

基于 CA-EfficientNetV2 的蘑菇图像分类算法研究

孟莉莎, 杨贤昭*, 刘惠康

武汉科技大学信息科学与工程学院冶金自动化与检测技术教育部工程研究中心, 湖北 武汉 430081

摘要 针对传统的蘑菇特征提取方法分类效率低且效果差的问题,提出了一种轻量型的蘑菇图像分类模型。由于实验所用数据集较小,所提分类模型在基于 Imagenet 数据集的迁移学习中初始化 EfficientNetV2 模型并修改全连接层。同时为了减少网络中参数影响,对原 EfficientNetV2 模型进行精简,去除了网络中重复的模块。最后用特征提取效果更好的 coordinate attention (CA) 注意力机制替代原来 MBConv 模块中的 squeeze-and-excitation 机制,得到了新的 CA-EfficientNetV2。实验结果表明:所提 EfficientNetV2 与经典 ResNet50 模型和 RegNet 相比分类准确率分别提高了 10 个百分点和 2 个百分点左右,并得到较高的泛化性能;相较于原始 EfficientNetV2,分类准确率提高了 3 个百分点。即 CA-EfficientNetV2 在蘑菇分类问题上具有更高的准确率,具有较高的分类性能。

关键词 图像处理; 轻量型; EfficientNetV2; coordinate attention; 泛化性能; 分类性能

中图分类号 TP392

文献标志码 A

DOI: 10.3788/LOP202259.2410005

Algorithm on Mushroom Image Classification Based on CA-EfficientNetV2

Meng Lisha, Yang Xianzhao*, Liu Huikang

Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China

Abstract In view of the low efficiency and poor effect of the traditional mushroom feature extraction method, a lightweight mushroom image classification model is proposed. In view of the small dataset used in the experiment, this classification model initializes the EfficientNetV2 model and modifies the full connection layer in the migration study based on the Imagenet dataset. At the same time, in order to reduce the parameter influence in the network, the original EfficientNetV2 model is streamlined to remove duplicate modules in the network. Finally, the squeeze-and-excitation mechanism in the original MBConv module is replaced with the coordinate attention (CA) attention mechanism with better feature extraction effect, and the new CA-EfficientNetV2 network is obtained. The experimental results show that compared with the classical ResNet50 model and RegNet, the classification accuracy of the proposed EfficientNetV2 is improved by about 10 percentage points and 2 percentage points respectively, and higher generalization performance is obtained; compared with the original EfficientNetV2, the classification accuracy is improved by 3 percentage points. That is, CA-EfficientNetV2 has higher accuracy and classification performance in mushroom classification.

Key words image processing; lightweight; EfficientNetV2; coordinate attention; generalization performance; classification performance

1 引言

随着深度学习的不断发展,图像的粗分类已不能满足人们的需求,人们不仅希望能够分辨出图像中的对象是花还是草,而且需要知道其具体种类。细粒度分类就是对属于同一基础类别的图像(如 birds 类、

aircrafts 类、dogs 类等)进行更加细致的子类划分^[1]。随着人们生活水平的提高,对食品原材料的要求也相应提高,不再满足于传统的食材,逐渐倾向于野生食材,比如菌类。与此同时,社会上蘑菇中毒事件层出不穷,人们对蘑菇食品安全越来越重视,因此对蘑菇的准确识别分类就显得至关重要。

收稿日期: 2021-09-17; 修回日期: 2021-10-08; 录用日期: 2021-10-25

基金项目: 国家重点研发计划资助项目(2017YFC0805100)

通信作者: *yangxianzhao@wust.edu.cn

由于存储空间和计算量有限,轻量化网络能在不牺牲精度的前提下减少模型参数个数和计算复杂度,逐步受到深度学习领域的重视。Google 团队先后提出 MobileNetV1^[2]和 MobileNetV2^[3]:前者用深度可分离卷积(depthwise separable convolution)替代普通卷积,在牺牲少量精度的情况下,运算速度得到很大的提升;后者在前者基础上,采用倒残差结构(inverted residual block),即在残差结构中先升维,后降维,使得模型更小,提高了准确率。Zhang 等^[4]提出了 ShuffleNet 模型,该模型通过有序打乱各部分特征图的通道构成新的特征图来解决组卷积引起的信息流通不畅问题,减少了模型的参数量。Li 等^[5]提出了 selective kernel networks(SKNet),该网络在各个 SK 模块内对不同的感受野选择不同尺寸的卷积核,从而捕捉到大小不同目标物体的特征。Xie 等^[6]提出了 ResNext,该网络用分组卷积代替 ResNet 残差块^[7]中的卷积层,既减少了参数量,又提高了准确率,便于训练和移植。Han 等^[8]提出了 GhostNet,该网络使用常规卷积提取到含有少量特征的特征图,之后对其进行恒等映射和简单线性变换操作得出最后的输出特征图。这种方式在减少计算量的同时维持了精度的稳定性。Tan 等^[9]提出了 EfficientNetV1,该网络采用神经架构搜索(NAS)技术来搜索输入图像分辨率、网络深度及通道宽度参数的合理配置。Ji 等^[10]提出 ACNet 模型,构建一种树结构式的注意力模型,通过树层级的学习特性,使得网络学习到由粗到细的特征信息,利用二叉注意力卷积迫使网络注意不同的局部区域,以实现更好的性能。陈珺莹等^[11]基于分层双线性池化网络和显著性检测网络,提出一种显著增强的分层双线性池化方法,通过注意力图与特征提取网络之间的交互增

