浙江理工大学学报,第 49 卷,第 3 期,2023 年 5 月
Journal of Zhejiang Sci-Tech University
DOI:10.3969/j.issn.1673-3851(n).2023.03.003

基于改进 GhostNet 的轻量级手势图像识别方法

田秋红,孙文轩,章立早,施之翔,潘　豪,吴佳璐
(浙江理工大学计算机科学与技术学院,杭州 310018)

摘　要:卷积神经网络应用于复杂背景的手势图像识别时,存在深层模型参数量大、计算成本高、轻量级模型准确率低等问题,针对这些问题提出了一种基于改进 GhostNet 的轻量级手势图像识别方法。首先,在 Ghost 模块中添加通道混洗操作,建立 CS-Ghost 模块以提取手势图像中的手势特征;然后,选用 SMU(Smoothing maximum unit)激活函数优化模型在反向传播中的学习能力;最后,使用注意力机制中的轻量级通道注意力模块 ECA 去除特征中的噪声信息。该方法在 ASL 和 NUS-Ⅱ数据集上的实验平均准确率分别为 98.82%和 99.36%;在 OUHANDS 数据集上的实验平均准确率为 97.98%,参数量为 1.2 Mi,FLOPs 为 0.29 Gi。实验结果表明该方法参数量小,计算成本低,可有效提高手势图像识别的准确率。

关键词:手势图像识别;卷积神经网络;轻量级模型;注意力机制;激活函数
中图分类号:TP181　　**文献标志码**:A　　**文章编号**:1673-3851 (2023) 05-0300-10
引文格式:田秋红,孙文轩,章立早,等. 基于改进 GhostNet 的轻量级手势图像识别方法[J]. 浙江理工大学学报(自然科学),2023,49(3):300-309.
Reference Format: TIAN Qiuhong,SUN Wenxuan,ZHANG Lizao,et al. Lightweight gesture image recognition method based on improved GhostNet[J]. Journal of Zhejiang Sci-Tech University,2023,49(3):300-309.

Lightweight gesture image recognition method based on improved GhostNet

TIAN Qiuhong, SUN Wenxuan, ZHANG Lizao, SHI Zhixiang, PAN Hao, WU Jialu
(School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: When convolutional neural network is applied to the recognition of gesture images with complex backgrounds, the deep model has a large number of parameters and high computational cost, and the accuracy of the lightweight model is low. To solve these problems, a lightweight gesture image recognition method based on improved GhostNet was proposed in this paper. Firstly, channel shuffling operation was added to the Ghost module, and the CS-Ghost module was designed to extract gesture features from gesture images. Then, SMU (smoothing maximum unit) was selected to activate the function to optimize the learning ability of the model in the back propagation. Finally, the lightweight channel attention module ECA in the attention mechanism was used to remove the noise information in the feature. The experimental average accuracy of the proposed method on ASL and NUS-Ⅱ datasets are 98.82% and 99.36%, respectively. The experimental average accuracy on the OUHANDS dataset is 97.98%, the parameter quantity is 1.2 Mi, and the FLOPs is 0.29 Gi. The experimental results show

收稿日期:2022-10-31　　网络出版日期:2023-01-16
基金项目:国家自然科学基金项目(51405448);浙江省教育厅一般科研项目(Y202250600);浙江省大学生科技创新活动计划大学生科技创新项目(2022R406A014)
作者简介:田秋红(1976-),女,辽宁兴城人,教授,博士,主要从事机器学习、模式识别和图像处理与识别方面的研究。

that the proposed method has small parameters, low computational cost, and effectively improves the accuracy of gesture image recognition.

Key words: gesture image recognition; convolution neural network; lightweight model; attention mechanism; activation function

0 引　言

手势是一种自然形态的交互方式,表达意义丰富;手势图像识别在人机自然交互中可以为用户提供更加真实的交互体验[1]。近年来,手势图像识别在机器控制、虚拟现实和辅助驾驶等领域中发挥着重要的作用。现有手势图像识别任务的解决方法主要分为基于机器学习技术的传统方法和基于卷积神经网络为主的深度学习方法[2]。

传统手势图像识别方法针对特定数据集,人工设计手势特征进行手势建模[3]。Tian 等[4]使用 YCbCr 特征提取出有效的手臂区域,并使用 SVM 分类器进行手势分类;该方法对简单背景的手势图像识别准确率较高,在复杂背景的手势图像中识别效果较差。Sadeddine 等[5]提出了一种基于梯度局部自相关描述符、Gabor 小波变换和快速离散曲线变换的静态手势识别方法,识别率达 94%。以上方法采用的特征易于提取,但提取特征较为单一,复杂手势图像的识别准确率不高。为了提高手势识别的准确率,一些学者采用更丰富的手势特征,并通过特定机器学习方法进行手势图像识别。杨述斌等[6]提取手势图像中的 HOG 特征并进行 PCA 降维,再将特征归一化处理,识别准确率高于一般机器学习方法。以上传统手势图像识别的方法需要对特征进行针对性调整,且容易受到背景与光照等因素的影响,要求数据集中手势动作简单,背景噪声较小,难以推广使用。

