



计算机应用  
*Journal of Computer Applications*  
ISSN 1001-9081, CN 51-1307/TP

## 《计算机应用》网络首发论文

题目：基于跨通道交叉融合和跨模块连接的轻量级卷积神经网络  
作者：陈力，丁世飞，于文家  
收稿日期：2020-06-19  
网络首发日期：2020-10-23  
引用格式：陈力，丁世飞，于文家. 基于跨通道交叉融合和跨模块连接的轻量级卷积神经网络[J/OL]. 计算机应用.  
<https://kns.cnki.net/kcms/detail/51.1307.TP.20201022.1005.008.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于跨通道交叉融合和跨模块连接的轻量级卷积神经网络

陈力, 丁世飞\*, 于文家

(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

(通信作者电子邮箱dingsf@cumt.edu.cn)

**摘要:** 针对传统卷积神经网络参数量过多, 计算复杂度高的问题, 提出了基于跨通道交叉融合和跨模块连接的轻量级卷积神经网络架构 C-Net。首先, 提出了跨通道交叉融合的方法, 它在一定程度上克服了分组卷积中各分组之间存在缺乏信息流动的问题, 简单高效的实现不同分组之间的信息通信; 其次, 提出了一种跨模块连接的方法, 它克服了传统轻量级架构中各基本构建块之间彼此独立的缺点, 实现了同一阶段内具有相同分辨率特征映射的不同模块之间的信息融合, 增强了特征提取能力; 最后, 基于提出的两种方法设计了一种新型的轻量级卷积神经网络架构 C-Net。C-Net 在 Food\_101 数据集上的准确率为 69.41%, 在 Caltech\_256 数据集上的准确率为 63.93%。实验结果表明: 与目前先进的轻量级卷积神经网络模型相比, C-Net 在降低存储开销和计算复杂度方面具有较好的性能。消融实验在 Cifar\_10 数据集上验证了提出的两种方法的有效性。

**关键词:** 卷积神经网络; 轻量级; 分组卷积; 跨通道交叉融合; 快捷连接; 跨模块连接

**中图分类号:** TP391.4

**文献标志码:** A

## Lightweight convolutional neural network based on cross-channel fusion and cross-module connection

CHEN Li, DING Shifei\*, YU Wenjia

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou Jiangsu 221116, China)

**Abstract :** In order to solve the problems of too many parameters and high computational complexity of traditional convolutional neural networks, a lightweight convolutional neural network architecture C-Net based on cross-channel fusion and cross-module connection was proposed. Firstly, a method called cross-channel fusion was proposed. It overcomes the shortcomings of lack of information communication between different groups of channels of group convolution to a certain extent and efficiently and easily enables information communication between different groups of channels. Then, another method called cross-module connection was proposed. It overcomes the shortcomings of independence between the basic building blocks in the traditional lightweight architecture and enables information fusion between different modules with the same resolution within the same stage, enhancing feature extraction capabilities. Finally, C-Net was designed based on the two proposed methods. The accuracy of C-Net on the Food\_101 dataset is 69.41%, and the accuracy on the Caltech\_256 dataset is 63.93%. Experimental results show that the C-Net achieves competitive performance with greatly reduced memory costs and computational complexity in comparison with the state-of-the-art lightweight convolution neural network models. The ablation experiment verified the effectiveness of the two proposed methods on the Cifar\_10 dataset.

**Keywords :** convolutional neural network; lightweight; group convolution; cross-channel fusion; shortcut connection; cross-module connection

## 0 引言

2006 年, 多伦多大学的 Hinton 教授等<sup>[1]</sup>在《Science》杂志上发表了题为“Reducing the dimensionality of data with neural networks”的文章, 标志着人工智能的复兴。计算机视

觉技术也随之取得巨大进步。卷积神经网络 (Convolutional Neural Network, CNN)<sup>[2]</sup>作为计算机视觉领域的基础模型之一, 是当前许多研究和应用的基础。自从 AlexNet<sup>[3]</sup>在 ILSVRC (ImageNet Large Scale Visual Recognition Challenge)<sup>[4]</sup>竞赛中取得了巨大成功, 目前已经提出了许多高性能的 CNN 模型, 例如 VGGNet<sup>[5]</sup>、ZFNet<sup>[6]</sup>、GoogLeNet<sup>[7]</sup>、ResNet<sup>[8]</sup>、

收稿日期: 2020-06-19; 修回日期: 2020-08-24; 录用日期: 2020-09-21。

基金项目: 国家自然科学基金 (61672522, 61976216, 61379101)。

**作者简介:** 陈力 (1993—), 男, 山东邹城人, 硕士研究生, 主要研究方向: 深度学习、图像处理; 丁世飞 (1963—), 男, 山东青岛人, 教授, 博士, CCF 会员, 主要研究方向: 人工智能、机器学习、深度强化学习; 于文家 (1994—), 男, 辽宁本溪人, 硕士研究生, 主要研究方向: 深度学习。

DenseNet<sup>[9]</sup>, SE-Net<sup>[10]</sup>等。但是, 准确率的提高并不一定会使网络在模型尺寸和速度方面更有优势。目前移动设备和嵌入式设备被广泛使用, 这些设备的计算资源和存储资源非常有限, 不利于大规模网络的部署。

