GSDCPeleeNet：基于 PeleeNet 的高效轻量化卷积神经网络

倪伟健，秦会斌
（杭州电子科技大学 电子信息学院 新型电子器件与应用研究所，浙江 杭州 310018）

摘 要：卷积神经网络在各个领域都发挥着重要的作用，尤其是在计算机视觉领域，但过多的参数数量和计算量限制了它在移动设备上的应用。针对上述问题，结合分组卷积方法和参数共享、密集连接的思想，提出了一种新的卷积算法 Group-Shard-Dense-Channle-Wise。利用该卷积算法，在 PeleeNet 网络结构的基础上，改进出一种高效的轻量化卷积神经网络——GSDCPeleeNet。与其他卷积神经网络相比，该网络在具有更少参数的情况下，几乎不损失识别精度甚至识别精度更高。该网络选取 1×1 卷积层中卷积核信道方向上的步长 s 作为超参数，调整并适当地选取该超参数，可以在网络参数量更小的情况下，拥有更好的图像分类效果。

关键词：图像分类；卷积神经网络；轻量化；密集连接；参数共享；分组卷积
中图分类号：TN911.73；TP399　　**文献标识码**：A　　DOI：10.16157/j.issn.0258-7998.200844

中文引用格式：倪伟健，秦会斌. GSDCPeleeNet：基于 PeleeNet 的高效轻量化卷积神经网络[J]. 电子技术应用，2021，47(3)：22-26，30.
英文引用格式：Ni Weijian，Qin Huibin. GSDCPeleeNet：efficient lightweight convolutional neural based on PeleeNet[J]. Application of Electronic Technique，2021，47(3)：22-26，30.

GSDCPeleeNet：efficient lightweight convolutional neural based on PeleeNet

Ni Weijian，Qin Huibin
（Institute of New Electronic Devices and Applications，School of Electronic Information，Hangzhou Dianzi University，Hangzhou 310018，China）

Abstract：Convolutional neural network plays an important role in various fields，especially in the field of computer vision，but its application in mobile devices is limited by the excessive number of parameters and computation. In view of the above problems，a new convolution algorithm，Group-Shard-Dense-Channle-Wise，is proposed in combination with the idea of grouping convolution and parameter sharing and dense connection. Based on the PeleeNet network structure，an efficient lightweight convolutional neural network，GSDCPeleeNet，is improved by using the convolution algorithm. Compared with other convolutional neural networks，this network has almost no loss of recognition accuracy or even higher recognition accuracy under the condition of fewer parameters. In this network，the step size s in the channel direction of convolution kernel in the 1×1 convolutional layer is selected as the super parameter. When the number of network parameters is smaller，better image classification effect can be achieved by adjusting and selecting the super parameter appropriately.

Key words：image classification；convolutional neural network；lightweight；dense connectivity；parameter sharing；grouping convolution

0 引言

随着深度学习理论的提出和硬件设备计算速度的不断突破，卷积神经网络在近年来得以迅速发展。2012 年，AlexNet[1]在 ImageNet 大规模视觉识别挑战赛中获得了图像分类冠军。之后，为了提高网络模型的准确率，研究人员不断地加深卷积网络的深度，相继提出了性能更加优越的卷积神经网络，如 VGG16[2]、GoogLeNet[3]和 DenseNet[4]等。

这些网络准确率普遍较高，但是有着非常复杂的模型和很深的层次，参数量十分巨大。在实际生活的应用中，模型往往需要在资源有限的嵌入式设备和移动设备上运行。因此，研究人员开始着手研究，并且相继提出了更高效的轻量级卷积神经网络。它们保持了网络的性能，大大减少了模型的参数量，从而减少计算量，提升了模型速度。

旷视科技的 ShuffleNet 在 ResNet[5]单元上进行了改进，有两个创新的地方：逐点分组卷积和通道混洗[6]。WANG R J 等提出的 PeleeNet 是一种轻量级网络，它在 DenseNet 基础上进行了改进和创新，主要有五个方面的结构改进[7]。ZHANG J N 等提出了一种卷积核及其压缩算法，通过这种卷积方法，ZHANG J N 等发明了轻量卷积神经网络 SDChannelNets[8]。

