

蔡润基,江方湧,郑涛涛,等.深度模型融合数据合成机制的长尾目标识别[J].华中农业大学学报,2023,42(3):271-280.
DOI:10.13300/j.cnki.hnlkxb.2023.03.031

深度模型融合数据合成机制的长尾目标识别

蔡润基,江方湧,郑涛涛,刘东霖,徐初东

华南农业大学电子工程学院(人工智能学院),广州 510642

摘要 为解决昆虫数据集数据分布不均衡导致识别模型在数据量少的尾部类别的识别性能差的问题,提出1种融合数据合成的卷积识别网络模型(synthetic samples combined model,SSCM),该模型包含图像分割与重组模块、主干网络模块和数据纠正分支模块等3个模块。通过图像分割与重组模块对训练的图片进行分割并重组,得到新的训练数据并加入训练集;再使用ResNet-50作为网络主干提取图片的特征,同时数据纠正分支模块采用均方误差与交叉熵计算合成图像与原图像之间的误差,以减少合成图像对尾部数据的不利影响。构建包含300个蝴蝶类别共26 045张图片的数据集验证模型性能,结果显示,SSCM模型在该数据集上的准确率较DRC、BBN、RIDE等主流长尾目标识别模型分别高3、2.14、2.71个百分点。采用公开昆虫数据集IP102进一步验证SSCM算法的有效性,结果显示,SSCM模型准确率比DRC、BBN、RIDE等模型分别高18.94、3.02、3.36个百分点。

关键词 昆虫识别;数据合成;图像分割;农业监测;长尾识别;生态环境监测;生物多样性保护;害虫防治

中图分类号 TP391.41; TP311.13 **文献标识码** A **文章编号** 1000-2421(2023)03-0271-10

蝴蝶(*Rhopalocera*)是自然界种类最丰富的昆虫之一^[1]。由于多数蝴蝶对其生存的自然环境要求高,因此常被用来评判一个地区的生态环境。蝴蝶种类繁多,现有的蝴蝶鉴别主要以人工鉴别和分类为主,耗时耗力,且高度依赖专家的长期经验积累。鉴于生态环境监测、生物多样性保护和害虫防治等领域的需要,开展蝴蝶等昆虫自动识别十分必要^[2]。

近年来,随着深度学习的发展,卷积神经网络被成功应用到自动识别领域。从1990年代的LeNet开始,卷积神经网络经历了AlexNet^[3]、VGGNet^[4]、GoogLeNet^[5]、MobileNet^[6]、EfficientNet^[7]等一系列模型的推陈出新,在常规分类数据集如鸟、柑橘、鱼等方面取得不错的分类效果^[8-9]。由于物种的种类在自然界分布十分不均匀^[10],特定类别难以收集,导致收集到的数据集的数据往往呈现不平衡的分布,也称长尾分布。而上述模型大多基于均衡的数据集,在将其应用于长尾数据时,由于模型中头部类别的样本量远多于尾部类别的样本量,易导致模型在头部类别(多样本类别)过拟合、尾部类别(少样本类

别)欠拟合,进而影响模型的整体识别性能。为提升长尾识别性能,当前一般通过重采样^[11-13]和重加权策略^[13-15],对尾部的少样本类别进行过采样或者添加1个更大的权重,以改善尾部识别性能。重采样策略针对尾部数据不足的问题主要通过2种方式解决:一是少样本类别的过采样^[11],二是对多样本类别的欠采样^[12],2种方式均提高了模型整体的识别性能。但过采样会导致少样本类别的过拟合,无法学习到更鲁棒易泛化的特征,欠采样则会导致欠拟合情况的发生。在重加权策略上,Kang等^[13]提出的解耦合训练(decoupling representation and classifier,DRC),直接利用原始数据训练分类模型,再单独对分类器通过在损失上进行重加权的方式以提升尾部识别性能。为了进一步提升模型性能,Zhang等^[14]先通过调整训练集的分类分布对齐到重加权得到预测结果的方式,更有效地降低重加权策略的误差。以损失函数角度对尾部类别进行重加权,平衡了分类间距,但在一定程度上损害深度网络学习到的特征,导致每个类别的内部分布变得更加分散,对头部数据的性

收稿日期:2022-09-02

基金项目:广东省自然科学基金项目(2020A1515010634)