2017 年之后, ImageNet 竞赛不再举办, 对 CNN 的研究也从仅追求高精度的大规模网络逐渐过渡到研究更适合实际应用的轻量级架构。精度高、规模小、速度快的轻量级架构成为当前 CNN 研究的主流。近年来, 压缩卷积神经网络以实现准确率与模型尺寸之间的权衡已成为一种热门的研究方向。一方面, 可以对现有的卷积神经网络模型进行压缩, 使网络具有较少的参数和较低的计算复杂度。这些压缩方法<sup>[11]</sup>可以分为四类: 参数剪枝和共享<sup>[12]</sup>、低秩分解<sup>[13]</sup>、迁移/紧凑卷积滤波器<sup>[14]</sup>和知识蒸馏<sup>[15]</sup>。基于参数剪枝和共享的方法主要是探索模型参数的冗余部分, 并删除对网络影响不大的参数以实现网络压缩。基于低秩分解的技术主要分解卷积核以减少冗余, 关键是如何设置卷积核的参数以及对哪个维度加以约束。基于迁移/紧凑卷积滤波器的方法设计具有特殊结构的卷积内核, 以减少存储消耗和计算复杂度。知识蒸馏的方法创建了一个与教师网络的特征分布相适应的学生网络, 以便小型网络可以学习大型网络的知识。通常, 学生网络是一个轻量级网络, 而教师网络是一个较大的网络。另一方面, 除了压缩现有的网络模型, 还可以使用一些高效的方法来设计新的网络结构, 即轻量级卷积神经网络模型设计。轻量级的 CNN 可以使更多的计算机视觉研究成果应用到工业产品和服务中。

针对一些常用的轻量级模型设计方法存在的明显不足, 本文在提出相应的解决方法的同时提出了一种轻量级卷积神经网络模型 C-Net, 旨在保证高准确率水平的前提下减少模型参数数量和降低计算复杂度。本文主要工作如下:

- 1) 针对分组卷积各分组之间缺少信息交流的问题, 提出了一种跨通道交叉融合的方法, 在增加深度和减少参数的同时实现了不同分组跨通道信息交流。
- 2) 针对特征提取不充分和特征图利用不充分的问题, 提出了一种跨模块连接方法, 在同一阶段内部实现特征重用和更好的上下文建模来增强特征的显著性。
- 3) 基于提出的两种方法设计了一种新型的轻量级卷积神经网络 C-Net。

## 1 相关工作

### 1.1 分组卷积

对于传统卷积<sup>[16]</sup>来说, 每个输出通道都与每个输入通道互相连接, 即通道之间是密集连接。例如, 如果输入特征映射的尺寸为  $C \times H \times W$ , 其中:  $C$  表示输入特征映射的通道数,  $H$  表示输入特征映射的高,  $W$  表示输入特征映射的宽, 如果

卷积核个数为  $N$ , 则输出特征映射的数量与卷积核的个数相同, 每个卷积核的尺寸为  $C \times K \times K$ , 总参数数量为  $N \times C \times K \times K$ 。

对于分组卷积, 输入通道和输出通道分成  $g$  组。每组的输出通道仅连接到相应组的输入通道而与其他通道无关。具体为首先对输入特征映射进行分组, 然后将每组分别卷积。

假设输入特征映射的尺寸为  $C \times H \times W$ , 输出特征映射的尺寸为  $N \times H \times W$ , 其中,  $H$  和  $W$  为特征映射的长度和高度,  $C$  和  $N$  分别为通道数。如果将其分为  $G$  组, 则每组中的输入

特征图的数量为  $\frac{C}{G}$ , 每组中的输出特征图的数量为  $\frac{N}{G}$ 。卷

积核的尺寸为  $\frac{N}{G} \times K \times K$ 。卷积核的个数仍为  $N$ 。每组卷积

核的个数为  $\frac{N}{G}$ , 卷积核仅与同组的输入特征映射进行卷积,

总参数数量为  $N \times \frac{C}{G} \times K \times K$ , 可见总参数数量减少到原来的

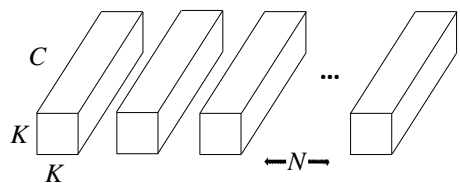
$\frac{1}{G}$ 。

显然, 尽管分组卷积可以显著减少参数数量, 但是存在不同分组之间缺乏信息交流的问题, 这将很大程度上影响模型的性能。本文提出了一种称为跨通道交叉融合的方法来实现不同分组之间信息交流。

### 1.2 深度可分离卷积

标准卷积运算实现了空间和通道维度上特征的联合映射, 但是计算复杂性高, 内存占用大, 并且权重参数过多。深度可分离卷积的基本假设<sup>[17]</sup>是输入的特征映射可以解耦。因此, 卷积过程可分成两个部分, 首先使用给定大小的卷积核分别对每个通道进行卷积并组合结果, 该部分称为深度卷积。然后使用  $1 \times 1$  卷积核<sup>[18]</sup>进行标准卷积并输出特征映射, 这部分称为逐点卷积。

深度可分离卷积的作用是将图像的空间信息和通道信息解耦, 主要目的是将空间互相关信息与通道互相关信息分离, 通过使用一个一维卷积核和一个二维位置卷积核, 分别学习图像中的通道信息和位置信息。即通过  $3 \times 3$  卷积对每个独立通道的空间区域提取特征, 然后使用  $1 \times 1$  卷积融合从不同通道中提取到的特征。