可以看出，上述轻量卷积神经网络均存在一定不足。在使用分组卷积时，为了解决分组卷积导致的信息丢失问题，需要增加额外的操作。在运用 1×1 卷积时，会导致 1×1 卷积的参数量在网络总参数量中占据大部分。通过分析，这些网络需要通过调整相应的超参数来提高网络识别精度。这些操作往往会大大增加网络模型参数量。

为了解决这个不足，本文结合参数共享、密集连接的卷积方法和分组卷积，基于 PeleeNet 网络，提出了轻量级卷积神经网络架构 GSDCPeleeNet。适当调节超参数，在损失较小准确度甚至拥有更高准确度的情况下，减小了模型的参数量。

1 GSDCPeleeNet 网络模型的设计

1.1 SD-Channel-Wise 卷积层

标准卷积层扫描一个宽度和高度均为 D_f、通道数为 m 的输入特征图的每个位置，输出一个宽度和高度均为 D_g、通道数为 n 的输出特征图。标准卷积层卷积核的参数在通道方向上是相互独立的，所以标准的卷积层的卷积核的大小为 $D_W \cdot D_H \cdot m$，一共有 n 个，其中 D_W 和 D_H（一般相等）是卷积核的宽度和高度。因此，标准卷积层的参数量数量为：

$$D_W \cdot D_H \cdot m \cdot n \tag{1}$$

其中，D_W 和 D_H 分别是卷积核的宽度和高度，m 和 n 分别是输入和输出特征图的通道数。

从上述分析中，标准卷积层参数的大小取决于输入特征通道数 m 和输出特征通道数 n 的乘积 $m \cdot n$ 的大小。为了能够减少卷积层的参数，SD-channel-wise 卷积层提出运用了参数共享、密集连接思想，它可以有效避免 $m \cdot n$ 的大小对卷积层参数的影响。

常规卷积层的卷积核数量为 n 个，而 SD-channel-wise 卷积层的卷积核只有 1 个。输入特征图每次只和这个卷积核的一小部分进行卷积运算，这一小部分上的参数是共享的，即权值相等。标准卷积层中的数个卷积核在卷积运算中是相互独立的。而该卷积核运用了密集连接的思想，在卷积过程中，有一部分的参数是重叠的，重叠的部分的参数的权值也是相同的。

为了更好地解释这个卷积层，先从一个高度和宽度均为 1、通道数为 m 的输入特征图说明。这个通道数量为 m 的输入特征图会逐次和这个卷积核的第 $(1+x \cdot s)$~第 $(m+x \cdot s)$ 通道进行卷积运算，最终获得一个输出通道数为 n 的 1×1 输出特征图。其中，x 是小于 n 的自然数，s 是长卷积核通道方向上的步长。当输入宽度和高度不为 1 时，卷积层的卷积核会和输入特征图的各个空间位置进行卷积运算。SD-channel-wise 卷积层参数量的数量为：

$$D_W \cdot D_H \cdot (m+(n-1) \cdot s) \tag{2}$$

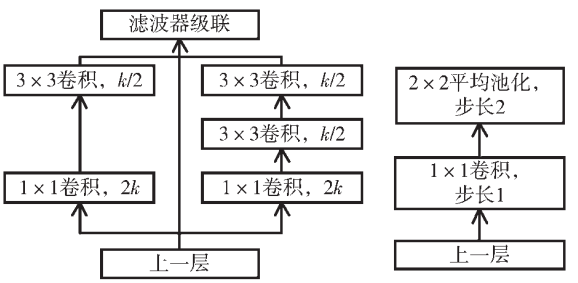
经过这样参数共享和密集连接的卷积操作，SD-channel-wise 卷积层的参数量不再取决于 $m \times n$ 的大小。因此，该卷积层的参数相比于标准卷积层，得以大幅度的下降。需要特别注意的一点是，当通道方向上的步长 s 等于或者大于输入特征图的通道数 m 时，该卷积层相当于标准的卷积层。