近年来,深度学习在图像识别领域有着广泛的应用,其中卷积神经网络(Convolutional neural network, CNN)由于其无需人工设计特征受到了广泛关注。Pardasani 等[7]将 CNN 应用到机器人上,识别人类的简单手势,在美国手语数据集上达到 85%的准确率。Khotimah 等[8]使用 CNN 对动态和静态两个场景的手势进行分类,平均准确率为 89%。以上两种方法通过简单的 CNN 实现了手势图像识别,但准确率不高,因此一些学者使用更复杂的模型进行识别。Kwolek 等[9]提出了一种基于生成性对抗网络和 ResNet 模型的方法对日本手语图像进行分类。Xie 等[10]使用 Inception V3 模型对表达 24 个英文字母的手势数据集进行分类,采用两阶段训练策略对模型进行微调,准确率达到 91.35%。Tao 等[11]提出了一种利用 CNN 进行多视角增强的手语识别方法,该方法具有较高的识别精度,但模型的计算成本较高。Singh 等[12]构建了基于 VGG16 的手势图像识别系统,该系统对手势图像的识别率为 96.7%。以上使用复杂 CNN 的方法能够提升手势图像识别的准确率,但随着网络加深,模型的计算成本越来越高,为了加快模型识别速度,一些学者采用轻量级模型进行手势图像识别。辛文斌等[13]提出了一种 ShuffleNetv2 作为主干网络的 YOLOv3 模型,同时采用 CBAM 模块优化特征提取,能够得到较快的识别速度。Wang 等[14]提出了一种改进的轻量级模型 E-MobileNetv2 进行手势图像识别,准确率达到了 96.82%,并且减少了 30%的参数量。Ansari 等[15]提出了一种使用 MobileNetV2 与 SSD 相结合的方法进行手势图像识别,大幅减少了模型计算成本,但识别的准确率只有 44.7%。上述基于轻量级模型的方法能够有效降低计算成本,但提取到的特征不够丰富且存在较多的噪声信息,手势图像识别的准确率较低。

为了提高轻量级模型在手势图像识别任务中的准确率,本文提出了一种基于改进 GhostNet 的轻量级手势图像识别方法。该方法在 Ghost 模块的基础上加入通道混洗操作,设计了能够对不同通道的特征进行重新分配的 CS-Ghost (Channel shuffle ghost)模块。该模块可以增强通道间的信息交流从而提取丰富的特征信息;同时,采用 SMU 激活函数避免 ReLU 函数中的神经元死亡问题,加强模型在训练过程中的特征学习能力;最后,使用轻量级通道注意力模块 ECA 去除特征中的噪声信息,以增强有效特征的表达能力。本文提出方法对 GhostNet 结构进行优化,在减少计算成本的同时,进一步提高手势图像识别的准确率。

1　方法设计

对于包含特定手势的图像,手势图像识别任务需要理解图像内容,排除背景干扰,强化手势特征并准确识别出手势类型。本文建立了轻量级模型 CS-GhostNet,对复杂手势图像中的手势进行分类。首先,在 Ghost 模块中加入通道混洗操作,设计了 CS-Ghost 模块,该模块能够提取更丰富的手势特征;其

次,使用CS-Ghost模块和SMU激活函数搭建CS-Ghost瓶颈层,增强模型的学习能力;然后,利用ECA模块减少特征中的噪声信息;最后,构建出CS-GhostNet模型,实现手势图像识别。

1.1 网络结构

CS-GhostNet网络结构示意图如图1所示,该模型建立在GhostNet的基础上,网络结构为:首先使用一层卷积层提取尺寸为224×224×3的手势图像特征;再将特征输入10层CS-Ghost瓶颈层和6个CS-Ghost-ECA模块中,输出尺寸为7×7×160的特征到卷积层中;接着经过一层平均池化层、一层卷积层和一层全连接层,最终得到形状为1×1×1280的特征进行手势分类。CS-Ghost-ECA模块由CS-Ghost瓶颈层和ECA模块组合得到,具体结构如图1所示。为了减少ECA模块对模型增加的计算成本,本文只使用6个CS-Ghost-ECA模块。其中5个模块在传递过程中改变特征尺寸,该操作可以有效利用ECA模块的注意力机制增强手势特征表达能力,第6个CS-Ghost-ECA模块用于在分类前强化手势特征,增强模型的分类能力。