(a) 标准卷积核

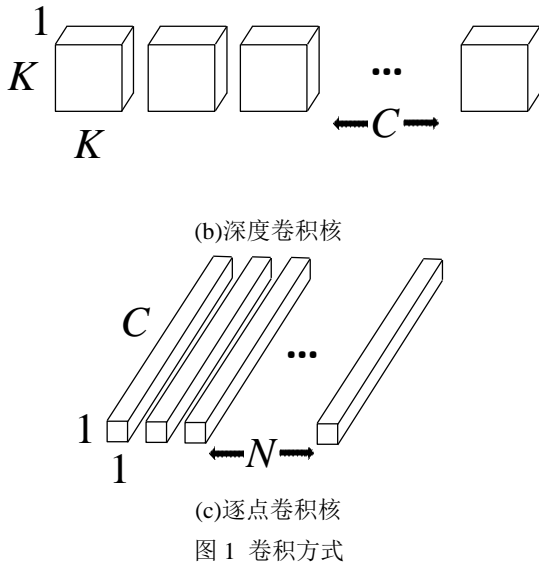


Fig.1 Convolution method

假设输入特征映射的尺寸为  $C \times H \times W$ , 输出特征映射的尺寸为  $N \times H \times W$ , 其中  $H$  和  $W$  为特征映射的长度和高度,  $C$  和  $N$  分别为通道数。标准卷积层使用大小为  $C \times K \times K$  的卷积核, 计算复杂度为  $CK^2HWN$ , 参数数量为  $CK^2N$ 。深度可分离卷积包括一组  $K \times K \times 1 \times C$  的卷积核和一组  $1 \times 1 \times C \times N$  的卷积核。深度可分离卷积的计算复杂度为  $CK^2HW + CHWN$ 。这只是标准卷积的  $\frac{1}{N} + \frac{1}{K^2}$ 。考虑输入和输出通道数数量上的差异, 深度可分离卷积的权重约为标准卷积权重的 10% 到 25%<sup>[17]</sup>。通过使用深度可分离卷积代替传统卷积可以减少模型参数, 实现模型的轻量化。

### 1.3 快捷连接

神经网络的深度在图像分类中起着至关重要的作用。随着网络层数的增加, 从理论上, 网络可以学习更多的特征, 拟合更复杂的函数, 从而提高识别的准确率。但是, 由于使用基于梯度的 BP 算法<sup>[19]</sup>优化时, 层数很深的网络在反向传播过程中容易出现梯度消失<sup>[20]</sup>或梯度爆炸<sup>[21]</sup>, 造成模型训练过程中难以收敛。

ResNet<sup>[8]</sup>是计算机视觉领域一项突破性工作。ResNet 通过添加快捷连接以引入残差学习来解决深层网络的退化问题, 从而可以训练出层数更深的网络。ResNet 的核心思想是从输入中学习残差函数而不是原始函数。快捷连接默认为恒等映射, 即直接添加原始输入而不进行任何改变直接与输出做加和作为下一层的输入。通过添加快捷连接, 使得深层网络更容易优化, 并得到更好的性能。包含一个快捷连接的多层网络称为一个残差块。

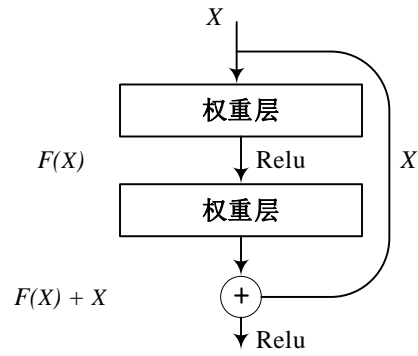


Fig. 2 Shortcut connections

将输入表示为  $x$ , 底层映射表示为  $H(x)$ , 残差映射定义为:

$$H(X) = F(X) + X \quad (1)$$

其中:  $x$  代表残差块的输入,  $H(x)$  代表残差块的输出,  $F(x)$  代表残差映射。进一步的:

$$F(X) = H(X) - X \quad (2)$$

即变成使用神经网络来拟合输入和输出之间的残差  $F(x)$  映射。

$$x = U(x) = \sigma(F(x, W) + h(x)) \quad (3)$$

其中:  $x$  代表残差块的输出,  $h(x)$  代表恒等映射且  $h(x) = x$ ;  $W$  代表一组权重 (省略偏差以简化符号),  $F(x, W)$  代表待学习的残差映射;  $\sigma(x)$  代表 Relu 激活函数,  $U(x)$  代表残差块函数。

残差映射比原始映射更易于优化。在极端情况下, 如果恒等映射是最佳输出则残差为 0, 比把恒等映射用一系列非线性层的堆叠去拟合更简单, 而且通过快捷连接实现恒等映射, 不会增加额外的参数和计算量。

ResNet 和快捷连接的提出解决了加深网络层数不能提高性能的问题。快捷连接在先进的轻量级 CNN 中均有应用。本文借鉴了残差学习的思想并加以改进, 使网络能够在同一阶段内部更好地进行局部残差学习, 该方法称为跨模块连接。

## 2 C-Net: 一种新型的轻量级卷积神经网络

### 2.1 通道交叉融合

分组卷积是轻量级卷积神经网络设计中的一种常用方法。分组卷积可以用较少的参数数量和较低的计算复杂度生成大量的特征映射, 而更多的特征映射意味着可以编码更多的信息以获得更丰富的特征。分组卷积也有明显的缺点。将输入特征映射分组导致输入信息不完整, 输出特征图不能包含输入特征图的所有信息。在 ShuffleNet<sup>[22]</sup>中, 使用通道重排来解决这个问题。



本文针对分组卷积提出了跨通道交叉融合的方法,以实现不同组通道之间的信息交流。具体分三步实现:首先,将输入特征图划分为 $k$ 组,分别标记为 $G_1, G_2, \dots, G_k$ ,并对每个组进行卷积以获得 $G'_1, G'_2, \dots, G'_k$ ,尺寸减小到输入的 $\frac{1}{2}$ ;然后再次卷积以获得 $G''_1, G''_2, \dots, G''_k$ ,尺寸保持不变;最后, $G''_1, G''_2, \dots, G''_k$ 与 $G'_1, G'_2, \dots, G'_k$ 中的相邻组合并(此处使用加法运算)作为下一层的输入,即 $G''_1 + G'_2, G''_2 + G'_3, \dots, G''_k + G'_1$ 。分组数越多,模型的参数量越少,但实现信息交流的组数就越少;分组数量越少,实现信息交流的通道数就越多。关于分组数的讨论如表1所示,其中 $k$ 为分组数。