1.2 GSD-Channel-Wise 卷积层

在 SD-Channel-Wise 卷积的基础上，本文提出了分组 SD-Channel-Wise 卷积，命名为 GSD-Channel-Wise。在这里将给定的输入特征图分成 g 组，每组都独立进行 SD-Channel-Wise 卷积操作。每个分组将会得到 n/g 个输出特征图，通过将它们级联来获得整个卷积层的输出特征图。由于该卷积层的参数共享且连接非常密集，经过分组后提取到的特征会更多，分组卷积的优势得以进一步发挥。

在分组卷积的过程中，相同分组中不同通道可能含有一样的信息，这样就有可能会丢失部分信息，导致输出特征图得到的信息非常有限。ShuffleNet 为了解决这个问题，增加了额外的操作——通道混洗，很好地解决了这个问题，但是也相应增加了网络的复杂度。

本文着手研究其他卷积神经网络，发现 PeleeNet 的两路密集层和转换层能够很好地解决这个问题，它们结构如图 1 所示。其中，k 为增长率，两路密集层在 Inception 结构基础上进行改进，由两路不同的网络分支组成，用来获取不同尺度感受野。第一路经过一层标准的 1×1 标准卷积缩减通道数量，再经过一层 3×3 卷积层学习特征；另一路则是在用 1×1 卷积减少通道数量后，经过两层 3×3 层卷积层来学习不同的特征。转换层作为过渡，保持输入输出通道一致。

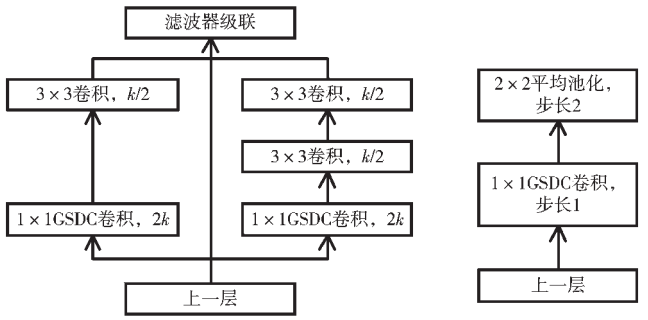


(a)两路密集层　　(b)转换层

图 1　PeleeNet 中两路密集层和转换层

通过分析 PeleeNet 网络结构，发现两路密集层和转换层占据了网络的主要部分，且 1×1 卷积的输入通道或者输出通道数目比较大，3×3 卷积的输入通道或者输出通道数目比较小，导致 1×1 卷积的参数量占据了 PeleeNet 网络的参数量的 60%以上。本文分析提出，将上述两层结构的 1×1 卷积替换成 1×1GSD-Channel-Wise，3×3 卷积使用标准卷积。经过这样的操作，不同分组间的信息可以重新流通。同时因为运用了新的卷方法，新网络的参数数量在 PeleeNet 的基础上大幅度减少。改进后的两路密集层和转换层如图 2 所示。

《电子技术应用》http://www.chinaaet.com



(a)改进的两路密集层　　(b)改进的转换层

图 2　GSDCPeleeNet 中改进的两路密集层和转换层

不同于常规的分组卷积，1×1GSD－Channel－Wise 卷积层中由于只有一个卷积核，参数并没有大幅度地减少。分成 g 组的 1×1GSD－Channel－Wise 卷积层参数量大小为：

$$1\cdot 1\cdot\left(\frac{m}{8}+\left(\frac{n}{8}-1\right)\cdot s\right)\cdot g=1\cdot 1\cdot(m+(n-g)\cdot s) \tag{3}$$

通过将标准 1×1 卷积层替换成参数共享的 1×1GSD－Channel－Wise 卷积层，将参数缩减为原来的：

$$\frac{1\cdot 1\cdot(m+(n-g)\cdot s)}{1\cdot 1\cdot m\cdot n}=\frac{1}{n}+\frac{s}{m}-\frac{g\cdot s}{m\cdot n} \tag{4}$$

1×1GSD－Channel－Wise 卷积层的具体计算如图 3 所示。可以分析出，通过调整超参数卷积核通道方向上的步长 s，模型的参数大小会相应改变。而参数量的改变在一定程度上会影响准确率改变。当选取较小的 s 和适当的 g 时，1×1GSD－Channel－Wise 相比于标准的 1×1 标准卷积层，参数量可以减少为原来的数十分之一。

