文章编号：2096-1472(2021)-10-23-05
DOI:10.19644/j.cnki.issn2096-1472.2021.010.006

基于轻量型卷积神经网络的菜品图像识别

姚华莹，彭亚雄

(贵州大学大数据与信息工程学院，贵州 贵阳 550025)
✉huayingyao97@163.com; 515154900@qq.com

摘　要：使用卷积神经网络分析研究识别菜品，能够帮助人们了解食物，根据不同的需求选择适合的菜品；同时也能被使用在自助餐厅结算系统中，提高结算效率。由于卷积神经网络有大量的卷积计算，大量参数致使卷积模型体积庞大，不利于将模型嵌入移动设备中，因此设计了一种轻量型卷积神经网络MobileNetV2−pro分类菜品。通过引入通道混洗、注意力机制提高网络的检测能力；利用随机擦除等图像预处理技术对菜品图像进行处理，提高系统的泛化能力。实验结果表明，该新结构网络能显著提高菜品分类准确率。

关键词：卷积神经网络；轻量化；菜品分类；注意力机制
中图分类号：TP391.41　**文献标识码**：A

Dishes Image Recognition based on Lightweight Convolutional Neural Network

YAO Huaying, PENG Yaxiong

(*College of Big Data and Information Engineering, Guizhou University, Guiyang* 550025, *China*)
✉huayingyao97@163.com; 515154900@qq.com

Abstract: Convolutional neural network can be used to analyze and recognize dishes, helping people know about food and choose suitable dishes according to different needs. At the same time, it can also be used in cafeteria settlement system to improve settlement efficiency. A large number of convolution calculations and parameters in the convolutional neural network make the convolution model bulky, which is not conducive to embedding the model in a mobile device. This paper proposes to design a lightweight convolutional neural network MobileNetV2-pro to classify dishes. Channel shuffling and attention mechanism are introduced to improve the detection ability of the network. Image preprocessing techniques such as random erasure are used to process the image of dishes to improve the generalization ability of the system. Experimental results show that the new structure network can significantly improve the accuracy of dish classification.

Keywords: convolutional neural network; lightweight; dishes classification; attention mechanism

1 引言(Introduction)

随着人们生活质量的提高，菜品种类变多，利用卷积神经网络能高效地实现菜品的分类。首次应用了卷积神经网络(Convolution Neural Networks，CNN)的AlexNet[1]在ImageNet图像分类竞赛中取得了优异的成绩，由此卷积神经网络得到研究人员的广泛关注，并衍生出新的网络结构(如GoogLeNet[2]、VGG[3]、ResNet[4]等)。虽然这些网络在图像分类上的精度不断提高，但是新的问题是卷积网络结构大多使用卷积层与全连接层的组合，用来提取图片特征，全连接层训练的网络模型内存占用高，大量卷积层导致计算量巨大。近几年，一些学者提出了轻量神经网络(Lightweight Neural Network)，如MobileNet[5]采用深度可分离卷积减少卷积运算量；ShuffleNet[6]提出通道混洗，打乱原有的通道顺序并重新分组，有效地提高了特征的提取。类似的轻量神经网络模型还有SqueezeNet[7]、Xception[8]等。轻量神经网络模型是专门针对嵌入式视觉应用终端设计的轻量且高效的神经网络模型[9]，

基金项目：贵州省科技成果转化项目([2017]4856).

这类模型具有计算资源需求少，模型简单的优点，能够有效提高计算机视觉的性能。

本文提出一种新轻量化神经网络模型，体积更小，运算量更少，易于应用在各类移动端用于识别菜品。该网络基于MobileNetV2[5]基础模型，结合ShuffleNet[6]提出的通道混洗思想，引入通道注意力机制加强特征学习能力，在训练网络时利用随机擦除技术对图片部分像素进行擦除，多方面对基础模型进行改进，提高了模型在菜品分类上的准确率。

2 相关工作(Related work)