蔡润基,E-mail:13600374676@163.com

通信作者:徐初东,E-mail:cd79cd@126.com

能产生负面影响。上述方法往往通过牺牲头部识别性能以换取尾部识别性能的提升,在总体性能上虽然获得了提升,但上述方法对尾部类别识别易存在过拟合,仍是长尾识别亟需解决的难题。

在生态监测领域,目前与害虫有关的数据集较少,且公开的害虫数据集存在类别少^[16]和类内数据少^[17]的问题。针对这个问题,本研究收集了1个包含300种的蝴蝶数据集,由于数据集里特定类别的蝴蝶难以收集,导致该数据集的类别数据呈现长尾分布,并存在类内差异大、类间差异小的细粒度识别问题。针对上述问题,本研究提出融合数据合成的卷积识别网络模型(synthetic samples combined model, SSCM),通过数据合成策略提升尾部数据以进一步缓解尾部类别过拟合情况,并且在蝴蝶数据集和公开的IP102数据集^[18]进行分析验证,旨在为农业监测和蝴蝶识别等领域提供新的思路。

1 材料与方法

1.1 SSCM模型

SSCM识别模型主要由用于数据预处理的图片分割与重组模块、基于ResNet-50的主干特征提取模

块以及数据纠正分支模块3部分组成,整体框架如图1所示。在分割重组阶段,输入的图片会被平均分割为4份,并打乱重组,再将重组得到的合成图像输入到网络中进行训练,以使网络学习到的特征量达到特征重用。主干网络模块基于ResNet-50,该模块由4个特征提取层(Layer1、Layer2、Layer3、Layer4)和全连接层组成。此外,构建数据纠正分支模块以计算分割重组模块产生的新数据与原始图像数据之间的误差,并与主干网络共享特征层,以增强不同层次的特征利于共同学习,同时进一步缓解尾部类别过拟合。在识别阶段,对分支模块和主干模块的输出结果,采用平衡加权的分类策略以增强模型的性能。

1.2 图片分割与重组模块设计

受到基于局部特征进行数据深度学习的方法启发^[19],本研究设计了1个图片分割与重组模块以提升数据分布的平衡性,模块结构如图2所示。在图片输入网络之前,通过将图片平均切割为4份小图像,以使网络可以将注意力集中到图片的不同局部特征当中,并且将切割好的图片打乱重组输入到网络中训练,实现全局特征与局部特征的融合。