表 1 不同分组数在 cifar\_10 的对比

Tab. 1 Comparison of different grouping numbers on cifar\_10

模型	准确率
C-Net( $k=2$ )	90.08%
C-Net( $k=3$ )	89.98%
C-Net( $k=4$ )	89.89%
C-Net( $k=5$ )	89.85%

通过对 $k$ 的不同取值发现当 $k=2$ 时,即将输入分为两组时效果最好。 $k=2$ 现了全通道的信息交流。同时,如文献[23]所建议,过多的分组会增加MAC,使用两个组也是合理的。在因此,应用于C-Net中设置 $k=2$ ,即将输入通道数分为两组,并使用 $1 \times 1$ 卷积实现两阶段的降维操作。当然,也可以用其他大小的卷积核替换它。两阶段降维方法加深网络的深度的同时减少了参数量,提高了模型的非线性建模能力。首先,将输入特征图分为通道数相同的两组 $G_1$ 和 $G_2$ 。两组中的通道数均为输入的一半,对两组使用 $1 \times 1$ 卷积得到 $G'_1$ 和 $G'_2$ ,通道数减少了一半。显然, $G'_1$ 和 $G'_2$ 包含了所在分组中所有通道的信息。然后再次卷积得到 $G''_1$ 和 $G''_2$ ,尺寸保持不变;最后使用跨通道交叉融合,将 $G'_1$ 和 $G'_2$ 添加到 $G''_1$ 和 $G''_2$ 中的相邻组中合并在一起,结果用作下一层的输入。因此,输入特征图的维数减少是通过两阶段的分组卷积和跨通道交叉融合实现的。C-Net中用于分组卷积的跨通道融合如图3所示。假设输入特征映射的尺寸为 $C \times H \times W$ ,直接使用标准 $1 \times 1$ 卷积来减小尺寸实现降维,则参数量为 $1 \times 1 \times C \times \frac{1}{2}C = \frac{1}{2}C^2$ ,使用本文的方法则参数量为: $2 \times \left(1 \times 1 \times \frac{1}{2}C \times \frac{1}{4}C\right) + 2 \times \left(1 \times 1 \times \frac{1}{4}C \times \frac{1}{4}C\right) = \frac{3}{8}C^2$ ,因此参数量减少了 $\frac{1}{8}C^2$ 。通道数 $C$ 通常比较大,因此,实现用于分组卷积的跨通道交叉融合以提高模型的泛化能力的同时减少到原始参数量的75%。

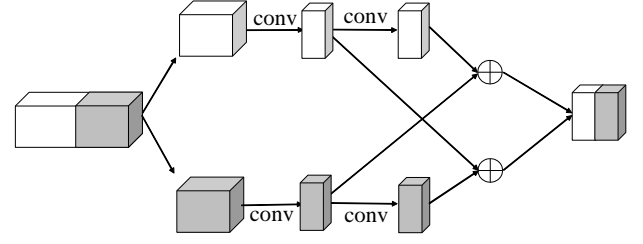


图 3 C-Net 中的跨通道交叉融合

Fig. 3 Cross-channel fusion in C-Net

## 2.2 跨模块连接

如文献[24]中建议的那样,为了提高网络的准确性,增加基数比增加网络的深度和宽度更有效,其中基数是指并行堆叠的某些模块的数量。目前最先进的轻量级CNN架构[22-23],[25-27]有一个共同的特征,即网络结构由基本构建块堆叠而成。把具有相同输入输出分辨率的几个基本构建块的堆叠称为一个阶段,相邻阶段通过池化[28-29]或步长为2卷积运算实现下采样。同一阶段内通常由许多有相同结构的基本构建块组成。通常,快捷连接通常是在单个基本构建块内部实现的,一般是在某个基本构建块的输入和输出之间添加了快捷连接。快捷连接破坏了网络的对称性并改善了网络的表征能力。通过增加快捷连接,参数量保持不变,分类精度得到进一步提高,网络的可控制性和可扩展性也得到改善。传统架构中的快捷连接如图4所示。

考虑到同一阶段内的基本构建块的输入输出具有相同的结构,为了进一步提高模型的特征提取和泛化能力,提出了一种称为跨模块连接的方法。在同一阶段内相邻的基本构建块之间,上一个基本构建块的通过两阶段降维后的特征映射和下一个基本构建块的降维后的特征映射之间添加快捷连接,然后将融合后的结果用作深度可分卷积的输入。这有助于增强模型的特征提取能力,实现上下文更好的建模。跨模块连如图5所示。

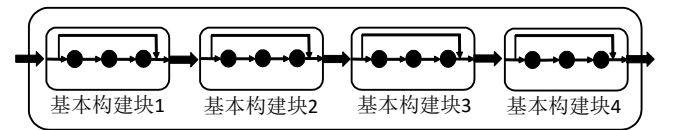


图 4 传统架构中的快捷连接

Fig.4 Shortcut connections in traditional architecture

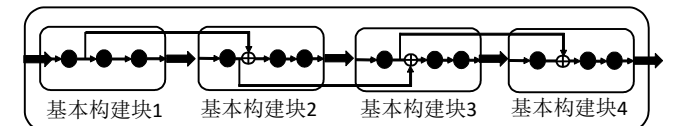


图 5 跨模块连接

Fig.5 Cross-module connections

## 2.3 C-Net 基本构建块

利用提出的两种方法,设计了一个轻量级的特征提取单元作为C-Net基本构建块。这是从残差块启发而来,采用先

压缩和后扩张策略<sup>[30]</sup>。它分为两部分：降维层和特征提取层。降维层使用两阶段跨通道交叉融合通过卷积进行输入特征映射的降维处理，降维为原始输入维的一半。之后深度可分离卷积用于特征提取。首先，使用深度卷积进行特征提取，然后使用逐点卷积增加输出通道的维度。基本构建模块如图 6 (a) (b) 所示。