强显著区域信息来提升分类效果。白瑜颖等^[12]通过双线性注意力池化将 ResNet50 网络和 1×1 卷积提取的特征图和注意力图进行拼接,再利用注意力机制指导改进的混合裁剪数据增强,取得了良好的效果。

以上网络虽然在算法上有很大创新,但是基本上都是在大型数据集上进行训练的。实际生产生活中也存在需要对小批量数据进行分类的情况,这时,较多的网络参数会造成上述网络不能充分训练或过拟合。本文所使用的数据集较小,为减少过拟合带来的影响,提出了一种改进的 CA-EfficientNetV2,采用 EfficientNetV2^[13]作为特征提取网络并通过数据增强的方法,帮助扩展训练集的多样性。随后,对模型进行精简并用 coordinate attention(CA)注意力机制^[14]替换 EfficientNetV2 中的移动倒置瓶颈(MBConv)结构中的 squeeze-and-excitation(SE)注意力模块^[15]。最后,采用改进渐进式训练方法在蘑菇数据集上进行实验。

2 基本原理

2.1 EfficientNetV2 模型

EfficientNetV2 是在 EfficientNetV1 的基础上进行改进的,主要针对 EfficientNetV1 浅层网络 MBConv 模块(图 1)中使用深度(DW)卷积导致速度比较慢这一点。EfficientNetV2 在浅层网络引进 Fused-MBConv 模块,即将原来的 MBConv 结构主分支中 1×1 的升维卷积和 3×3 的 DW 卷积替换成一个普通的 3×3 卷积,然后使用 NAS 技术去搜索两种结构的最佳组合,并非均匀地缩放模型,如图 2 所示。两者具体区别如下:EfficientNetV2 使用了 Fused-MBConv 和 MBConv 的组合;EfficientNetV2 使用较小的 expansion ratio 来减少内存访问开销;EfficientNetV2

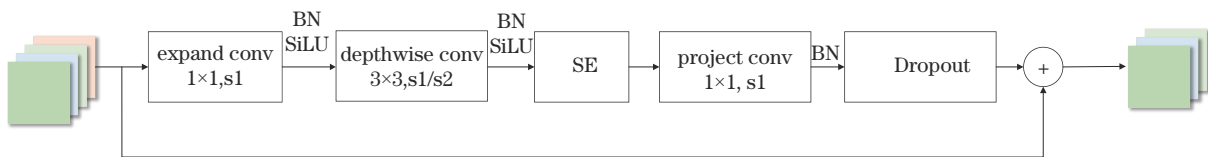


图 1 MBConv 模块

Fig. 1 MBConv module

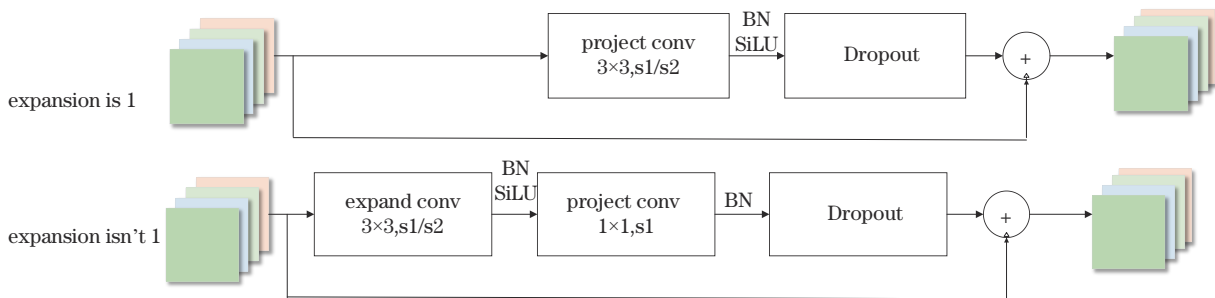


图 2 Fused-MBConv 模块

Fig. 2 Fused-MBConv module

倾向使用更小(3×3)的卷积核,因此需要多层堆叠来增加感受野;除了EfficientNetV1中最后一个步距为1的Stage,因为它的参数数量过多并且内存访问开销过大。