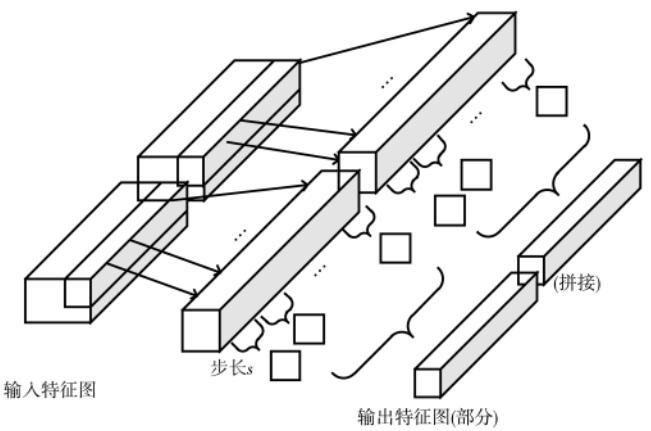


图 3　1×1GSD－Channel－Wise 卷积层卷积计算图示

1.3 GSDCPeleeNet

本文遵循了 PeleeNet 的基本架构，并对上述方法进行改进，将两路密集层和转换层 1×1 标准卷积层替换成 1×1GSD－Channel－Wise，提出了 GSDCPeleeNet。每一层输出维度和 PeleeNe 保持一致。GSDCPeleeNet 的结构如表 1 所示。

ShuffleNet v2[9]中提出的高效网络设计实用准则中指出，较大的分组卷积会提升内存访问成本，导致模型的速度反而降低。综合考虑精度和速度等因素的影响，将 1×1GSD－Channel－Wise 卷积中设置分组数量为 2。在该卷积层中，选择长卷积核通道方向上的步长 s 作为超参数，该超参数可以根据所需的精度和参数数量进行相应的调整。本文设计了 GSDCPeleeNe－s1、GSDC－PeleeNe－s32、GSDCPeleeNe－s64、GSDCPeleeNe－s192 4 种模型，其在通道方向上的步长分别为 1、32、64 和 192。它们的参数总数在 1.11 M~1.808 M 之间，占 PeleeNet (2.8 M)的 39.6%~64.5%。

表 1　GSDCPeleeNet 网络结构

输入维度			224×224×3
步骤		层	输出维度
步骤 0	茎块	/	56×56×32
步骤 1	GSDC 密集块	GSDC 密集层×3	28×28×128
	GSDC 转换层	1×1 GSDC 卷积，步长 1 2×2 平均池化，步长 2	
步骤 2	GSDC 密集块	GSDC 密集层×4	14×14×256
	GSDC 转换层	1×1 GSDC 卷积，步长 1 2×2 平均池化，步长 2	
步骤 3	GSDC 密集块	GSDC 密集层×8	7×7×512
	GSDC 转换层	1×1 GSDC 卷积，步长 1 2×2 平均池化，步长 2	
步骤 4	GSDC 密集块	GSDC 密集层×6	7×7×704
	GSDC 转换层	1×1 GSDC 卷积，步长 1	
分类层		7×7 全局平均池化 1×1×704	1×1×704

网路模型的复杂度常用浮点运算数衡量，可以理解为计算量(FLoating－point OPerations，FLOPS)。对于卷积层来说，计算量公式为：

$$\mathrm{FLOPs}=(2\cdot m\cdot K^2-1)\cdot H\cdot W\cdot n \tag{5}$$

其中，m 是输入特征图通道数，K 是卷积核大小，H、W 是输出特征图大小，n 是输出通道数通道数。不考虑偏置 bias 时有－1，考虑时没有。对于全连接层，计算量公式为：

$$\mathrm{FLOPs}=(2\cdot I-1)\cdot O \tag{6}$$

其中，I 是输入神经元数，O 是输出神经元数。经过计算，GSDCPeleeNet 的计算量为 178.6 MFLOPs，为 PeleeNet (508 MFLOPs)的 35.1%。