2.1 深度可分离逆残差卷积块

本文为尽可能减少卷积过程中的运算量，采用了深度可分离卷积(Depthwise Sparable Convolution)替代传统卷积，用一个深度卷积和一个点卷积替换标准卷积，有效减少了卷积运算量。首先进行深度卷积，即对每个输入的通道各自用单个卷积核进行对应的卷积运算，每个通道各自得到的卷积结果则为深度卷积的最终结果；然后是一个1×1卷积，即点卷积，负责将深度卷积过程输出的卷积结果线性组合，构建新的特征[10]。如果不考虑偏置参数，深度分离后的卷积参数运算量为：

$$D_K\cdot D_K\cdot M\cdot D_F\cdot D_F+M\cdot N\cdot D_F\cdot D_F \tag{1}$$

标准卷积计算量为：

$$D_K\cdot D_K\cdot M\cdot N\cdot D_F\cdot D_F \tag{2}$$

其中，$D_K\cdot D_K$为卷积核尺寸，$D_F\cdot D_F$为输入图像尺寸，M和N分别是输入通道数量和输出通道数量。图1中对比了深度可分离卷积和传统卷积过程。

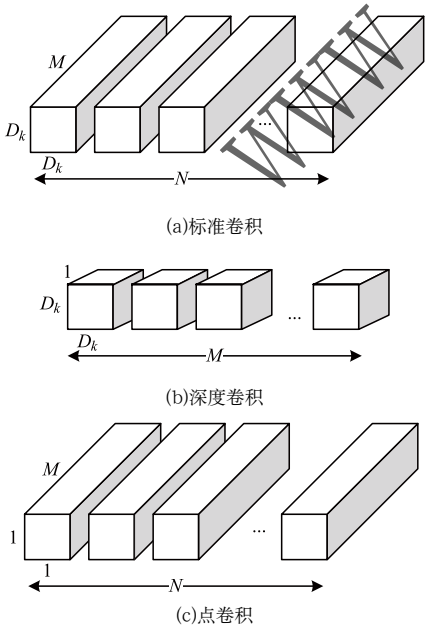


图1 传统卷积与深度可分离卷积对比

Fig.1 Comparison of traditional convolution and depth separable convolution

传统卷积的计算量是深度可分离卷积的$(1/M+1/K^2)$倍，当卷积核大小为3×3时，计算量相比传统卷积减少了九倍多。

在新的模型中，采用残差模块提高特征提取能力，浅层网络与深层网络所包含的特征量不同，通过“特征映射”和跳跃式的连接形式，可以融合不同分辨率的特征。

图2使用了一种“逆残差结构”，对输入特征通道先扩充后缩减，用1×1卷积核代替3×3卷积核，减少计算量。由于1×1卷积核得到的信息少于3×3卷积核，模型准确度受到了一定程度的影响，因此，使用逆残差结构用来保证得到的特征量足够至不影响模型精度[11]。表1为逆残差结构的卷积实现架构。

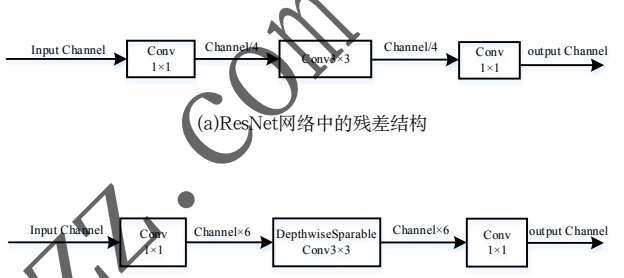


图2 残差结构与逆残差结构对比

Fig.2 Comparison of residual structure and inverse residual structure

表1 残差结构卷积实现架构

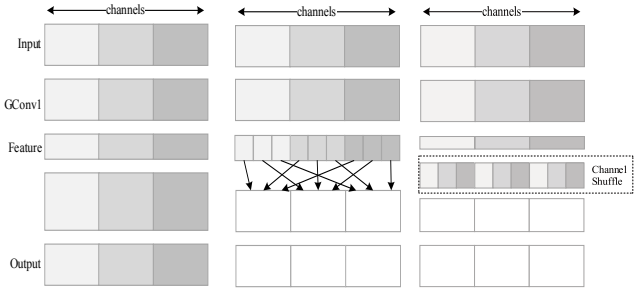
Tab.1 Architecture of residual structure convolution

输入	操作	输出
h×w×N	SE block	h×w×N
h×w×N	1×1Conv2d,ReLU6	h×w×tN
h×w×tN	3×3DW,s=s,ReLU6	h/s×w/s×tN
h/s×w/s×tN	Linear 1×1Conv2d	h/s×w/s×M