EfficientNetV2-S结构如表1所示。Layers表示Operator中的模块重复堆叠的次数,Channels为经过该Stage所有Layers中的Operator后所输出的特征矩阵的通道数。Operator为在该Stage中使用的模块结构:其中的Conv3×3表示3×3的卷积核加SiLU激活函数和批标准化(BN);Fused-MBConv和MBConv后面跟的数字表示倍率因子 n ,即使输入特征矩阵的通道数扩充 n 倍; $k3\times3$ 表示卷积核大小为3×3,只有expansion ratio不为1时才有expand conv模块。SE0.25表示使用了SE模块且该模块中第1个全连接层的节点个数是输入该MBConv模块特征矩阵通道数的1/4。Stride表示步距,在每个Stage中会多次反复堆叠Operator中的模块,Stride一般默认为1,而仅有第1个Operator模块的步距是严格依照表格中Stride设置的。在图1和2中,当步长为1且输入输出Channels相等时才有shortcut连接。当有shortcut连接时才有Dropout层,而且这里的Dropout层是Stochastic Depth,即会随机丢掉整个block的主分支(只剩捷径分支,相当于直接跳过了这个block),减少了网络的深度。EfficientNetV2-S有8个Stage。

表1 EfficientNetV2-S结构表
Table 1 EfficientNetV2-S structure table

Stage	Operator	Stride	Channels	Layers
0	Conv3×3	2	24	1
1	Fused-MBConv1, k3×3	1	24	2
2	Fused-MBConv4, k3×3	2	48	4
3	Fused-MBConv4, k3×3	2	64	4
4	MBConv4, k3×3, SE0.25	2	128	6
5	MBConv6, k3×3, SE0.25	1	160	9
6	MBConv6, k3×3, SE0.25	2	272	15
7	Conv1×1&Pooling&FC		1792	1

2.2 CA注意力机制

自从SENet出现以来,就以其良好的性能使SEBlock被广泛使用。但是,它的不足之处在于只单纯对通道关系建模来重新确定每个通道的重要性,却严重忽视了因位置信息确定生成空间选择性带来的影响^[16]。对此,本实验组选用了一种能够全面考虑特征图通道和空间的位置信息间关系的CA注意力机制,它可以在基本没有增加额外计算开销的情况下灵活地插入移动网络中,如图3所示。

CA模块依照准确的位置信息分成两个阶段来给通道关系编码。第1阶段为coordinate信息嵌入操作。在对通道注意编码时,一般采用全局池化对空间信息进行全局编码,但这种操作会把全局空间信息压缩到

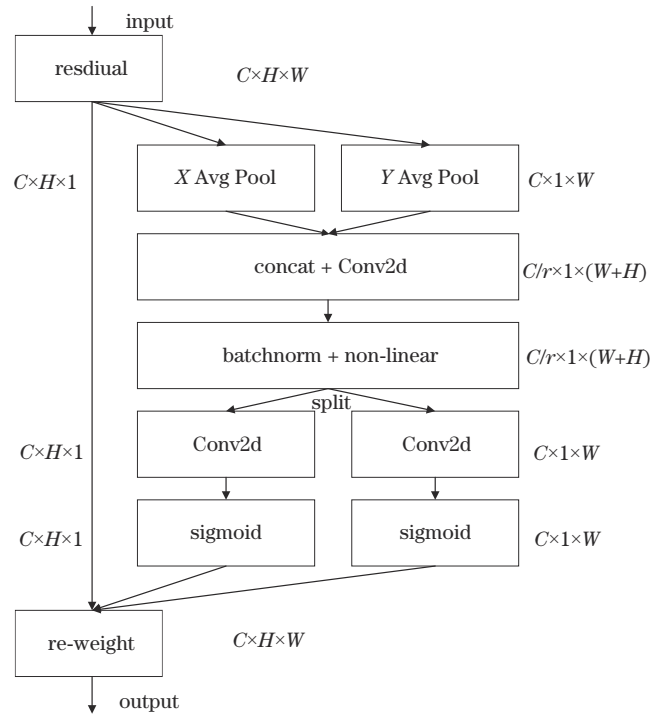


图3 CA模块流程图

Fig. 3 CA module flowchart

通道描述符中,造成位置信息丢失。本实验组把全局池化分解成一对一维特征编码操作,以此来激励注意力模块接收到具有精确位置信息的远程空间交互:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j). \quad (1)$$

