操作，具体步骤如下：第一步，对每一张图像进行水平翻转、垂直翻转、逆时针旋转90°、逆时针旋转180°、逆时针旋转270°共5种操作，扩充数据集至原来的6倍；第二步，在训练过程中引入随机旋转、随机翻转和随机擦除操作。部分图像数据及数据增强后的效果如图6所示，其中图6(a)为原图，图6(b)—图6(f)为第一步数据增强的结果，图6(g)—图6(i)为第二步数据增强的结果。

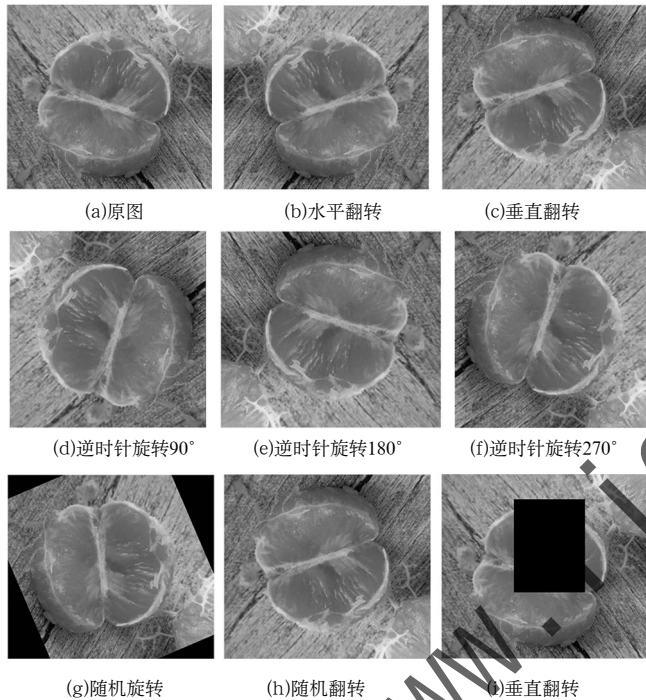


图6 原数据图像及数据增强效果

Fig.6 Original image and data enhancement effect

4.3 训练参数设置

本文实验使用Adam作为优化器，损失函数选择交叉熵函数。在数据增强后的数据集上进行实验，设置30个迭代周期，每批次训练32张图像。为保证模型的收敛，实验采用变学习率的训练方式，初始学习率设置为0.002，每经过10个迭代周期降低为原来的一半。除此之外，本文的部分实验采用迁移学习策略，将原始模型在公开数据集ImageNet的子集ImageNet100上进行训练，训练后的模型权重将用于相应模型的初始化。

4.4 结果与分析

在实验过程中，多尺度卷积模块在提升模型性能的同时必定会导致参数量的上升；调整模型的宽度因子，参数量和分类准确率会同时下降。为了寻找最优的多尺度组合和宽度因子，以MobileNetV2为基准模型，引入一个评价标准Value，Value的表达式如下：

$$Value = \frac{A_c - A_s}{P_c - P_s}$$

式中， A_c 和 A_s 分别是当前模型与基准模型的分类准确率， P_c 和 P_s 分别是当前模型和基准模型的参数量。

4.4.1 多尺度卷积核对比实验

对多组多尺度卷积模块进行对比实验，由于点卷积无法充分提取图像特征，过大的卷积核如 9×9 、 11×11 会导致模型参数量剧增，因此仅对 3×3 、 5×5 和 7×7 的卷积核进行组合。

对比实验结果如表2所示，由Value的定义可知，本实验应选择Value值最大的组合。(1)多尺度的深度卷积模块的性能优于单尺度，并且尺度越多，模型性能越好，这说明引入多尺度卷积能够提高模型的泛化能力；(2)相较于基准模型，多尺度模型的参数量有了不同程度的增加，权衡模型的参数量和分类准确率，选用Value值最大的 $\{3 \times 3, 5 \times 5\}$ 卷积组合；(3)本文选用的卷积组合 $\{3 \times 3, 5 \times 5\}$ 在仅增加0.107 M参数量的情况下，提高了1.33%的分类准确率。

表2 多尺度卷积核对比实验

Tab.2 Comparative experiment of multi-scale convolutional kernel

卷积核组合	参数量/M	准确率/%	Value
$\{3 \times 3\}$ (基准)	2.229	86.04	—
$\{3 \times 3, 7 \times 7\}$	2.371	87.11	7.54
$\{5 \times 5, 7 \times 7\}$	2.407	87.17	6.35
$\{3 \times 3, 5 \times 5, 7 \times 7\}$	2.435	87.51	7.16
$\{3 \times 3, 5 \times 5\}$ (本文)	2.336	87.37	12.43

4.4.2 宽度因子调整

为了验证本文宽度因子 α 的合理性，将本文方法($\alpha=0.50$)与 α 分别设置为0.25、0.35、0.75、1.00的模型进行对比实验。对比实验结果如表3所示，由Value的定义可知，本实验应选择Value值最小的模型。本文方法能够在仅牺牲0.40%分类准确率的前提下，将模型的参数量压缩至0.587 M，参数量约为基准模型的0.26倍。