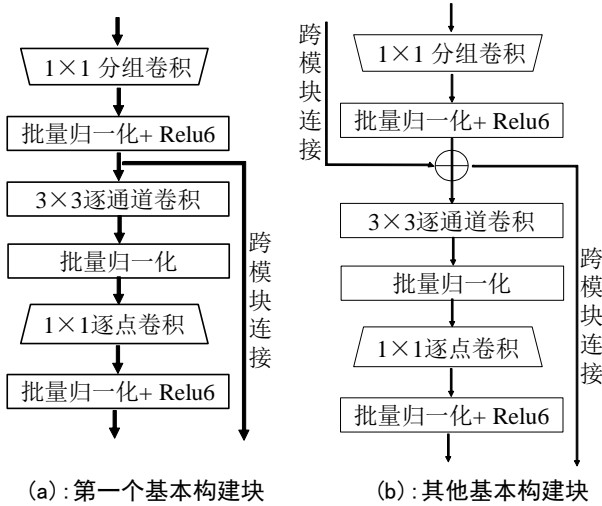


图 6 某阶段内部的基本构建块

Fig.6 The basic building block in a certain stage

通常，卷积操作后会添加非线性。最常用的激活函数是 ReLU。ReLU 的公式 (图 7) 表示如下：

$$ReLU = \max(x, 0) \quad (4)$$

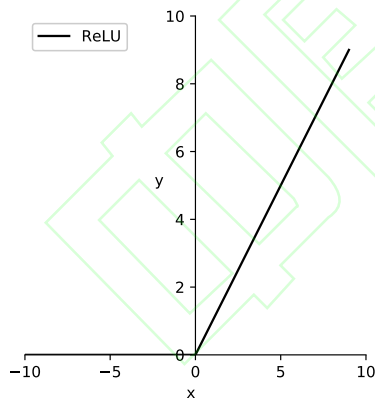


图 7 激活函数 ReLU

Fig.7 Activation function ReLU

如图 7 所示，如果对 ReLU 的激活范围没有限制，则输出范围为 0 到正无穷大；如果激活值很大并且分布在很大的范围内，则低精度时无法很好地精确描述造成精度损失。考虑到这些因素，使用 ReLU6<sup>[26]</sup>作为 C-Net 中的激活函数。ReLU6 是普通的 ReLU，但将最大输出值限制为 6。移动终端设备 float16 的精度较低，ReLU6 可以提供良好的数值分辨率。将最大值限制为 6 也有助于使网络为定点推理做好准备。ReLU6 的公式 (图 8) 表示如下：

$$ReLU = \min(\max(x, 0), 6) \quad (5)$$

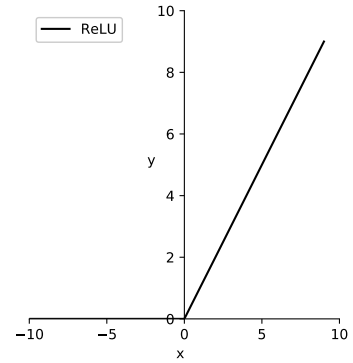


图 8 激活函数 ReLU6

Fig.8 Activation function ReLU6

Xception<sup>[17]</sup>中实验证明深度卷积后增加非线性会降低性能。一种猜测是深度卷积输出通道的数量太少，非线性的应用会造成信息丢失。因此在 C-Net 基本构建块的设计中，去除了深度卷积后的激活函数，输出为线性。

## 2.4 网络架构

C-Net 架构主要由 C-Net 基本构建块堆叠组成。网络结构如表 2 所示，其中  $k$  为类别数。除了最后的全局平均池化<sup>[31]</sup>，整个网络再不使用池化，但每个阶段的第一个基本构建块都使用步长为 2 的卷积进行下采样。全局平均池化用于防止过拟合并实现轻量级。此外，Dropout<sup>[32]</sup>和 BatchNormalization<sup>[33]</sup>用于训练。

表 2 C-Net 网络架构

Tab. 2 Overall architecture of C-Net

层		输出尺寸	输出通道数	卷积核	步幅
输出		224×224	3	-	-
Conv1		112×112	32	3×3	2
阶段 2	基本构建块 1	56×56	64	-	2
阶段 3	基本构建块 2	28×28	128	-	2
	基本构建块 3	28×28	128	-	1
	基本构建块 4	28×28	128	-	1
	基本构建块 5	28×28	128	-	1
	基本构建块 6	28×28	128	-	1
阶段 4	基本构建块 7	14×14	384	-	2
	基本构建块 8	14×14	384	-	1
	基本构建块 9	14×14	384	-	1
	基本构建块 10	14×14	384	-	1
	基本构建块 11	14×14	384	-	1
阶段 5	基本构建块 12	7×7	768	-	2
Conv6		7×7	1024	1×1	1
全局平均池化		1×1	1024	7×7	1
全连接层		k	-	-	-

## 3 实验与结果分析

### 3.1 实验设置

实验环境: 整个实验在 Ubuntu16.04LTS 和 Python 3.6, Pytorch 1.2.0<sup>[34]</sup>和 Opencv3.1 的软件环境下进行。CPU 是具有 8 核 3.7GHz 的 Intel®Xeon®E5-1620, 内存为 32GB。使用的 GPU 是具有 CUDA 10.0 和 cuDNN 7.6 的 NVIDIA-RTX2080ti。相关数据集如下:

1) cifar\_10 数据集由 10 个类别的 60 000 张彩色图像组成, 每个类别有 6000 张图像。其中有 50000 张训练图像和 10000 张测试图像。

2) food\_101 数据集包含 101 种食物的 101 000 张图片, 每张包含 1000 张图片。每个类别有 250 张测试图像和 750 张训练图像, 因此共有 75 750 张训练图像和 25 250 张测试图像。