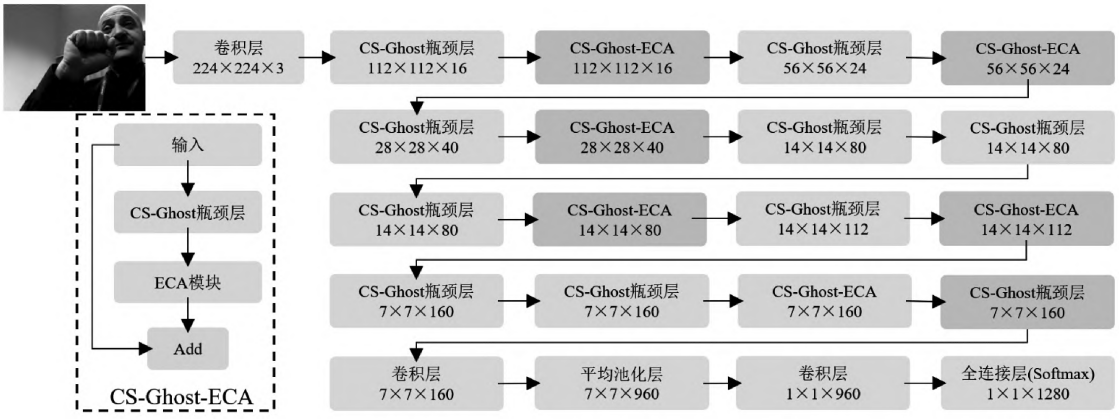


图1 CS-GhostNet网络结构示意图

1.2 CS-Ghost模块

手势图像中会存在一些非手势的干扰物品,卷积操作会从图像中提取出手势特征和非手势特征。为了强化其中的手势特征,需要将同一手势的多幅图像输入网络进行训练,这种操作会产生大量的特征图。Han等[16]发现,在特征图中存在部分相似的特征,这些相似特征可增强模型对输入数据的理解程度,有助于提升模型的准确率;但这些相似特征在卷积过程中产生的冗余映射会消耗大量计算资源。为了减少冗余映射带来的计算量,目前普遍采用轻量级模型GhostNet通过简单的线性操作生成特征图。GhostNet由多个Ghost模块组成,Ghost模块的具体结构如图2(b)所示,每个Ghost模块通过三步操作获得与普通卷积一样数量的特征图。第一步操作是少量卷积,相对图2(a)中的普通卷积操作,少量卷积只使用相当于普通卷积一半数量的卷积核,减少了一半的计算量;第二步,对特征图进行廉价操作ϕ,其中$\phi_1,\phi_2,\cdots,\phi_m$表示对$m$个通道中的特征图逐个进行线性变换,线性变换会选择计算成本低的深度可分离卷积操作;第三步,对恒等映射后的特征图和线性变换后的特征图进行拼接,得到最终的输出特征。

在Ghost模块中会生成两组特征图,其中第二组特征图由第一组特征图通过线性变换得到。由于两组特征图中存在较多的相似特征且通道结构一致,模型在训练的过程中只能学习到其中一组特征图的主要信息,而另一组信息被忽略。因此,本文设计了CS-Ghost模块,使用ShuffleNetV2[17]中的通道混洗操作来增强两组特征图不同通道之间的信息交流,具体结构如图2(c)所示。其中通道混洗操作是在通道的层面上打乱特征的顺序,首先假设一组特征图中有N个特征通道,将其看作一个$(1,N)$的一维数组并重塑成$(g,N/g)$的多维数组,其中g为分组的数量,值为2;然后对多维数组进行转置,构成$(N/g,g)$的数组;最后对其进行重塑,将特征数组变回$(1,N)$,完成通道混洗操作。通过打乱特征通道的位置顺序,CS-Ghost模块能够同时学习到两组特征图的信息,从而提升模型的特征提取能力。

1.3 基于SMU激活函数的CS-Ghost瓶颈层

ReLU激活函数具有快速的收敛能力。ReLU函数对负的特征值直接归零,特征值在原点不可微的特性使得下一层出现更多的负值特征,最终超过50%的神经元在模型训练期间死亡。相比ReLU函数,SMU函数在原点处可微,在模型训练时能够

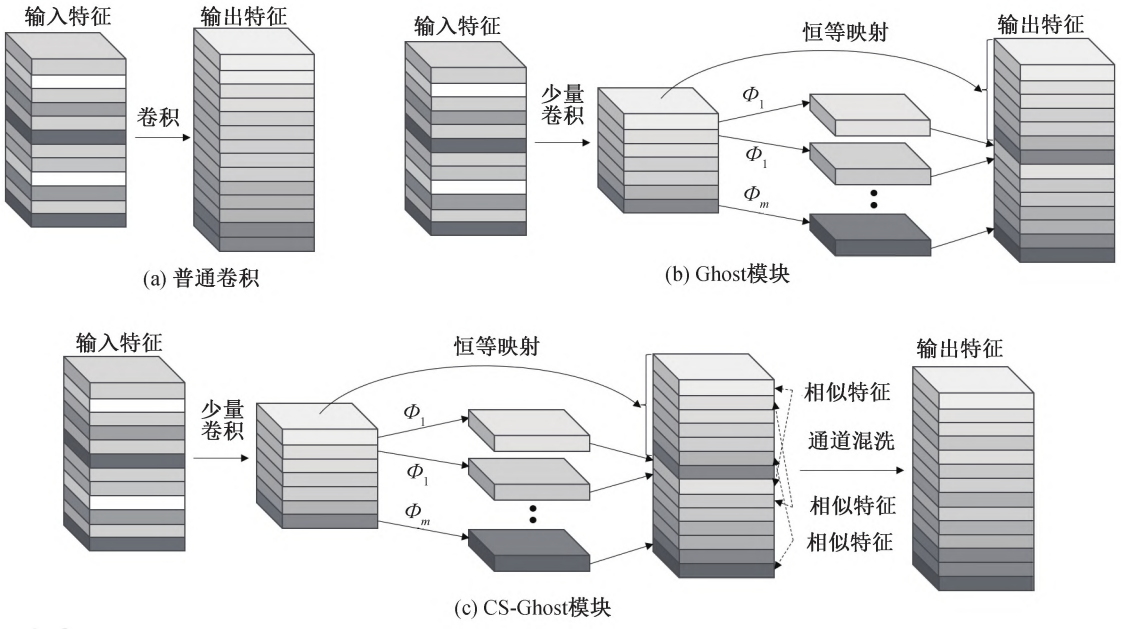


图 2　普通卷积、Ghost 模块和 CS-Ghost 模块结构示意图

更加平滑地传递特征到下一层,有效避免 ReLU 函数导致的神经元死亡问题。应用 SMU 函数的模型可以在训练过程中更好地从手势图像中学到手势特征。因此,为了提升模型在训练时的稳定性,本文选用 SMU 激活函数[18]代替 ReLU 函数,该函数的公式可用式(1)表示:

$$f_{\mathrm{SMU}}(x)=\frac{(1+a)x+(1-a)x\cdot erf(u(1-a)x)}{2} \tag{1}$$

其中:a 是一个超参数,默认值设为 0.25;u 是一个可训练参数,初始化值为 1000000;$erf()$是高斯误差函数,定义为:

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t \tag{2}$$

本文参照 ResNet 中的残差结构[19],通过 CS-Ghost 模块和 SMU 激活函数构建 CS-Ghost 瓶颈层,如图 3 所示。CS-Ghost 瓶颈层分为步长为 1 和步长为 2 两种结构,每种结构主要由两个 CS-Ghost 模块组成。对于步长为 1 的 CS-Ghost 瓶颈层,在第一个 CS-Ghost 模块后添加一个批量归一化层(BN)和一个 SMU 激活函数;同时根据 MobileNetV2[20]的建议,在第二个 CS-Ghost 模块后使用一个批量归一化层而不使用激活函数,以避免信息损失;最后使用 Add 操作将输入特征与经过两个 CS-Ghost 模块后的特征进行相加,得到输出特征。对于卷积步长为 2 的 CS-Ghost 瓶颈层,需要使用步长为 2 的深度可分离卷积(Depthwise separable convolutions,DWConv)对特征进行空间下采样,其他结构与步长为 1 的 CS-Ghost 瓶颈层相同。使用 SMU 激活函数代替 ReLU 函数,CS-Ghost 瓶颈层在模型训练时可以接收到更多的有效特征。

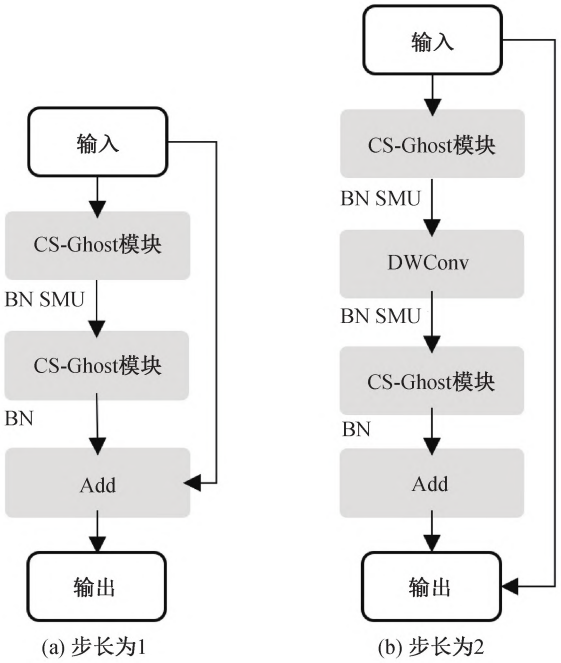


图 3　不同步长的 CS-Ghost 瓶颈层结构示意图

1.4　ECA 模块

在复杂背景下,手势图像中一些环境干扰因素,例如光照以及背景中不同的物品等,在训练过程中这些因素会产生影响模型准确率的不利特征,并且在训练过程中被传播放大,可视为噪声。通道注意机制削弱了这些背景特征的通道,降低了它们的权重[21],从而减少干扰因素对模型的负面影响。大多

数注意力机制模块无法兼顾计算成本和识别性能[22],例如SE模块[23]在通道之间交换信息并进行特征降维,这个操作会增大模型的计算成本,且对通道注意力的预测产生负面影响[24]。

ECA(Efficient channel attention)模块是一种轻量的通道注意力模块,使用一维卷积组合相邻通道上的特征进行特征加权,以补偿特征维数减少所造成的缺陷,避免了特征降维操作的负面影响。此外,ECA模块采用了跨通道交互,在保持性能的同时不会过多增加模型的计算开销。因此,轻量级模型适合引入ECA模块,在提高模型特征提取能力的同时保持模型的轻量级特性。本文在CS-Ghost瓶颈层的Add操作之前嵌入ECA模块,对残差模块生成的特征进行校准,加强模型的识别能力。ECA模块的具体结构如图4所示,其中:W是特征图的宽度,H是特征图的高度,C是通道的数量,GAP(Global average pooling)表示全局平均池化层。ECA模块能够根据通道数自适应地确定卷积核大小K,从而节省计算资源。K的计算公式如式(3)所示:

$$K=\psi(C)=\left|\frac{\log_2 C}{\gamma}+\frac{b}{\gamma}\right|_{\text{odd}} \tag{3}$$

其中:b和γ是固定数值的系数,其值分别为1和2;$|\,|_{\text{odd}}$代表取最接近其值的奇数。

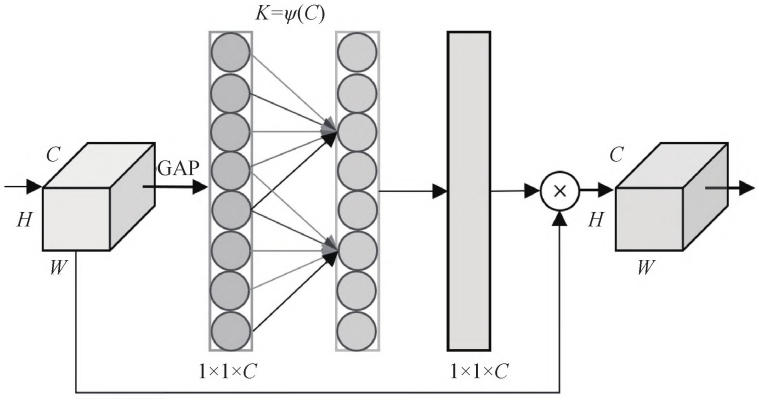
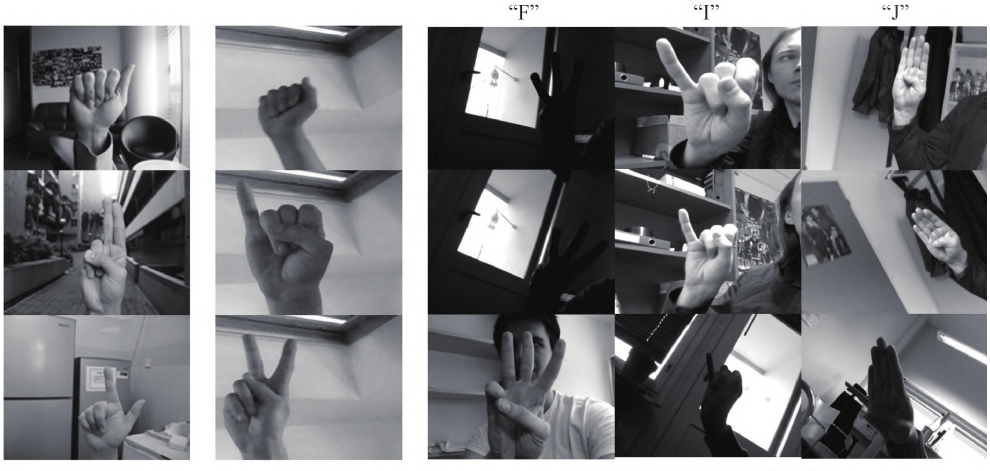