2 实验研究

2.1 数据集和实验设置

实验主要在数据集 CIFAR－10 和 Fashion－MNIST 进行。CIFAR－10 数据集中的图像均为彩色，像素大小为 32×32。它的训练集包含一共 5 万张不同的图像，测试集包含一共 1 万张不同的图像，共分为 10 个不同的类别。Fashion－MNIST 数据集均是灰度图像，像素大小为 28×28。它的训练集一共包含 6 万张图像，测试集包含 1 万张图像，与 CIFAR－10 一样，被划分成为 10 个不同的类别。

每张图片均进行一定的预处理。在 CIFAR－10 数据

集，本文使用通道均值和标准差对它们进行归一化处理；在 Fashion-MNIST 上，每张图像中的每个像素点均除以 255 进行归一化处理。同时数据集里的每张图片进行了数据增强，主要包括水平随机翻转和平移。

所有网络的训练方式均是随机梯度下降法[10]，动量设置为 0.9[11]，权值衰减为 5×10^{-4}。训练批量大小设置成 64，CIFAR-10 趟数设置成 100，Fashion-MNIST 趟数设置成 60；初始的学习率设置成 0.1，在趟数 25 和 50 上分别除以 10。每一层后均增加批归一化 BN(Batch Normalization)[12]层。因为只是在各个网络上进行对比，每一层都没有设置 Dropout[13]层。实验训练所有数据集中的图片，并且在训练结束后记录测试集的准确率。

2.2 实验结果与分析

本文在 CIFAR-10 和 Fashion-MNIST 数据集上使用相同的分组数 2 和不同的长卷积核通道方向上的步长 s 训练 GSDCPeleeNet，再用同样条件训练其他不同的卷积神经网络。在训练结束后，记录下两个数据集的识别正确率，结果如表 2、表 3 所示。由于两个数据集的的像素大小分别为 32×32×3 和 28×28×1，因此 GSDCPeleeNet 和 PeleeNet 都没有使用茎块，增长率(growth rate)设置为 32；DenseNet 中的密集层为 3 层，转换层为 2 层。

表 2 各模型在 CIFAR-10 上的识别准确度

网络模型	参数数量($\times10^6$)	识别正确率/%
Network in NetWork[14]	0.972	89.60
Resnet110	1.742	90.52
DenseNet100(k=12)	0.793	90.06
MobileNetV2[15]	2.289	91.17
ShuffleNetV2(1.0×)	1.275	90.28
PeleeNet	2.103	90.76
GSDCPeleeNet-s1	0.404	88.58
GSDCPeleeNet-s32	0.518	90.91
GSDCPeleeNet-s64	0.634	91.38
GSDCPeleeNet-s192	1.110	90.81

表 3 各模型在 Fashion-MNIST 上的识别准确度

网络模型	参数数量($\times10^6$)	识别正确率/%
Network in NetWork	0.972	94.02
Resnet110	1.742	94.50
DenseNet100(k=12)	0.793	94.52
MobileNetV2	2.289	94.61
ShuffleNetV2(1.0×)	1.275	94.18
PeleeNet	2.103	94.35
GSDCPeleeNet-s1	0.404	94.01
GSDCPeleeNet-s32	0.518	94.46
GSDCPeleeNet-s64	0.634	94.61
GSDCPeleeNet-s192	1.110	94.39

从表 2 可以看出，在 CIFAR-10 数据集中，GSDCPeleeNet 中长卷积核信道方向上的步长 s 取 1 时，该网络结构的参数大幅度减少，只有 0.404M，在 Peleenet 的基础上，大约减少了 80%，准确率只大约降低了 2%。当步长 s 逐渐增加时，网络结构的参量数量逐渐增大，而识别准确度有一个先上升后下降的过程。这与 SDChannelNets 中步长 s 和识别准确度呈正相关不同。这是因为，输入特征图经过分组卷积后，过大的步长 s 会使得更多的 1×1GSD-Channel-Wise 转换成标准 1×1 卷积，导致参数共享和密集连接与分组卷积结合的效果减弱。因此，选取合适的步长 s，可以在网络模型参数增加不大的情况下有一个很好的精准度。当步长 s 选取 64 时，网络的识别准确率最高，为 91.38%，优于 PeleeNet 的 90.76%和其他模型。