2.2 通道混洗卷积

依据卷积过程中数据仅在固定通道之间流动这一特点，在本文的新网络结构中引入通道混洗[6](Channel Shuffle)，它是基于通道分组卷积实现的通道混合卷积。通道混洗基于分组卷积技术，将输入通道分为g组，每组分别与对应的1个卷积核卷积，这样做使计算量降低为普通卷积的$1/g$，对每组通道进行打乱重组，原本封闭固定的通道经过打乱重组后特征得到交流，解决了由于分组固定导致特征融合效果差的问题。图3(a)为普通分组卷积，分组固定，特征无法交流；图3(b)表示对每个组内通道再次分组；图3(c)为通道混洗，将图

3(b)中的每一小组通道组合起来。



(a)普通分组卷积　(b)通道再次分组　(c)通道混洗

图3 Channel shuffle架构

Fig.3 Architecture of channel shuffle

2.3 通道注意力机制

注意力机制类似人眼，将重点关注特征明显的区域，运用在卷积过程中，能将不重要的背景因素剔除，本文使用了通道注意力机制，更多地关注菜品的特征部分。通道注意力机制[12]关注通道间的联系，有一个SE块由压缩(Squeeze)和激发(Excitation)两个部分构成。经过SE块后的特征被赋予不同的权重，表示出特征之间不同的重要程度，引入了注意力机制的网络能提高学习特征的能力，进一步提高识别的准确率[13]。

图4为本文使用了注意力机制和未使用注意力机制的MobileNet的菜品特征图，可以明显看出，本文的网络处理的图片白色亮点区域更多，说明提取到图片特征点更多。通过对网络部分卷积特征层的可视化，不同的卷积层的注意力响应程度不一，可以看到在conv_4后的高层卷积，都对菜品中鸡蛋的部位响应更加强烈，而对碗这种与菜品关系弱的部分响应较弱。

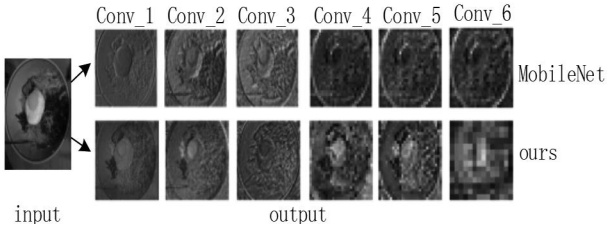


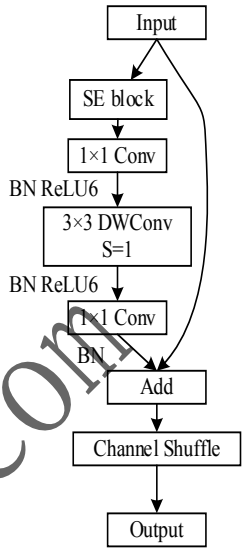
图4 基础网络与采用注意力机制网络的特征图对比

Fig.4 Comparison of feature maps between the basic network and the network using the attention mechanism

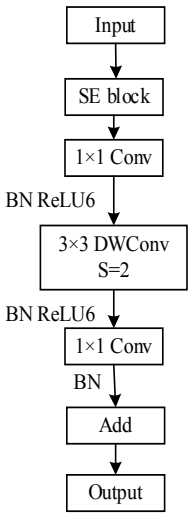
2.4 菜品识别网络模型

本文新模型的架构针对输入的特征图会首先进行一次通道注意力机制处理，此操作能够对输入的通道进行加权处理，得到不同通道中特征的重要程度。

如图5(a)所示，新的残差结构在步距S=1时，在模块最后增加了一个Channel Shuffle层，加强通道间的特征交流；如图5(b)所示，由于在步距S=2阶段没有残差结构，遂不经进行混洗操作。最后将得到的菜品特征信息通过全连接层进行分类。