具体地,若给定一个输入 \mathbf{X} ,先以尺寸为 $(H, 1)$ 的池化核沿着水平方向或 $(1, W)$ 的池化核沿着垂直方向编码每个通道。因而,高度为 h 和宽度为 w 的第 c 通道的输出分别可以表示为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i), \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w). \quad (3)$$

该变换方式沿着两个不同的空间方向来聚合特征,从而生成一对方向感知的特征图,能够得到全局感受野并对位置信息进行精确编码。第2阶段为CA注意力生成操作。先将第1阶段的变换结果进行concatenate操作,之后利用 1×1 卷积变换函数 F_1 对其进行变换操作:

$$f = \delta \{ F_1 [z^h, z^w] \}, \quad (4)$$

式中: $[\cdot, \cdot]$ 表示沿空间方向连接操作; δ 表示非线性激活函数; f 是分别在水平和垂直方向对空间信息进行编码的中间特征映射。 r 表示缩减率,可以用来控制SEblock大小,即可以通过控制 r 来减少 f 的通道数,使得模型更加精简并减少算力。随后,分别沿空间方向将 f 分解为两个独立的张量 $f^h \in \mathbf{R}^{C/r \times H}$ 和 $f^w \in \mathbf{R}^{C/r \times W}$ 。用其他两个 1×1 卷积变换 F_h 和 F_w 分别

将 f^h 和 f^w 变换为通道数相同的张量到输入 X , 得到输出 g^h 和 g^w :

$$g^h = \sigma[F_h(f^h)], \quad (5)$$

$$g^w = \sigma[F_w(f^w)], \quad (6)$$

式中: σ 为 Sigmoid 激活函数。接着分别对 g^h 和 g^w 进行扩展并作为注意力权重。最后的输出 Y 可以表示为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j). \quad (7)$$

2.3 CA-EfficientNetV2 模型

与之前的网络相比, EfficientNetV2 的优点在于有更高的训练速度和参数效率, 但是由于网络参数很多, 每个 Fused-MBConv 和 MBConv 内部结构复杂, 对小样本来说并不太实用, 计算量相对依然很大, 对计算机硬件设备要求高。本实验组对原 EfficientNetV2 进行以下改进: 1) 精简原 EfficientNetV2。去除原来网络中重复堆叠的 Fused-MBConv 和 MBConv 模块, 只保留每个 Stage 中的第 1 个 Layer, 即 3 个 Fused-MBConv 和 3 个 MBConv。原 EfficientNetV2 参数较多, 网络结构较为精细, 因此耗费的算力资源也多, 而本研究的数据集较小, 模型无须太大, 对网络进行精简, 保留网络主干框架中最核心的结构就能实现较高的分类结果。2) 将 MBConv 模块中的 SE 注意力机制改进为效果更好的 CA 注意力模块, 以此来提高网络对于重要特征图的注意力, 从而提高模型准确率。

图 4 为改进后的 EfficientNetV2 网络模型总体结构, 网络所包含的子模块的数目较原网络大大减少了。改进后网络的内部结构如图 5 所示: 其中左图为使用

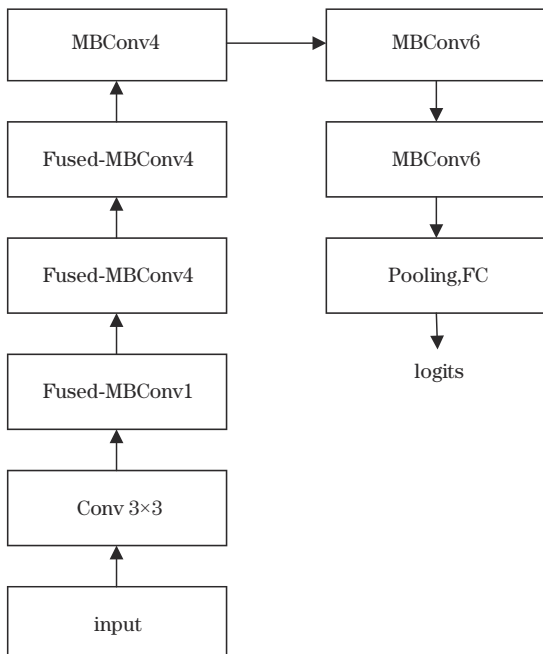


图 4 改进后的 EfficientNetV2 模型总体框架

Fig. 4 Overall framework of the improved EfficientNetV2 network model

了 CA 注意力机制的 MBConv6 模块, 输入先经过一个 1×1 的升维卷积, 通道数扩展 6 倍, 再经过 3×3 的 DW 卷积, 恢复部分特征, 之后经过 CA 模块对特征进行加强, 最后再降维并随机丢弃一些参数; 右图为不含 CA 模块的 Fused-MBConv 内部结构。从改进后的 EfficientNetV2 网络结构中可以看出, 改进后的网络结构与原 EfficientNetV2 相比精简了重复模块, 网络的参数相应减少很多, 计算速度得到提升。