图4 ECA模块结构示意图

2 结果与讨论

2.1 实验数据集

本文使用NUS-Ⅱ和ASL手势图像数据集进行实验,图5(a)和图5(b)分别为NUS-Ⅱ和ASL手势数据集的示例图像。NUS-Ⅱ数据集由50名受试者在不同背景下采集制作,包含10种不同的手势,共有2000幅图像,ASL数据集提供了一套手势代替从"A"到"Z"的26个字母以及"Delete""Space"和"Nothing"字符,共29个手势类别,包含80000幅图像。



(a) NUS-Ⅱ (b) ASL (c) OUHANDS(主要数据集)

图5 NUS-Ⅱ、ASL和OUHANDS手势数据集示例图像

为了验证本文方法的有效性,本文将复杂程度更高的OUHANDS手势图像数据集作为主要数据集。OUHANDS数据集由深度传感摄像头拍摄捕捉,23名受试者,包含从"A"到"K"(不包含"G")10

种不同的手语动作。该数据集中图像背景较为复杂,共包含 28 种不同的背景。图 5(c)中展示了“F”“I”和“J”3 种手语动作,每种手语动作选取 3 幅图像放在同一列中,每列的前两幅图像处于同一背景下,第三幅图像和前两幅图像的背景不同。手势“F”中两种背景图像分别处于暗光和正常光环境下,其中正常光环境中的手势放在人像前面,存在肤色干扰因素。对于手势“J”,相同背景的两幅图像中手势的位置和角度不同,不同背景的图像之间光源位置不同。OUHANDS 中拥有 3000 幅 RGB 图像,每幅图像数据的分辨率为 640×480,本文将其中的 80%划分为训练集,20%划分为测试集。

2.2　实验准备

实验环境如下:服务器操作系统为 Ubuntu16.04,Python 版本 3.7.2,使用的深度学习框架为 TensorFlow2.3,显卡为 Nvidia GeForce GTX 2070Ti,实验选用 Adam 算法作为模型参数优化器,BatchSize 的大小设置为 16,训练周期为 100 次。在训练之前对图像进行预处理,先对读入的原始手势图像进行尺寸归一化,变成 224×224×3 的三通道 RGB 图像,再对三通道 RGB 图像进行标准化,将三通道 RGB 图像的像素从 0～255 的整数映射为 0～1 的浮点数,最后输入模型进行训练和测试。

2.3　激活函数对比实验

为了验证 SMU 激活函数的有效性,本文在 CS-GhostNet 模型中使用 5 种激活函数在 OUHANDS 数据集上进行对比实验,对于每种激活函数,本文进行了 20 次测试,最终求出每种激活函数对应的平均准确率及方差。采用不同激活函数的模型平均准确率如表 1 所示,使用 SMU 激活函数的模型平均准确率为 97.98%,相比 Sigmoid 函数和 Tanh 函数分别提高了 0.42%和 0.76%。在反向传播的过程中,Sigmoid 函数和 Tanh 函数饱和区域接近于 0 且非常平缓,容易出现梯度消失的问题,导致网络中神经元的权重无法即时更新。SMU 函数在超参数确定的情况下,正输入时得到的结果是线性的,能够完整传递梯度,可以避免梯度消失问题。同时,SMU 函数的平均准确率相比 ReLU 函数和 Leaky ReLu 函数分别提高了 0.06%和 0.12%。由于 ReLU 函数解决了梯度消失问题,所以其平均准确率相对 Sigmoid 函数和 Tanh 函数有所提升,但 ReLU 函数在输入负值的情况下存在神经元坏死的问题,Leaky ReLU 函数在负半轴添加了一个小的正斜率,确保神经元的权重在负值输入的情况下仍然可以更新。但 Leaky ReLU 函数中使用的斜率很小,影响权重更新的速度,最终会影响模型的平均准确率。SMU 函数通过平滑逼近的方式更新权重,在避免神经元坏死问题的同时加快模型的收敛速度,在模型训练过程中能够传递更多的有效参数,得到的特征更加契合手势图像,最终提高了模型的平均准确率。

为了验证激活函数对模型稳定性的影响,本文计算了 5 种激活函数的准确率方差。从表 1 中可以看出,Tanh 函数的方差最大,达到了 0.32%,当输入较大或较小时,Tanh 函数的输出较为单一,不利于权重更新,最终影响了模型的收敛速度使得平均准确率不够稳定。Sigmoid 函数和 Leaky ReLU 函数的方差较为接近,分别为 0.23%和 0.24%,Sigmoid 函数中的梯度消失问题影响了模型收敛速度,Leaky ReLU 函数在负值输入下使用小斜率不利于权重更新。ReLU 函数的方差为 0.19%,其在正输入时输出是线性的,负输入时输出直接为 0,计算速度快且不存在梯度消失的问题。SMU 函数的方差最小,只有 0.16%,SMU 函数能够平滑地传递特征到下一层,权重更新快,模型收敛速度加快,使得平均准确率波动幅度小,从而增强模型的稳定性。