图 4 是各个 GSDCPeleeNet 模型在 CIFAR-10 上数据集上的测试准确率曲线对比。

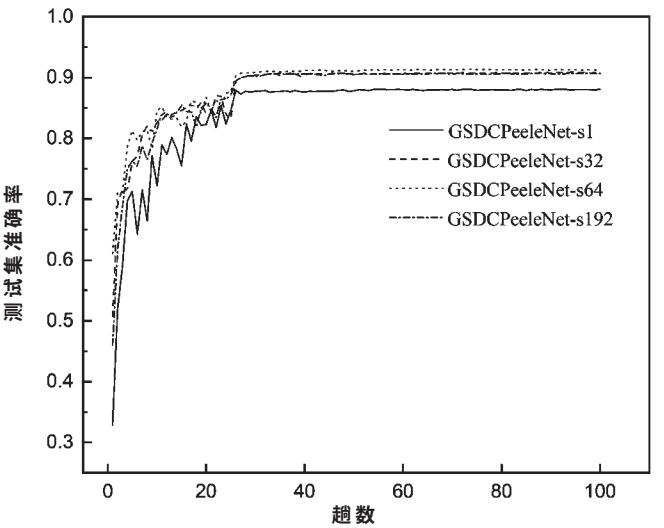


图 4 GSDCPeleeNet 各模型在 CIFAR-10 的准确率曲线

从表 3 可以看出，该模型在不同的数据集中也表现出了较强的识别能力。尤其是在 Fashion-MNIST 这样的简单数据集中，GSDCPeleeNet-s1 也表现出了不错的识别能力，准确率达到 94.01%。GSDCPeleeNet-s64 的识别准确率仍然最高，达到了 94.61%，优于 PeleeNet 的 94.35%和其他模型。图 5 是 GSDCPeleeNet 各模型在 Fashion-MNIST 的准确率曲线。

3 结束语

本文结合 SD-Channel-Wise 卷积算法和分组卷积的方法，提出改进的 GSD-Channel-Wise 卷积方法；并结合 PeleeNet 网络的结构，用该卷积方法代替网络中标准的 1×1 卷积，改进出了一种新型网络——GSDCPeleeNet。通过调整长卷积核通道方向上的超参数步长 s，可以改变网络模型的参数量和准确度。通过一定的实验，发现网络的识别准确度和步长 s 之间没有成正比。这表明，选择合适的步长 s 可以在一个较少的模型参数量上取得更高的准确率。实验结果也表明，与其他卷积神经网络相比，该卷积神经网络参数更少，而且识

《电子技术应用》http://www.chinaaet.com

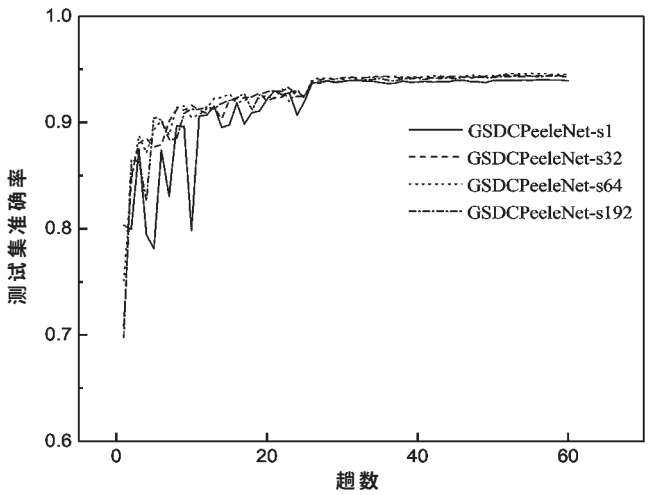


图 5 GSDCPeleeNet 各模型在 Fashion-MNIST 的准确率曲线

别效果差距不大甚至更佳。但由于 PeleeNet 网络结构的限制，GSDCPeleeNet 的网络参数和计算量还是较大。在接下去的工作中，希望通过结合其他轻量卷积神经网络结构和压缩方法，继续改进该网络。在进一步的工作中，还将会研究该网络在其他计算机视觉领域的应用，如目标检测和图像识别等。

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] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]Nnternational Journal of Computer Vision, 2015, 115(3): 211-252.