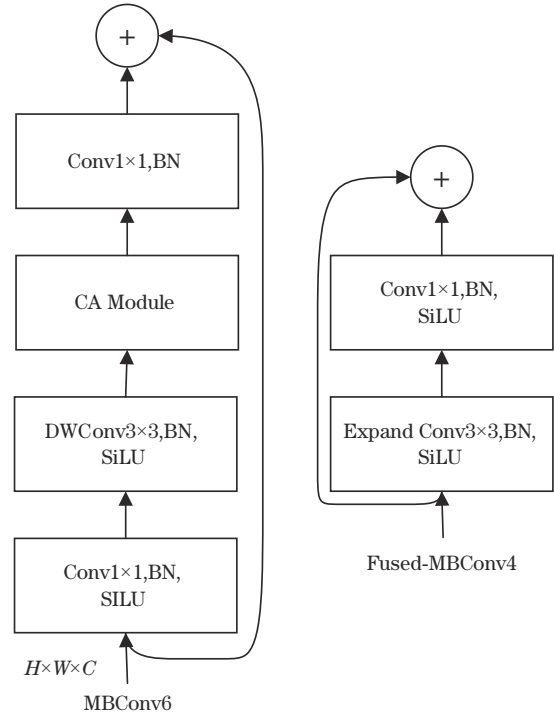


图 5 改进后的 EfficientNetV2 主要模块内部结构图

Fig. 5 Internal structure diagram of main modules of improved EfficientNetV2

3 分析与讨论

实验采用 Python 语言编写网络训练程序, 基于 Pytorch 深度学习框架对模型进行构建和训练, 在 Windows 系统下的 Pytorch 1.7 版本的深度学习框架中进行。处理器采用的是 11th Gen Intel Core i7-11700, 其主频为 2.50 GHz, 内存为 16 GB, 显卡为 NVIDIA GeForce RTX 3080, Cuda 版本为 10.1。

3.1 数据集描述

所采用的数据集来自于 Kaggle 公开蘑菇数据集, 该数据集共有 6707 张图片, 分为 9 个类别。该数据集主要有两个特点: 1) 各类样本数量严重不平衡, 例如乳菇种类数量高达 1563, 而落叶松蕈、绯红湿伞和乳牛肝菌却不到 400; 2) 图片分辨率大小不一致, 包含多种分辨率的图像。采用 EfficientNetV2-S 提取图像特征, 首先通过预处理将图片尺寸统一为 300×384 , 然后使用随机裁剪、颜色扭曲、图像旋转等进行图像增强, 如

图 6 所示,并将数据集按照 6:2:2 的比例划分为训练集、验证集和测试集。

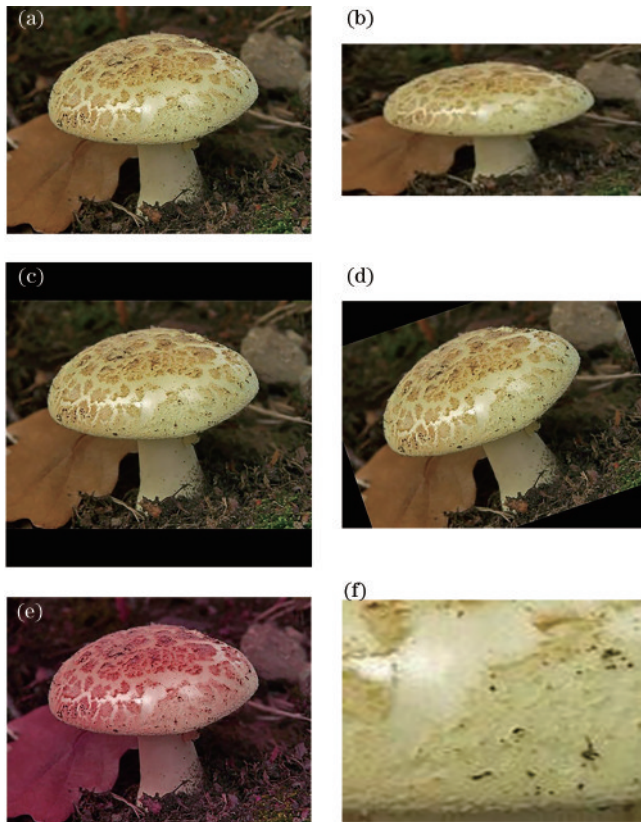


图 6 数据增强结果。(a)原图;(b)随机比例缩放;(c)填充;
(d)随机旋转;(e)颜色抖动;(f)随机剪裁

Fig. 6 Data enhancement results. (a) Original picture; (b) random resize; (c) padding; (d) random rotation; (e) color jitter; (f) random crop