表 1　采用不同激活函数的模型平均准确率

激活函数	平均准确率/%
Sigmoid	97.56±0.23
Tanh	97.22±0.32
ReLU	97.92±0.19
Leaky ReLU	97.86±0.24
SMU	97.98±0.16

2.4　通道混洗机制和 ECA 模块的性能验证实验

为了探索通道混洗机制和 ECA 模块对模型产生的影响,本文将对 GhostNet 和 CS-GhostNet 嵌入不同注意力模块进行对比实验,实验结果如表 2 所示。从表 2 中可以看出,CS-GhostNet 相对 GhostNet 平均准确率有显著提升,在不加注意力模块的情况下,CS-GhostNet 相对 GhostNet 提高 0.92%的平均准确率,参数量没有变化,FLOPs 增加了 0.01 Gi。将 SE 模块、CBAM 模块和 ECA 模块分别加入模型后,CS-GhostNet 相对 GHostNet 平均准确率分别提高了 0.84%、0.80%和 0.90%,参数量和 FLOPs 没有显著变化,这是因为通道混洗机制用于增强模型的特征提取能力且花费的计算成本较低,能够进一步提升模型的性能。加入 SE 模块与 CBAM 模块的 CS-GhostNet 平均准确率分

别提升了0.50%和0.26%,但SE模块将模型的参数量从1.19 Mi增加到4.36 Mi,FLOPs从0.29 Gi增加到0.32 Gi,CBAM模块增加了1.34 Mi的参数量和0.02 Gi的FLOPs。SE模块和CBAM模块虽然提升了模型的平均准确率,但增大了模型的计算成本。加入ECA模块的CS-GhostNet在平均准确率提升了0.52%的同时参数量只增加了0.01 Mi,且FLOPs没有增加。与实验中的其他注意力模块对比,ECA模块在提升平均准确率的同时不会显著影响模型的参数量和计算成本,有效提升模型的性能。

表2 嵌入不同注意力模块时CS-Ghost模块和ECA模块的平均准确率、参数量和FLOPs

模型	注意力模块	平均准确率/%	参数量/Mi	FLOPs/Gi
GhostNet	无	96.54	1.19	0.28
GhostNet	SE	97.12	4.35	0.32
GhostNet	CBAM	96.92	2.42	0.31
GhostNet	ECA	97.08	1.19	0.29
CS-GhostNet	无	97.46	1.19	0.29
CS-GhostNet	SE	97.96	4.36	0.32
CS-GhostNet	CBAM	97.72	2.43	0.31
CS-GhostNet	ECA	97.98	1.20	0.29

2.5 CS-GhostNet的性能验证实验

为了验证提出的CS-GhostNet模型的有效性,将此模型与主流的分类模型ResNet50、VGG16、ShuffleNetV2以及MobileNetV2在OUHANDS数据集上进行实验,在平均准确率和FLOPs两个方面作对比分析,实验结果如表3所示。

表3 CS-GhostNet与主流模型在OUHANDS数据集上的平均准确率和FLOPs

模型	平均准确率/%	FLOPs/Gi
ResNet50	97.32	4.10
VGG16	97.38	3.13
ShuffleNetV2	95.43	0.41
MobileNetV2	93.40	0.59
CS-GhostNet	97.98	0.29

由表3可见,本文提出的CS-GhostNet模型在平均准确率上比ResNet50高0.66%,比VGG16高0.60%,且FLOPs为0.29 Gi,远低于ResNet50的4.1 Gi和VGG16的3.13 Gi,CS-GhostNet在保持模型轻量特性的同时平均准确率能够高于VGG16和ResNet50这些计算成本较高的模型。与ShuffleNetV2和MobileNetV2相比,CS-GhostNet不仅平均准确率分别提升了2.55%和4.58%,FLOPs也分别降低了0.12 Gi和0.30 Gi,CS-GhostNet的平均准确率提升幅度较大且计算成本也低于两个轻量级模型实验结果表明,CS-GhostNet在保持模型低计算成本的同时实现了较高的平均准确率,是一种性能优良的网络模型。

2.6 训练性能对比实验

为了测试本文提出的CS-GhostNet模型在训练时的性能,将该模型与VGG16、ResNet50、MobileNetV2、ShuffleNetV2和GhostNet模型在训练时的准确率变化情况进行对比分析,这6种模型在OUHANDS数据集上的准确率变化曲线如图6所示。由图6可知,CS-GhostNet在60次迭代后基本收敛,VGG16和ResNet50在70次迭代后收敛,GhostNet、MobileNetV2和ShuffleNetV2都在80次迭代后才开始收敛,且在80到100次迭代中准确率曲线仍然存在一定程度的波动。轻量级模型GhostNet、MobileNetV2和ShuffleNetV2的收敛速度较慢且稳定性较差。ResNet50和VGG16的稳定性较好,但收敛速度一般。CS-GhostNet的收敛速度最快且稳定性好,这是因为通道混洗操作和ECA模块能够让模型在训练前期提取到更多的有效特征,加快了模型的收敛速度;SMU函数增强了模型在反向传播中的学习能力,提高了模型的稳定性。观察图6中模型的准确率变化曲线,可以发现CS-GhostNet的准确率最高,VGG16和ResNet50准确率略低,GhostNet、ShuffleNetV2和MobileNetV2的准确率较低。实验结果表明,在模型训练的过程中,CS-GhostNet在收敛速度和稳定性方面表现较好,识别准确率达到最高,模型整体性能优秀。