(a)步距S=1时残差结构



(b)步距S=2时残差结构

图5 菜品识别网络模型

Fig.5 Dishes recognition network model

表2显示了MobileNetV2-pro模型每层的输出形状和参数量。多次叠加使用深度可分离逆残差卷积块减少卷积计算量；在浅层卷积块中使用注意力机制快速确定菜品位置和特征点，有效降低了自然环境中背景对菜品定位的影响；深层卷积层利用通道混洗技术提高数据之间的汇通融合能力，在深层的特征图中最大限度地学习菜品特征。

表2 MobileNetV2-pro整体网络架构

Tab.2 Overall network architecture of MobileNetV2-pro

卷积层	输出尺寸	重复次数	参数量
Input layer	224 × 224 × 3	–	0
Conv2d	112 × 112 × 32	–	864
Bottleneck1	112 × 112 × 16	1	960
Bottleneck2	56 × 56 × 24	2	14,384
Bottleneck3	28 × 28 × 32	2	25,648
Bottleneck4	14 × 14 × 64	2	77,888
Bottleneck5	14 × 14 × 96	2	191,552
Bottleneck6	7 × 7 × 160	2	492,672
Bottleneck7	7 × 7 × 320	1	475,200
Conv2d 1 × 1	7 × 7 × 1280	–	412,160
Avgpool 7 × 7	1 × 1 × 1280	–	0
FC & Softmax	1×1×10	–	12,810

3 实验结果与分析(Experimental results and analysis)

3.1 实验评估

为了验证本文提出的新型网络在菜品分类上的有效性，使用数据集进行验证。FOOD-101是包含101 种菜品的图像数据集，包含101,000 张图像，每类菜品包含250 张验证集和750 张训练集，图片最大边长为512 像素。图6为数据集中的部分菜品图像。

图6 FOOD-101数据集部分菜品

Fig.6 Part of dishes in FOOD-101 dataset

3.2 实验环境

使用NVIDIA Geforce RTX 1060、pytorch 1.5，在Windows 10环境下训练网络。Batchsize为64，共设置200 个epoch，初始学习率为0.001，在epoch分别达到50、80时调整学习率为上一阶段的一半。

3.3 参数分析

对训练图像做预处理，将输入网络的图片随机翻转和裁剪为224×224大小，使用不同的擦除概率讨论最优值，用于训练网络。

本文利用随机擦除对图像做预处理，对图片中的部分像素进行擦除，模拟自然环境中的遮挡情况，在此过程中将生成擦除面积不同的图片，能够增加数据集训练数据。通过预处理后，网络具有更高的鲁棒性。随机擦除的实现步骤如下：

(1)设置擦除的概率 P，则不被擦除的概率为1-P，假设图片大小为：

$$S = H \times W \tag{3}$$

(2)设置擦除矩形区域的参数，可以得到擦除的面积为：

$$S_e = S \times rand(s_l, s_h) \tag{4}$$

s_l 和 s_h 是设置的最小擦除面积和最大擦除面积，随机擦除矩形长宽比为 r_e，此值随机产生。随机擦除的矩形高和宽为：

$$\begin{cases} H_e = \sqrt{S_e \times r_e} \\ W_e = \sqrt{S_e / r_e} \end{cases} \tag{5}$$

(3)在图像中随机选择一个点 $A(x_e, y_e)$，被擦除的区域为 $(x_e, y_e, x_e + W_e, y_e + H_e)$，对选择的区域随机赋值，其中点$A$需满足：

$$\begin{cases} x_e = rand(0, W) \\ y_e = rand(0, H) \\ x_e + W_e \leqslant W \\ y_e + H_e \leqslant H \end{cases} \tag{6}$$

随机擦除效果如图7所示，分别是未进行擦除的原图及最大擦除概率为0.2、0.4的效果图。随机在原图像中生成同原尺寸比例为0.2或0.4的矩形块，模拟遮挡情况和提高模型的泛化能力。

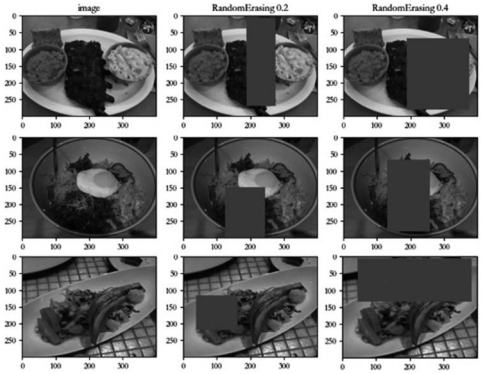


图7 随机擦除效果

Fig.7 Random erasure effect

对菜品进行随机擦除，模拟自然环境下被遮挡的情况，被遮挡部分在学习过程中卷积计算值为零，减少了卷积运算量。从图8可以明显看出，在数据集FOOD-101上，随着随机擦除比例的改变，模型分类准确率逐步上升，我们把随机擦除概率设置为0.4时，模型在食物数据集上的分类准确率最高。