3.2 训练策略

采用了一种新颖的渐进式训练策略:在早期进行训练时,选用尺寸较小的图像和弱正则化方式进行训练来学习简单的特征,随后逐步增大图像尺寸并增强正则化。基于渐进式调整大小,但通过动态调整正则化,可以加快训练速度,而不会导致准确率下降。具体训练步骤如下:

- 1) 输入蘑菇图像,其形状为 $[N, H, W, C]$,并进行预处理;
- 2) 搭建 EfficientNetV2-S 特征提取网络;
- 3) 加载 ImageNet 预训练参数;
- 4) 对模型进行精简并加入 CA 注意力机制重构网络;
- 5) 训练网络;
- 6) 在训练集损失函数最低处保存网络模型;
- 7) 输出图像分类结果。

3.3 实验结果与分析

实验时,先去除 EfficientNetV2 最后一层全连接层,然后使用在 ImageNet 数据集上训练完成的 EfficientNetV2-S 权重将其初始化,并对模型的相关参

数进行微调,包括将分类类别数目改为 9、利用随机赋值的方法对最后一层的参数进行初始化并训练。实验具体参数设置如表 2 所示。

表 2 实验参数设置

Table 2 Experimental parameter settings

Experimental parameter	Scene setting
Activation function	SiLU/Sigmoid
Learning initial rate	0.01
Learning decayl rate	0.01
Drop_connect_rate	0.2
Loss function	Cross entropy
Momentum	0.9
Weight_decay	1×10^{-4}
Optimization	SGD
Batch size	32
Number of iterations	150
Dropout rate	0.2
Activation function	Softmax

表 2 中的 Drop_connect_rate 表示 MBConv 结构中 Dropout 层使用的随机丢弃率,而 Dropout_rate 为整个网络结构中最后一个全连接层前的 Dropout 层的随机失活率。改进后的 EfficientNetV2 结构参数如表 3 所示。

表 3 改进后的 EfficientNetV2 模型结构参数

Table 3 Structural parameters of improved EfficientNetV2 model

Stage	Kernel size /pooling size	Output
Input		$224 \times 224 \times 3$
Conv	3×3	$112 \times 112 \times 24$
Fused-MBConv	3×3	$112 \times 112 \times 24$
Fused-MBConv	3×3	$56 \times 56 \times 48$
Fused-MBConv	3×3	$28 \times 28 \times 64$
MBConv	3×3	$14 \times 14 \times 128$
MBConv	3×3	$14 \times 14 \times 160$
MBConv	3×3	$7 \times 7 \times 272$
Ave-pooling	3×3	$7 \times 7 \times 272$
Softmax	1×1	$1 \times 1 \times 9$

为了验证 EfficientNetV2 的性能,分别在蘑菇数据集上与 ResNet50 和 RegNet^[17]进行了实验对比,准确率和损失函数仿真对比结果分别如图 7 和图 8 所示。

从图中可以看出:这 3 个网络在 100 个 epoch 左右逐渐趋于平稳,EfficientNetV2 由于采用了更高效网络特征提取模块,其在验证集的准确率相比于 ResNet50 和 RegNet 分别提高了 8 个百分点和 2 个百分点左右;损失也比另外两个网络低。这说明了 EfficientNetV2