2.7 时间性能验证实验

为了测试提出方法的时间性能,本文将不同方法的训练时间和预测时间进行对比,实验结果如表4所示,其中训练时间是指模型从开始训练到100个周期训练完成所花费的时间,预测时间是指已经训练完成的模型对于预测一幅分辨率为640×480的手势图像花费的时间。从表4中可以看出,CS-GhostNet的训练时间为0.72 h,相对MobileNetV2和ShuffleNetV2分别减少了0.05 h和0.08 h,比训练时间最长的Xception减少了0.62 h。CS-GhostNet对于单幅图像的预测时间只需要232 ms,比轻量级模型MobileNetV2和ShuffleNetV2分别减少6 ms和11 ms,比ResNet50和Xception分别减少了30 ms和81 ms。CS-GhostNet的训练时间较短,表明了模型具有较低计算成本。相比其他模

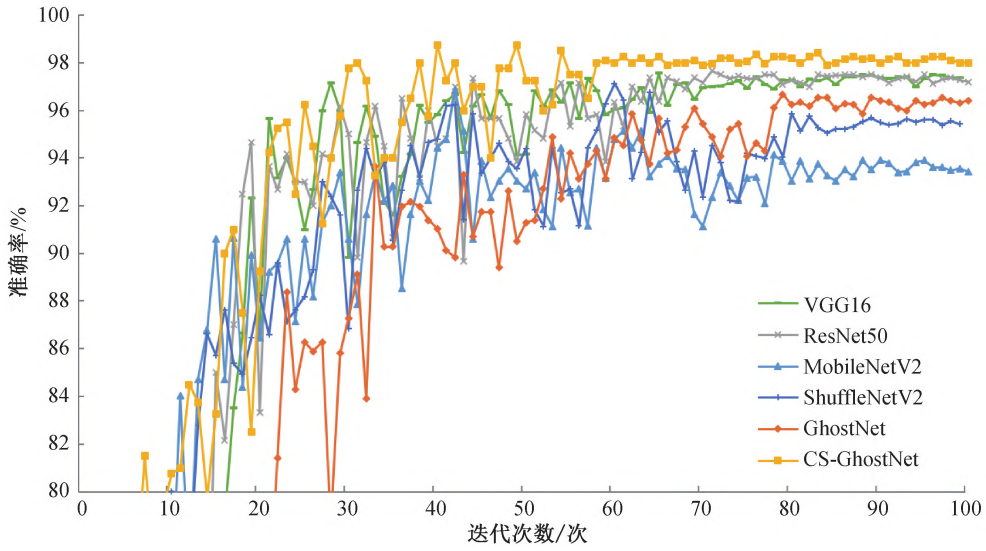


图 6　VGG16、ResNet50、MobileNetV2、ShuffleNetV2、GhostNet 和 CS-GhostNet 的准确率曲线

表 4　CS-GhostNet 的时间性能对比

模型	训练时间/h	预测时间/ms
ResNet50	0.82	262
VGG16	1.22	294
Xception[25]	1.34	313
MobileNetV2	0.77	238
ShuffleNetV2	0.80	243
CS-GhostNet	0.72	232

型,CS-GhostNet 对单幅图像预测的时间成本较低,在时间性能方面有一定的优越性。

2.8　不同数据集上的对比实验

为了测试提出模型的泛化性能,在 OUHANDS 数据集、ASL 数据集以及 NUS-Ⅱ数据集上进行 CS-GhostNet 和其他模型的对比实验。由表 5 可见,在 OUHANDS 数据集的实验结果中,CS-GhostNet 的平均准确率为 97.98%,在所比较的方法中平均准确率最高。ASL 数据集图像背景较为简单,VGG16 在该数据集上存在过拟合现象,平均准确率为 98.46%,ResNet50 在一定程度上解决了过拟合问题,得到了 99.2%的平均准确率,CS-GhostNet 作为参数量较少的轻量级模型,过拟合风险较低,平均准确率为 98.82%,略低于 ResNet50,但高于其他模型。在 NUS-Ⅱ数据集上,CS-GhostNet 达到了 98.36%的平均准确率,高于其他模型。通过分析可知,本文提出的 CS-GhostNet 能够在 3 个数据集上获得较高的平均准确率,泛化性能良好。

表 5　不同数据集上的平均准确率对比

方法	平均准确率/%		
	OUHANDS	ASL	NUS-Ⅱ
ResNet50	97.32	99.20	97.96
VGG16	97.38	98.46	97.82
ShuffleNetV2	95.43	98.06	96.43
MobileNetV2	93.40	97.55	95.56
EfficientNet[26]	96.42	97.33	92.15
DeepConv[27]	93.72	98.63	94.70
SegNet[28]	97.49	—	—
HyFiNet[29]	—	—	97.78
Two-branch CNN[30]	90.90	—	—
CS-GhostNet	97.98	98.82	98.36

3　结　论

本文提出一种基于改进 GhostNet 的轻量级手势图像识别方法,通过通道混洗操作改进 Ghost 模块,增强特征通道之间的信息交流;使用 SMU 激活函数加强模型的特征学习能力和训练时的稳定性;加入 ECA 模块减少特征中的噪声信息。实验结果表明,采用 CS-Ghost 模块、SMU 函数和 ECA 模块可以保证模型在轻量的特性下提高手势图像的识别准确率。本文提出方法在 ASL 和 NUS-Ⅱ数据集上分别得到了 98.82%和 98.36%的平均准确率,在 OUHANDS 数据集上平均准确率达到了 97.98%,参数量为 1.20 Mi,FLOPs 为 0.29 Gi,在准确率和计算成本方面与现有手势图像识别方法相比有明显优越性。

参考文献:

[1] Jiang D, Zheng Z J, Li G F, et al. Gesture recognition based on binocular vision[J]. Cluster Computing, 2019, 22(6): 13261-13271.