[3] SZEGEDY C, LIU W, JIA Y, et al. Going deeperwith convoutions[C]. CVPR 2015.7298594, 2015.

[4] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]. CVPR 2017, 2017: 4700-4708.

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

[6] ZHANG X, ZHOU X, LIN M, et al. Shufflent: an extremely efficient convolutional neural network for mobile devices[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

(下转第 30 页)

(上接第 21 页)

approach for post-processing in HEVC intra coding[C]. International Conference on Multimedia Modeling, 2017.

[5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[6] AIMAR A, MOSTAFA H, CALABRESE E, et al. NullHop: a flexible convolutional neural network accelerator based on sparse representations of feature maps[J]. IEEE Transactions on Neural Networks, 2019, 30(3): 644-656.

[7] RAJASEGARAN J, JAYASUNDARA V, JAYASEKARA S, et al. DeepCaps: going deeper with capsule networks[C]. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 10725-10733.

[8] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[9] HORI T, WATANABE S, ZHANG Y, et al. Advances in joint CTC-attention based end-to-end speech recognition with a deep CNN encoder and RNN-LM[C]. Interspeech 2017, 2017.

[10] Yang Yichen, Zhang Guohe, Liang Feng, et al. Design of FPGA based convolutional neural network co-processor[J]. Journal of Xi'an Jiaotong University, 2018, 52(7): 158-164.

[11] 崔小乐，陈红英，崔小欣，等. 一种软硬件协同设计工具原型及其设计描述方法[J]. 微电子学与计算机，2007，24(6)：28-30.

[12] 吴艳霞，梁楷，刘颖，等. 深度学习 FPGA 加速器的进展与趋势[J]. 计算机学报，2019(11)：2461-2480.

[13] 卢冶，陈瑶，李涛，等. 面向边缘计算的嵌入式 FPGA 卷积神经网络构建方法[J]. 计算机研究与发展，2018，55(3)：551-562.

[14] 魏浚峰，王东，山丹. 基于 FPGA 的卷积神经网络加速器设计与实现[J]. 中国集成电路，2019，28(7)：18-22.

[15] 施一飞. 对使用 TensorRT 加速 AI 深度学习推断效率的探索[J]. 科技视界，2017(31)：26-27.

[16] 杨一晨，张国和，梁峰，等. 一种基于可编程逻辑器件的卷积神经网络协处理器设计[J]. 西安交通大学学报，2018，52(7)：158-164.

[17] 张哲，孙瑾，杨刘涛. 融合多层卷积特征的双视点手势识别技术研究[J]. 小型微型计算机系统，2019，40(3)：646-650.

[18] 余子健，马德，严晓浪，等. 基于 FPGA 的卷积神经网络加速器[J]. 计算机工程，2017，43(1)：109-114.

(收稿日期：2020-08-13)

作者简介：

程佳风(1995-)，男，硕士研究生，主要研究方向：人工智能、硬件加速。

王红亮(1978-)，男，博士研究生，教授，主要研究方向：测试系统集成、目标检测与识别。

周期。在计算 CPU 搬移数据所需要的周期时不考虑进入中断和中断返回的开销。由图 6 可知，相同通道数的配置下，DMA 搬移数据所需要的周期数为 CPU 的三分之一，且 DMA 在进行数据搬移的过程中不需要 CPU 的干预，这样就提高了 CPU 的工作效率。

4 结论

本文实现了一款应用于导航 SoC 的专用 DMA，该 DMA 支持全系统全频点导航通道的数据搬移。与传统的 CPU 搬移数据方法进行比较，数据搬移所需要的周期降为 CPU 的三分之一，提高了 CPU 的工作效率。设计中采用了低功耗设计方法，使功耗降低为原来的 15%。