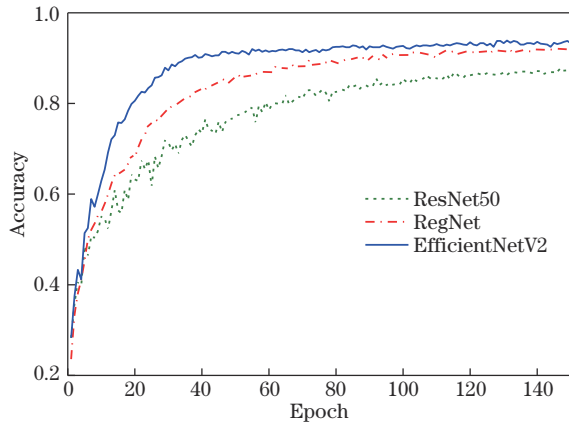


图7 EfficientNetV2与其他网络准确率对比曲线

Fig. 7 Accuracy comparison curve of EfficientNetV2 and other networks

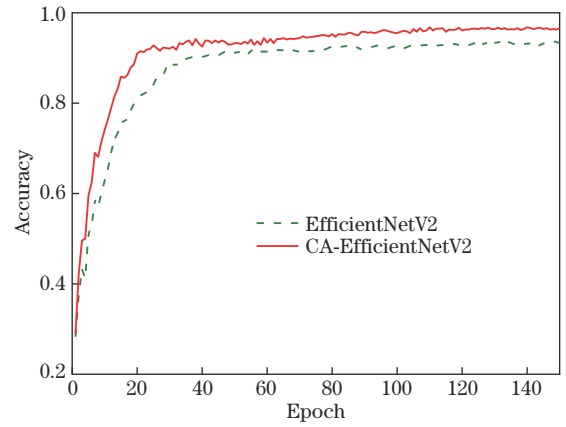


图9 改进前后 EfficientNetV2 准确率对比曲线

Fig. 9 Accuracy comparison curve of EfficientNetV2 network before and after improvement

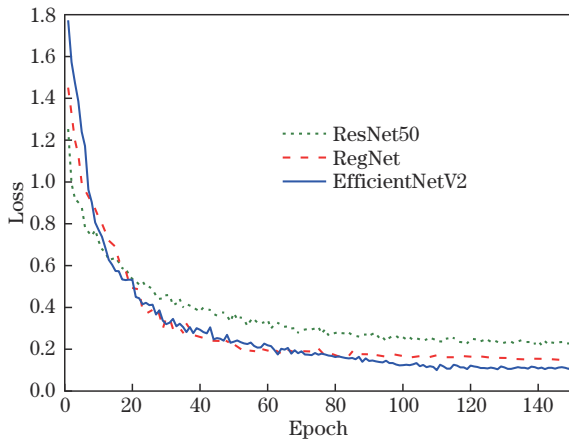


图8 EfficientNetV2与其他网络损失函数对比曲线

Fig. 8 Loss function comparison curve of EfficientNetV2 and other networks

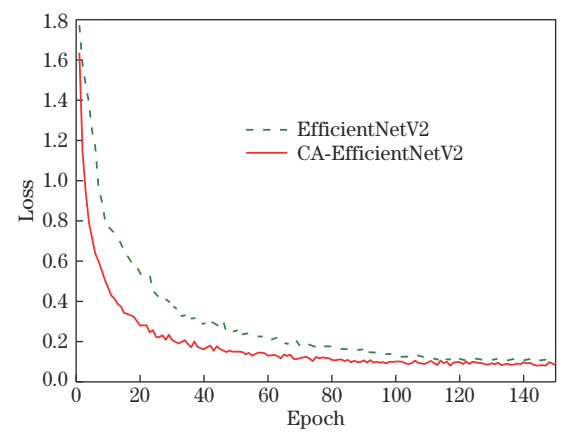


图10 改进前后 EfficientNetV2 损失函数对比曲线

Fig. 10 Loss function comparison curve of NetworkNetV2 network before and after improvement

在蘑菇分类数据集上的有效性。

为了证明所提改进的 CA-EfficientNetV2 的分类效果,对比了 EfficientNetV2 和 CA-EfficientNetV2 在验证集上的准确率和训练集的损失函数曲线,仿真结果如图 9 和 10 所示。从仿真结果可知,加入 CA 注意力机制后网络的识别精度明显高于原网络,且在前期训练速度也高于未改进的 EfficientNetV2。仿真结果表明,所设计的蘑菇分类方法有更好的识别效果。

为了更进一步说明 CA 注意力机制的有效性,采

用 Grad-CAM^[18]对结果图进行可视化,如图 11 所示。Grad-CAM 是一种利用网络反向传播的梯度计算出 feature map 每一个通道的权重从而得到热力图的方法。图 11 (a) 为输入原始图像,图 11 (b) 为 EfficientNetV2 最后一层特征层可视化结果,图 11 (c) 为加入 CA 注意力后的可视化结果。从图 11 可以发现,加入注意力机制后,特征提取更为精确,且该图像所属类别的置信度更高。这进一步说明了 CA 注意力机制在 EfficientNetV2 上的有效性。