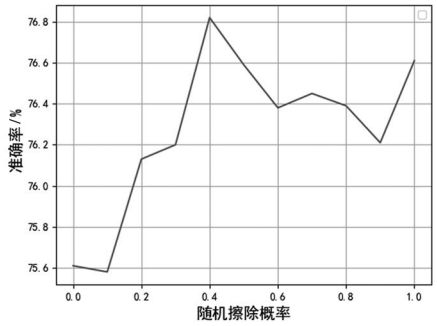


图8 随机擦除对准确率的影响

Fig.8 The effect of random erasure on accuracy

本文通过分类准确率和检测速度衡量模型性能，通过对基础网络增加通道混洗、注意力机制和随机擦除的数据增强，可以看出网络对菜品的分类准确率都有不同程度的提升。从表3中能看出，本文提出的模型相比基础网络，在模型体积上减少18.2%，参数和浮点计算都有相应的减少，在检测时间大致相同的情况下，准确率提高0.84%。本文的模型(Ours)准确率均高于其他网络，通过在数据集FOOD-101上训练和测试，对比其他网络的实验结果，可以得出本文提出的网络模型具有更好的效果。

表3 本文模型与不同模型在数据集FOOD-101上的实验对比

Tab.3 Experimental comparison between the proposed model and other models on the dataset FOOD-101

模型	参数量	浮点数/M	体积/M	准确率 FOOD-101	运行时间
MobileNetV2	2,236,682	318.70	8.53	75.98%	5′ 44″
ShuffleNet	1,825,690	132.41	6.96	69.29%	3′ 51″
SqueezeNet	787,429	771.45	3.00	66.50%	–
GhostNet	3,246,939	216.72	12.39	70.30%	6′ 10″
Ours(MobileNetV2-pro)	1,830,948	246.43	6.98	76.82%	5′ 40″

4 结论(Conclusion)

为了帮助人们在自然环境下更方便地分辨菜品，对菜品图像使用随机擦除方法，提高网络的特征提取能力。新的模型引入了通道混洗及注意力机制，缩减了网络的卷积层，将其命名为MobileNetV2-pro，新的网络体积更小。实验表明，本文网络能更快地提取特征，在菜品分类中表现更好。下一步工作将围绕网络处理更多种类菜品，以增强特征提取能力，提高准确率为主，对网络做进一步改进。

参考文献(References)

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.

[2] 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.

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

[4] 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.

[5] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Computer Society, 2018:4510-4520.

[6] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Computer Society, 2018:6848-6856.

[7] IANDOLA F, HAN S, MOSKEWICZ M, et al. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and<0.5 mbmodel size[C]// ICLR Organizing Committee. ICLR' 17 Conference Proceedings. Toulon: International Conference on Learning Representations, 2017:207-212.

[8] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society, 2017:1251-1258.

[9] 梁峰,董名,田志超,等.面向轻量化神经网络的模型压缩与结构搜索[J].西安交通大学学报,2020,54(11):106-112.

[10] 王韦祥,周欣,何小海,等.基于改进MobileNet网络的人脸表情识别[J].计算机应用与软件,2020,37(04):137-144.

[11] 程越,刘志刚.基于轻量型卷积神经网络的交通标志识别方法[J].计算机系统应用,2020,29(02):198-204.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// CVPR Organizing Committee. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE Computer Society, 2018:7232-7241.

[13] 张翔,史志才,陈良.引入注意力机制和中心损失的表情识别算法[J].传感器与微系统,2020,39(11):148-151.

作者简介：

姚华莹(1997-)，女，硕士生.研究领域：深度学习.本文通讯作者.

彭亚雄(1963-)，男，本科，副教授.研究领域：通信系统.