3) caltech\_256 数据集共包含 257 个类别的 30607 张图像, 每个类别中最少 80 张图像, 最多 827 张图像, 其中一个背景类别。在本实验中, 将背景类删除, 训练集和测试集以 4: 1 的比例随机分离。

在当前的图像识别研究中, 这些公开数据集被广泛使用。其中, cifar\_10 数据集主要用于消融实验。

数据预处理: 在训练之前对所有数据集进行预处理。所有数据集均使用数据增强增加训练样本。对于 cifar\_10, 所有图像使用像素 0 填充图像的边缘, 然后将图像随机裁剪为  $32 \times 32$ 。对于 food\_101 和 caltech\_256, 首先将所有图像缩放到  $256 \times 256$ , 然后随机裁剪为  $224 \times 224$ 。在训练过程中, 按 50% 的概率水平翻转和垂直翻转。最后对所有图像进行归一化处理。

超参数设置: 在使用 food\_101 和 caltech\_256 数据集进行的实验中, 数据处理、初始化、批处理大小和训练迭代次数等实验设置保持不变。在训练期间, 批处理大小为 128, 每个 epoch 后进行一次测试迭代。所有层使用相同的学习率并将初始学习率设置为 0.1。训练过程采取学习率衰减的策略,  $\gamma$  为 0.1。学习率衰减 3 次后停止训练。使用交叉熵损失函数, 优化器为 SGD, 动量为 0.9, 权重衰减参数为 0.0005。Food\_101 在 180 000 次迭代后结束训练, 而 Caltech\_256 在 90 000 次迭代后结束训练。

### 3.2 结果分析

为了评估 C-Net 的性能, 在 food\_101 数据集和 caltech\_256 数据集进行了实验。实验评估了 food\_101 数据集和 caltech\_256 数据集上具有代表性的不同类型的典型 CNN 模型的性能。Alexnet<sup>[3]</sup>和 Resnet<sup>[8]</sup>是传统 CNN 架构的代表, 其中 Alexnet 没有快捷连接, Resnet 包含快捷连接; Squeezenet<sup>[25]</sup>, MobileNetv2<sup>[27]</sup>和 ShuffleNetv2<sup>[23]</sup>是目前最先进的轻量级 CNN 的代表, 其中, Squeezenet 通过提出 FireModule 模块, 先利用  $1 \times 1$  卷积进行降维, 再利用  $1 \times 1$  卷积+ $3 \times 3$  卷积组合升维, 实现了参数优化; MobileNetv2 通过使用深度可分离卷积和 Inverted Residual Block, 可以用较

少的运算量得到较高的精度, 适用于移动端的需求; ShuffleNetv2 通过使用 channel split 和 channel shuffle 等方法, 在速度和准确度之间做了很好地平衡。实验结果如表 3 和表 4 所示。

表 3 各模型在 Food\_101 的对比

Tab. 3 Comparisons of different networks for classification on Food\_101

模型	参数量 /MB	计算量 /M	准确率
Alexnet	57.58	858.68	60.31%
Resnet-18	11.22	1780.54	70.32%
Squeezenet1.0	0.79	771.45	62.01%
1.0MobileNetv2	2.50	356.18	69.30%
ShuffleNetv2 1.5×	2.58	600.71	69.66%
<b>C-Net</b>	<b>2.34</b>	<b>409.53</b>	<b>69.41%</b>

表 4 各模型在 Caltech\_256 的对比

Tab. 4 Comparisons of different networks for classification on Caltech\_256

模型	参数量 /MB	计算量 /M	准确率
Alexnet	58.22	859.32	55.02%
Resnet-18	11.30	1781.76	65.54%
Squeezenet1.0	0.87	784.94	55.33%
1.0MobileNetv2	2.70	356.38	63.78%
ShuffleNetv2 1.5×	2.74	601.03	64.42%
<b>C-Net</b>	<b>2.50</b>	<b>409.69</b>	<b>63.93%</b>

从表 3 和表 4 可以看出, C-Net 在 food\_101 数据集和 caltech\_256 数据集上的准确率比 Alexnet 分别提高 8.67 个百分点和 8.91 个百分点, 但参数仅为 Alexnet 的 4% 和 4.2%, 计算复杂度仅为 Alexnet 的 47.69% 和 47.67%。这充分证明传统的 CNN 架构具有很大的冗余和缺陷, 可以通过使用更有效的卷积方法和设计高质量的网络架构来实现 CNN 性能的提高, 而不仅是单纯依靠网络规模的增加。此外, C-Net 的准确率比 Resnet-18 低 1.34 个百分点和 1.61 个百分点, 但参数数量和计算复杂度远小于 Resnet-18, 大约仅为 Resnet-18 的 20% 和 25%, 更加适合在存储和计算资源有限的设备上运行。

另外, 与 Squeezenet、MobileNetv2、ShuffleNetv2 这三个最先进的轻量级模型的对比实验表明, C-Net 不仅比传统的 CNN 模型具有优势, 而且与其他最先进的轻量级架构相比也具有很强的竞争力。与 Squeezenet 相比, C-Net 准确率分别高出 6.97 个百分点和 8.60 个百分点, 尽管参数量增加, 但计算复杂度大幅降低, 这主要得益于深度可分离卷积和 shortcut 带来的特征提取能力的增强。C-Net 在两个数据集的准确率表现均介于 1.0MobileNetv2 和 ShuffleNetv2 1.5× 之间, 参数