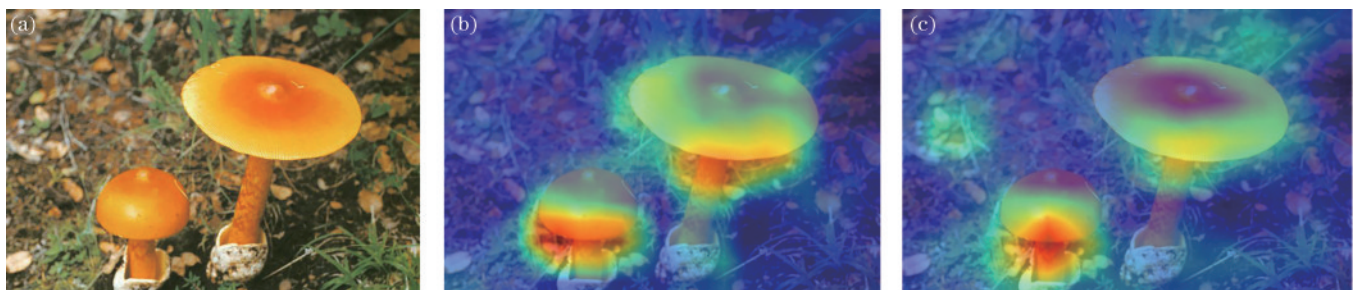


图 11 图像可视化结果对比图。(a)输入;(b) EfficientNetV2;(c) CA- EfficientNetV2

Fig. 11 Comparison of image visualization results. (a) Input; (b) EfficientNetV2; (c) CA- EfficientNetV2

最后,对比了各模型在测试集上的分类准确率,结果如表 4 所示。从表 4 可以看出,所提改进 CA-EfficientNetV2 模型在蘑菇上的识别精度比原 EfficientNetV2 有了很大提升。和其他模型识别方法相比,所提方法对蘑菇的识别效果也优于其他识别方法。这说明了所提蘑菇分类方法的有效性。

表 4 不同模型分类准确率对比
Table 4 Comparison of classification accuracy for different models unit:%

Model	Accuracy
ResNet50	86.7
RegNet	91.9
EfficientNetV2	93.8
CA-EfficientNetV2	96.8

4 结 论

简要介绍了经典的轻量化网络,对 EfficientNetV2 模块结构进行了研究并分析了它的优缺点,在 Pytorch 框架中对其进行改进。先简化原有的复杂网络模型,之后用提取特征效果更好的 CA 模块来搭建蘑菇分类网络模型,采用在不同训练阶段使用不同的训练尺寸以及正则化强度的方法的训练策略优化网络模型。与 ResNet50 和 RegNet 的对比实验表明了 EfficientNetV2 具有较高的准确率,适合轻量化部署。将改进后的网络与 EfficientNetV2 在蘑菇数据集上进行仿真对比,实验结果表明,CA-EfficientNetV2 较 EfficientNetV2 提高了 3 个百分点的准确率,且计算速度并未降低。所提 CA-EfficientNetV2 能够增强网络的特征提取能力,提高了分类准确率,对未来实现蘑菇分类自动化分拣有一定现实意义。

参 考 文 献

- [1] 罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, 43(8): 1306-1318.
Luo J H, Wu J X. A survey on fine-grained image categorization using deep convolutional features[J]. Acta Automatica Sinica, 2017, 43(8): 1306-1318.
- [2] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-01-12]. <https://arxiv.org/abs/1704.04861>.
- [3] 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, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [4] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on

Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.

- [5] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 510-519.
- [6] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.
- [7] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [8] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1577-1586.
- [9] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-04-05]. <https://arxiv.org/abs/1905.11946>.
- [10] Ji R Y, Wen L Y, Zhang L B, et al. Attention convolutional binary neural tree for fine-grained visual categorization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10465-10474.
- [11] 陈珺莹, 陈莹. 基于显著增强分层双线性池化网络的细粒度图像分类[J]. 计算机辅助设计与图形学学报, 2021, 33(2): 241-249.
Chen J Y, Chen Y. Saliency enhanced hierarchical bilinear pooling for fine-grained classification[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(2): 241-249.
- [12] 白瑜颖, 刘宁钟, 姜晓通. 结合注意力混合裁剪的细粒度分类网络[J]. 计算机技术与发展, 2021, 31(10): 38-42.
Bai Y Y, Liu N Z, Jiang X T. Fine grained image classification network combined with attention CutMix [J]. Computer Technology and Development, 2021, 31(10): 38-42.
- [13] Tan M X, Le Q V. EfficientNetV2: smaller models and faster training[EB/OL]. (2021-04-01)[2021-05-08]. <https://arxiv.org/abs/2104.00298v1>.
- [14] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [15] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(8): 2011-2023.
- [16] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient

- channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [17] Radosavovic I, Kosaraju R P, Girshick R, et al. Designing network design spaces[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10425-10433.
- [18] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.