表3 宽度因子调整实验

Tab.3 Experiment of width factor adjustment

宽度因子 α	参数量/M	准确率/%	Value
1.00(基准)	2.229	86.04	—
0.75	1.282	85.70	0.36
0.35	0.303	83.30	1.42
0.25	0.161	82.83	1.55
0.50(本文方法)	0.587	85.64	0.24

4.4.3 消融实验

为验证本文对MobileNetV2的多个改进点有利于模型的性能提升与轻量化，在数据增强后的华为垃圾分类数据集上进行消融实验，共设计6种实验进行评估：(1)基准模型；(2)基准模型引入DSDC模块；(3)设置基准模型的宽度因子为

0.50；(4)基准模型中引入GConv结合CS技术；(5)基准模型训练过程中使用迁移学习策略；(6)本文方法。各实验在测试集上的准确率随迭代次数的变化情况如图7所示。

消融实验的结果如表4所示。DSDC模块和宽度因子调整效果在本文“4.4.1”与“4.4.2”实验中已经得到验证；与基准函数对比，GConv结合CS技术不仅降低了0.953 M的参数量，还提升了0.67%的分类准确率；迁移学习策略在不增加参数量的前提下，将分类准确率提高了3.14%。最终，本文提出的DG-MobileNetV2在华为垃圾分类数据集上的分类准确率为90.58%，模型参数量仅0.403 M，相较于MobileNetV2，分类准确率提升4.54%，参数量减少约81.9%。

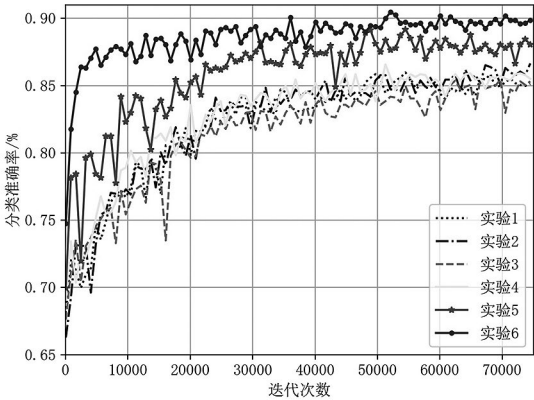


图7 测试集准确率曲线

Fig.7 Test set accuracy curve

表4 消融实验

Tab.4 Ablation experiment

实验序号	参数量/M	准确率/%
1(基准)	2.229	86.04
2	2.336	87.37
3	0.587	85.64
4	1.276	86.71
5	2.229	89.18
6(本文方法)	0.403	90.58

注：加粗字体表示各列最优结果。

4.4.4 不同模型性能对比实验

为验证本文模型的性能，在数据增强后的华为垃圾分类数据集上进行性能对比实验。将本文算法与VGG、GoogleNet、ResNet等经典模型及SqueezeNet、MobileNet、ShuffleNet等轻量型模型进行对比，以分类准确率和模型参数量为指标，实验结果如表5所示。从表5可以看出，本文算法的分类准确率为90.58%，分别高于VGG16、GoogleNet、SqueezeNet、MobileNetV2、ShuffleNetV2模型6.61%、5.21%、12.96%、4.54%、5.61%，仅低于ResNet50模型0.47%，分类效果良好；本文算法的参数量仅有0.403 M，远远低于VGG16、ResNet50等经典模型，与轻量型模型相比，本文算法的参量数约为SqueezeNet的54.7%、MobileNetV2的18.1%、ShuffleNetV2的32.0%，在参数量方面具有绝对的优势。

表5 不同模型性能对比实验

Tab.5 Performance comparison experiment of different models

网络	参数量/M	准确率/%
VGG16	134.277	83.97
GoogleNet	9.942	85.37
ResNet50	23.516	91.05
SqueezeNet	0.737	77.62
MobileNetV2	2.229	86.04
ShuffleNetV2	1.258	84.97
本文算法	0.403	90.58

注：加粗字体表示各列最优结果。

5 结论(Conclusion)

针对人工分拣生活垃圾过程存在的问题，本文在MobileNetV2的基础上进行改进，提出轻量型生活垃圾分类模型DG-MobileNetV2，具体的改进点如下：(1)引入DSDC模块，提高模型的特征提取能力；(2)调整宽度因子，压缩模型参数量；(3)引入通道混洗技术，保证特征图通道信息的融合，降低模型参数量的同时提升分类效果。实验过程中，采用数据增强、迁移学习、变学习率等训练策略，得出结果：DG-MobileNetV2在华为垃圾分类数据集上的分类准确率达到90.58%，参数量仅为0.403 M，充分证明了本文模型的鲁棒性与有效性。下一步工作是尝试将本文模型移植至移动端或嵌入式设备中，部署到实际的生活垃圾分类场景，发挥模型参数量小、轻量型的优点，辅助人们更好地对生活垃圾进行分类。

参考文献(References)

[1] 梁蕊,陈冠益,颜蓓蓓,等.城市生活垃圾智能分类技术研究与应用进展[J].中国环境科学,2022,42(01):227-238.

[2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature: International weekly journal of science, 2015, 521(7553):436-444.

[3] KAREN S, ANDREW Z. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014, 6(1):1-14.

[4] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Computer Society, 2015:1-9.

[5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society, 2016:770-778.

[6] LANDOLA F, HAN S, MOSKEWICZ M, et al. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and<0.5

(下转第24页)