参考文献

[1]《卫星应用》编辑部.2019 年中国卫星应用若干重大进展[J].卫星应用，2020(1)：8-13.
[2] 梁科，李国峰，王锦，等.通用多通道高性能 DMA 控制器设计[J].天津大学学报，2008(5)：621-626.
[3] 张路煜，李丽，潘红兵，等.SoC 系统中多端口 DMA 控制器的设计[J].电子测量技术，2014，37(9)：32-36.
[4] 张美迪，马胜，雷元武.基于 AXI 协议的 DMA 接口的设计与验证[C].第二十一届计算机工程与工艺年会暨第七届微处理器技术论坛论文集，2017：10.
[5] 张路煜.支持并行传输的多端口 DMA 控制器设计[D].南京：南京大学，2014.
[6] 吴瑶裔.基于 AMBA 总线的 DMA 控制器的设计[D].长沙：湖南大学，2012.
[7] HUANG Z，ZHANG S，GAO H，et al.A configurable multiplex data transfer model for asynchronous and heterogeneous FPGA accelerators on single DMA device[J].Microprocessors and Microsystems，2020，77：103174.
[8] KATZ D J，GENTILE R.嵌入式媒体处理[M].北京：电子工业出版社，2007.
[9] 吕广秋，李伟，陈韬，等.一种面向密码 SoC 的高性能全双工 DMA 设计[J].计算机工程，2020，46(5)：167-173，180.
[10] 张帅.一种支持多种传输模式的 DMA 主机模块设计与实现[D].长沙：国防科学技术大学，2014.
[11] 王俊，应忍冬.嵌入式音频处理器中 DMA 控制器的设计[J].信息技术，2011，35(3)：42-46.
[12] 姬强.基于时钟门控技术对内存控制模块的 RTL 级功耗优化[D].西安：西安电子科技大学，2017.
[13] 包志家.大规模集成电路低功耗技术分析[J].数字通信世界，2017(12)：63，279.
[14] 王凯龙.基于通用 DMAC IP 的功耗分析及优化[D].西安：西安电子科技大学，2019.
[15] ARM.Arm AMBA 5 AHB protocol specification[EB/OL].(2015-10-30)[2021-01-15].https://developer.arm.com/documentation/ihi0033/bb.

(收稿日期：2020-09-18)

作者简介：
秦爽(1996-)，男，硕士研究生，主要研究方向：数字集成电路设计。
李健(1981-)，男，博士，副研究员，主要研究方向：超大规模集成电路设计。
杨颖(1982-)，女，博士，副研究员，主要研究方向：卫星导航算法。

(上接第 26 页)

[7] WANG R J，Li Xiang，LING C X，et al.Pelee：a real-time object detection system on mobile devices[C].Conference on Neural Information Processing Systems.arXiv：1804.06882，2018.
[8] ZHANG J N，ZHOU J J，WU J F，et al.SDChannelNets：extremely small and efficient convolutional neural networks[J].IEICE Transactions on Information and Systems，2019，102(12)：2646-2650.
[9] MA N，ZHANG X，ZHENG H T，et al.Shufflenet V2：practical guidelines for efficient cnn architecture design[C].Proceedings of the European Conference on Computer Vision(ECCV)，2018.
[10] BOTTOU L，CURTIS F E，NOCEDAL J.Optimization methods for large-scale machine learning[J].SIAM Review，2016，60(2)：16M1080173.
[11] SUTSKEVER I，MARTENS J，DAHL G，et al.On the importance of initialization and momentum in deep learning[C].Atlanta：International Conference on Machine Learning，2013：1139-1147.
[12] IOFFE S，SZEGEDY C.Batch normalization：accelerating deep network training by reducing internal covariate shift[C].Lille：International Conference on Machine Learning，2015：448-456.
[13] SRIVASTAVA N，HINTON G，KRIZHEVSKY A，et a1.Dropout：a simple way to prevent neural networks bom overfittlng[J].The Journal of Machine Learning Research，2014，15(1)：1929-1958.
[14] LEE C Y，XIE S，GALLAGHER P，et al.Deeply-supervised nets[C].San Diego：Artificial Intelligence and Statistics，2015：562-570.
[15] SANDLER M，HOWARD A，ZHU M，et al.Inverted residuals and linear bottlenecks：mobile networks for classification，detection and segmentation[EB/OL].(2018-01-13)[2020-08-13].http://arxiv.org/abs/1801.04381.

(收稿日期：2020-08-13)

作者简介：
倪伟健(1995-)，男，硕士研究生，主要研究方向：图像处理。
秦会斌(1961-)，通信作者，男，教授，博士研究生导师，主要研究方向：电路与系统，E-mail：hangdian198@sina.com。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所