量与 1.0 MobileNetV2 和 ShuffleNetV2 1.5 $\times$  处于同一级别, 但均比它们小, 显示出 C-Net 在移动端部署的可行性。C-Net 的计算复杂度略高于 1.0 MobileNetV2, 但仅约为 ShuffleNetV2 1.5 $\times$  的大约 68.05%, 这也解释了 C-Net 在准确度上与 MobileNetV2 基本持平却低于 ShuffleNetV2, 主要是复杂度相较 ShuffleNetV2 大大降低, 这也充分说明模型复杂度在特征提取和建模过程中的重要性。这些轻量级架构是为资源有限的设备设计的, 与某些基准模型相比具有竞争优势。各模型在 Food\_101 和 Caltech\_256 训练过程中的准确率如图 9 和图 10 所示。

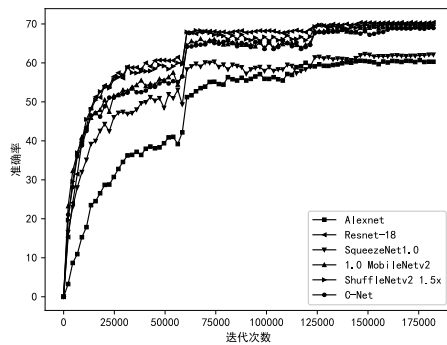


图 9 各模型在 Food\_101 的准确率曲线

Fig.9 The accuracy of each model on Food\_101

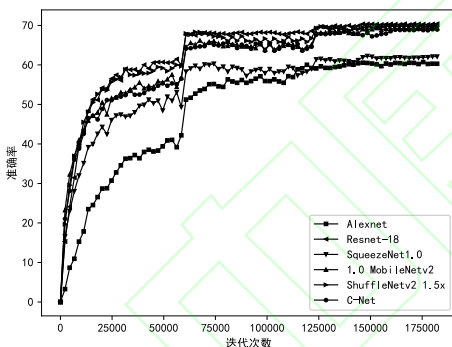


图 10 各模型在 Caltech\_256 的准确率曲线

Fig. 10 The accuracy of each model on Caltech\_256

总的来说, C-Net 不仅在参数量方面相较传统的 CNN 架构占用更少的硬件资源, 而且在不牺牲模型性能的情况下比其他几种先进的轻量级架构更轻巧。实验结果表明, C-Net 通过使用跨通道交叉融合进行分组卷积以减少参数数量并在同一阶段内部实现跨模块连接以增强特征表达, 与其他网络架构相比, C-Net 是一种高效且轻量的网络模型, 具有更少的参数和更高的识别精度。C-Net 更加便于硬件应用, 性能更高。

### 3.3 消融实验

为了对提出的两种方法进行评估, 使用 cifar\_10 数据集进行消融实验。为了适应数据集的变化, 对网络输入输出进行相应修改。输入大小改为  $32 \times 32$ , 最后一个全连接层的输

出更改为 10。此外, 由于输入分辨率的宽度和高度变小, 将阶段 2 的步幅为 2 的卷积更改为步幅为 1, 不进行下采样, 然后全局平均池化从  $7 \times 7$  变为  $4 \times 4$ 。其他参数保持不变。

#### 3.3.1 跨通道交叉融合 vs. 分组卷积

为了评估分组卷积中各分组间跨通道交叉融合对于分组卷积的重要性, 比较了是否包含跨通道融合的两组模型。由于没有引入额外参数, 因此两组模型的数量和计算复杂度相同, 其准确率曲线如图 11 所示。

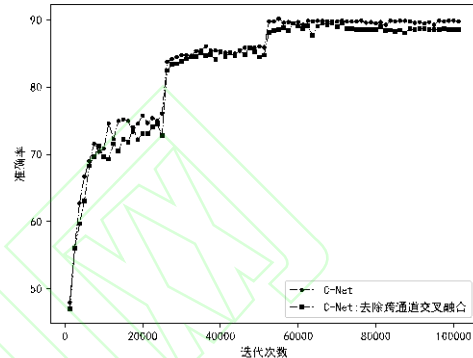


图 11 跨通道交叉融合 vs. 分组卷积

Fig.11 Cross-channel fusion for group convolution vs. grouped convolution

从图 11 可以看出, 使用跨通道交叉融合进行分组卷积的模型与使用普通分组卷积的模型在收敛速度方面没有太大差异, 但是最终精度比使用普通分组卷积的模型高 1%~2%。这表明跨通道交叉融合在实现了跨通道信息交流的同时, 提高了模型的特征提取能力和模型泛化性能。这说明了跨通道信息通信对分组卷积提高性能的重要性。

#### 3.3.2 跨模块连接 vs. 无连接

为了评估跨模块连接的重要性, 比较了是否包含跨模块连接的两组模型。由于跨模块连接也没有引入额外参数, 因此, 这两组模型的数量和计算复杂度相同, 其准确率曲线如图 12 所示。

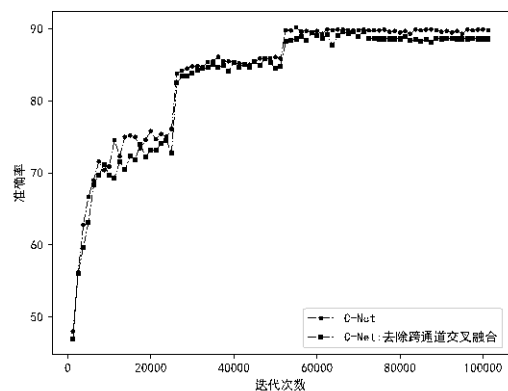


图 12 跨模块连接 vs. 无连接

Fig.12 Cross-module connections vs. no connections



如图 12 所示,使用跨模块连接的模型的收敛速度比不使用跨模块连接的模型快得多,并且其最终的分类精度明显高于不使用跨模块连接的模型,表明了跨模块连接更好地实现了特征重用和梯度反向传播,而特征提取能力的增强和反向传播的优化也带来了更好的模型泛化性能。这也显示了快捷连接对于在深层网络中梯度传播的重要性。