[2] 王银, 陈云龙, 孙前来. 复杂背景下的手势识别[J]. 中国图象图形学报, 2021, 26(4):815-827.

[3] 陈影柔，田秋红，杨慧敏，等. 基于多特征加权融合的静态手势识别[J]. 计算机系统应用，2021，30(2)：20-27.

[4] Tian Q H，Bao J X，Yang H M，et al. Improving arm segmentation in sign language recognition systems using image processing[J]. Technology and Health Care：Official Journal of the European Society for Engineering and Medicine，2021，29(3)：527-540.

[5] Sadeddine K，Chelali F Z，Djeradi R，et al. Recognition of user-dependent and independent static hand gestures：Application to sign language [J]. Journal of Visual Communication and Image Representation，2021，79：103193.

[6] 杨述斌，潘伟，蒋宗霖. 基于HOG特征与手部多特征信息融合的静态手势识别[J]. 自动化与仪表，2020，35(8)：47-51.

[7] Pardasani A，Sharma A K，Banerjee S，et al. Enhancing the ability to communicate by synthesizing american sign language using image recognition in a chatbot for differently abled[C]//2018 7th International Conference on Reliability，Infocom Technologies and Optimization(Trends and Future Directions). Noida，India：IEEE，2018：529-532.

[8] Khotimah W N，Suciati N，Benedict I. Indonesian sign language recognition by using the static and dynamic features [C]//2018 International Seminar on Intelligent Technology and Its Applications (ISITIA). Bali，Indonesia：IEEE，2018：293-298.

[9] Kwolek B，Baczynski W，Sako S. Recognition of JSL fingerspelling using deep convolutional neural networks[J]. Neurocomputing，2021，456：586-598.

[10] Xie B，He X Y，Li Y. RGB-D static gesture recognition based on convolutional neural network[J]. The Journal of Engineering，2018，2018(16)：1515-1520.

[11] Tao W J，Leu M C，Yin Z Z. American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion [J]. Engineering Applications of Artificial Intelligence，2018，76：202-213.

[12] Singh D K，Kumar A，Ansari M A. Robust modelling of static hand gestures using deep convolutional network for sign language translation [C] // 2021 International Conference on Computing，Communication，and Intelligent Systems (ICCCIS). Greater Noida，India：IEEE，2021：487-492.

[13] 辛文斌，郝惠敏，卜明龙，等. 基于ShuffleNetv2-YOLOv3模型的静态手势实时识别方法[J]. 浙江大学学报(工学版)，2021，55(10)：1815-1824.

[14] Wang W J，He M L，Wang X H，et al. Medical gesture recognition method based on improved lightweight network[J]. Applied Sciences，2022，12(13)：6414.

[15] Ansari Z A，Harit G. Nearest neighbour classification of Indian sign language gestures using kinect camera [J]. Sadhana，2016，41(2)：161-182.

[16] Han K，Wang Y H，Tian Q，et al. GhostNet：More features from cheap operations[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle，USA：IEEE，2020：1577-1586.

[17] Ma N N，Zhang X Y，Zheng H T，et al. ShuffleNet V2：Practical guidelines for efficient CNN architecture design[C]//Ferrari，V，Hebert M，Sminchisescu C，Weiss Y. Lecture Notes in Computer Science：European Conference on Computer Vision . Springer，Cham，2018，11218：116-131.

[18] Biswas K，Kumar S，Banerjee S，et al. SMU：Smooth activation function for deep networks using smoothing maximum technique[EB/OL]. (2022-10-31)[2021-11-08]. https://arxiv.org/abs/2111.04682.

[19] He K M，Zhang X Y，Ren S Q，et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas，NV，USA：IEEE，2016：770-778.

[20] Sandler M，Howard A，Zhu M L，et al. MobileNetV2：Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA：IEEE，2018：4510-4520.

[21] He L J，Gong X L，Zhang S，et al. Efficient attention based deep fusion CNN for smoke detection in fog environment[J]. Neurocomputing，2021，434：224-238.

[22] Gao R H，Wang R，Feng L，et al. Dual-branch，efficient，channel attention-based crop disease identification [J]. Computers and Electronics in Agriculture，2021，190：106410.

[23] Hu J，Shen L，Sun G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA：IEEE，2018：7132-7141.

[24] Wang Q L，Wu B G，Zhu P F，et al. ECA-Net：Efficient channel attention for deep convolutional neural networks [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle，WA，USA：IEEE，2020：11531-11539.

[25] Chollet F. Xception：Deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，

HI, USA: IEEE, 2017: 1800-1807.

[26] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C] // International Conference on Machine Learning. Long Beach, CA, USA: PMLR, 2019: 6105-6114.

[27] Adithya V, Rajesh R. A deep convolutional neural network approach for static hand gesture recognition [J]. Procedia Computer Science, 2020, 171: 2353-2361.

[28] Yadav K S, Kirupakaran A M, Laskar R H, et al. Design and development of a vision-based system for detection, tracking and recognition of isolated dynamic bare hand gesticulated characters[J]. Expert Systems, 2022: e12970.

[29] Bhaumik G, Verma M, Govil M C, et al. HyFiNet: Hybrid feature attention network for hand gesture recognition[J]. Multimedia Tools and Applications, 2022: 1-20.

[30] Wang S, Zhang S H, Zhang X W, et al. A two-branch hand gesture recognition approach combining atrous convolution and attention mechanism[J]. The Visual Computer, 2022: 1-14.

(责任编辑:康　锋)