## 4 结语

本文提出了一种新型的轻量级卷积神经网络架构 C-Net。首先,提出了跨通道交叉融合的方法,在一定程度上解决了分组卷积中不同分组间缺乏信息交流的问题;然后,提出了跨模块连接的方法,更好地实现同一阶段内部的特征提取和特征复用;最后,基于上述两种方法设计出轻量级网络模型 C-Net。C-Net 在 Food\_101 数据集上的准确率为 69.41%,在 Caltech\_256 数据集上的准确率为 63.93%,表明 C-Net 是一种高效的轻量级神经网络结构。同时,在 Cifar\_10 数据集上的消融实验验证了提出的两种方法都可以提高 CNN 的分类准确率。由于时间和实验条件的限制,C-Net 的性能仅在图像分类任务上得到了验证。下一步工作是探索如何进一步提高 C-Net 的性能,并将 C-Net 应用于其他复杂任务,例如场景分割<sup>[35]</sup>和目标检测<sup>[36]</sup>。

## 参考文献

- [1] HINTON G, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2014-09-04]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [6] DUMOULIN V, VISIN F. A guide to convolution arithmetic for deep learning [EB/OL]. [2016-03-23]. <https://arxiv.org/pdf/1603.07285.pdf>.
- [7] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.
- [8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [9] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2261-2269.
- [10] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).
- [11] 纪荣嵘, 林绍辉, 晁飞, 等. 深度神经网络压缩与加速综述 [J]. 计算机研究与发展, 2018, 55(9): 1871-1888. (JI R R, LIN S H, CHAO F, et al. Deep neural network compression and acceleration: a review [J]. Journal of Computer Research and Development, 2018, 55(9): 1871-1888.)
- [12] SETIONO R, LIU H. Neural-Network feature selector [J]. IEEE transactions on neural networks, 1997, 8(3): 654-662.
- [13] ZHANG Y, JIANG Z L, DAVIS L S. Learning structured low-rank representations for image classification. [C] // Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 676-683.
- [14] PENG C, ZHANG X, Yu G, et al. Large Kernel Matters--Improve Semantic Segmentation by Global Convolutional Network [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4353-4361.
- [15] BUCILA C, CARUANA R, NICULESCU-MIZIL A. Model compression [C] // Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006: 535-541.
- [16] GAO L, CHEN P Y, YU S. Demonstration of convolution kernel operation on resistive cross-point array [J]. IEEE Electron Device Letters, 2016, 37(7): 870-873.
- [17] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1800-1807.
- [18] LIN M, CHEN Q, YAN S. Network in network [EB/OL]. [2013-12-16]. <https://arxiv.org/pdf/1312.4400.pdf>.
- [19] RUMELHART D E, HINTON G, WILLIAMS R J, et al. Learning representations by back-propagating errors [J]. Nature, 1986, 323(9): 533-536.
- [20] HOCHREITER S. The vanishing gradient problem during learning recurrent neural nets and problem solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107-116.
- [21] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [J]. Journal of machine learning research, 2010, 9: 249-256.
- [22] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [EB/OL]. [2017-07-04]. <https://arxiv.org/pdf/1707.01083.pdf>.
- [23] MA N, ZHANG X, ZHENG H, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [EB/OL]. [2018-07-30]. <https://arxiv.org/pdf/1807.11164.pdf>.
- [24] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5987-5995.
- [25] IANDOLA F N, MOSKEWICZ M W, ASHRAF K, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5Mb model size [EB/OL]. [2016-02-24]. <https://arxiv.org/pdf/1602.07360.pdf>.
- [26] HOWARD A G, ZHU M L, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2017-04-17]. <https://arxiv.org/pdf/1704.04861.pdf>.
- [27] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [EB/OL]. [2018-01-13]. <https://arxiv.org/pdf/1801.04381.pdf>.
- [28] LEE C Y, GALLAGHER P W, TU Z. Generalizing pooling functions in convolutional neural networks: mixed, gated, and tree [C] // Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. Cambridge, MA: MIT Press, 2016: 464-472.

- [29] SCHERER D, MULLER A, BEHNKE S. Evaluation of pooling operations in convolutional architectures for object recognition[C] //Proceedings of the 20th International Conference on Artificial Neural Networks. Cham: Springer, 2010:92-101.
- [30] DONGYOON H, JIWHAN K, JUNMO K. Deep pyramidal residual networks[EB/OL].[2016-10-10]. <https://arxiv.org/pdf/1610.02915.pdf>.
- [31] ZHANG S, ZHANG S, ZHANG C, et al. Cucumber leaf disease identification with global pooling dilated convolutional neural network[J]. Computers and Electronics in Agriculture, 2019: 422-430.
- [32] 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.
- [33] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proceedings of the 2015 International Conference on Machine Learning. New York: ACM, 2015: 448-456.
- [34] NIKHIL K. Introduction to PyTorch [M]. Berkeley, CA: Apress, 2017: 195-208.
- [35] 张蕊, 李锦涛. 基于深度学习的场景分割算法研究综述[J]. 计算机研究与发展, 2020, 57(4): 859-875. (ZHANG R, LI J T. A survey on algorithm research of scene parsing based on deep learning[J]. Journal of Computer Research and Development, 2020, 57(4): 859-875.)
- [36] 黄继鹏, 史颖欢, 高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展, 2019, 56(2): 319-327. (HUANG J P, SHI Y H, GAO Y. Multi-scale faster-RCNN algorithm for small object detection[J]. Journal of Computer Research and Development, 2019, 56(2): 319-327.)

This work is partially supported by the National Natural Science Foundation of China (61672522, 61976216, 61379101).

**CHEN Li**, born in 1993, M. S. candidate. His research interests include deep learning, image processing.

**DING Shifei**, born in 1963, Ph. D., professor. His research interests include artificial intelligence, machine learning, deep reinforcement learning.

**YU Wenjia**, born in 1994, M. S. candidate. His research interests include deep learning.