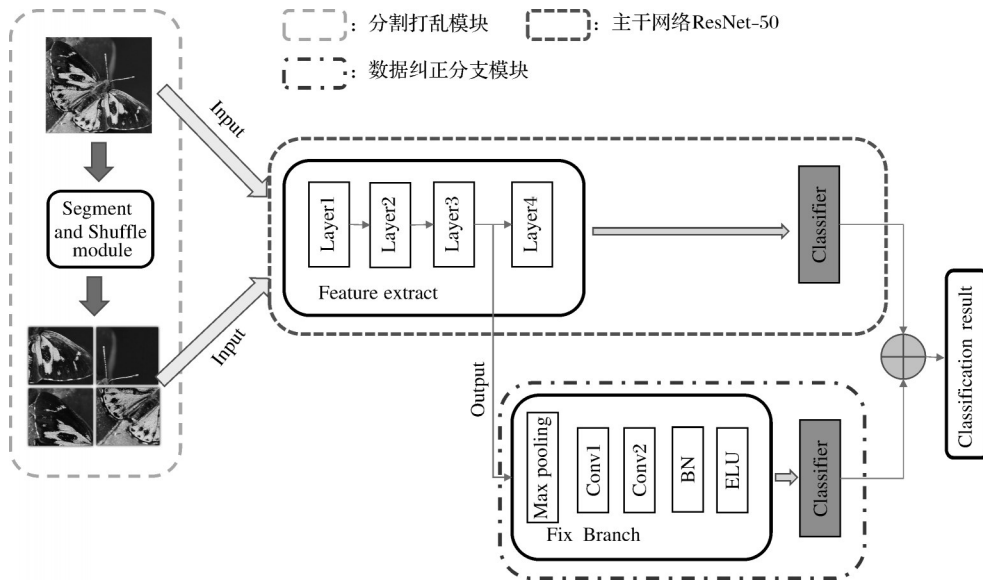


图1 SSCM模型蝴蝶识别网络结构

Fig.1 Butterfly recognition network structure of SSCM model

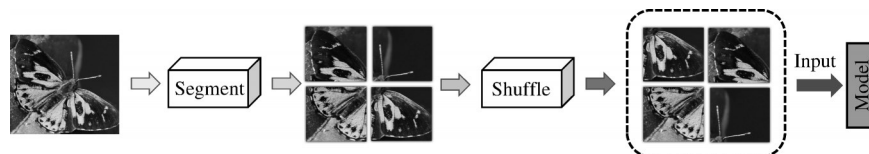


图2 图片分割与重组模块

Fig.2 Segment and shuffle module

对于输入大小为 224 像素×224 像素的 RGB 蝴蝶图像,使用该模块将输入图片切割为大小 112 像素×112 像素的分块,再将分割好的图片随机打乱重组输入到网络中训练。由于数据集的分布是不平衡的,受到DRC^[13]的启发,将类别下 30 张以下的图片定义为少样本类,100 张以上的图片定义为多样样本类。相对于多样样本类,在少样本类中提升数据量,相较于多样样本类可以更显著提升模型在尾部数据的识别性能。在网络训练阶段,通过该模块为训练集里的每张图片生成 1 张对应的重组图片(生成的图片不会保存到原数据集中),这种方式可以增加训练集的规模,同时不影响原训练集的分布。此外,少样本类

通过该模块获得了 1 倍的训练数据,提升了网络在少样本类中学习到的特征数。

1.3 主干网络模块

SSCM 模型选择 ResNet-50^[20]作为模型的主干网络,以高效地提取图片特征。模型结构如图 3 所示。输入首先通过 1 个 7×7 的卷积层,接着通过池化层,紧接着经过 16 个堆叠的残差块以学习图像的特征,再经过全连接层得到最终输出。为了适合数据集训练,需要将 ResNet-50 最后的原全连接层修改为适应试验数据集类别的全连接层,改动的全连接层不加载预训练参数,剩余的网络层均加载了预训练参数,以加快训练时间并进一步提升模型性能。

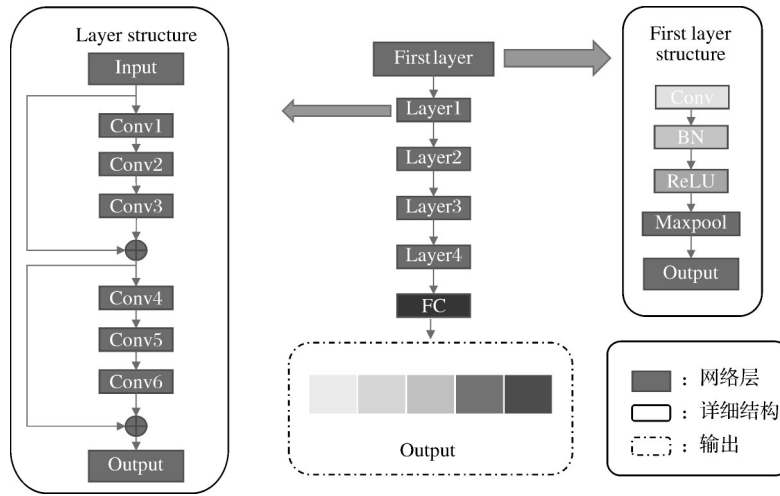


图 3 ResNet-50 结构

Fig.3 Structure of ResNet-50

1.4 数据纠正分支模块

由于生成的新图像与原图像存在特征位置、目标边界等不同,由此学习到的特征与原图像存在一定的误差。针对这个问题,受到 Hariharan 等^[21]的启发,设计了 1 个数据纠正分支模块,如图 4 所示。该

模块与主干网络模块共享 Layer4 前面网络层的特征,并采用与主干网络不同的损失网络来判别分割模块得到的新数据与原数据的差异,再通过共享特征层,提升总体的分类能力。

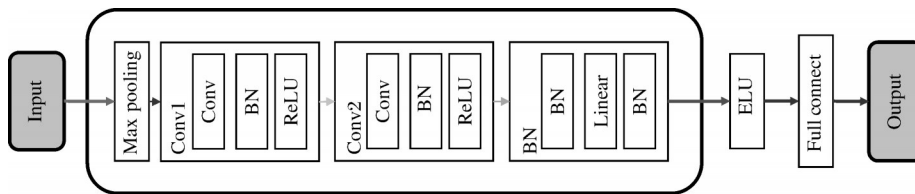


图 4 数据纠正分支模块

Fig.4 Data fix branch

改进的数据纠正分支模块由 2 个卷积层和 1 个全连接层组成,2 个卷积层采取不同大小的卷积核以学习与原来主干网络不同的深层特征。第一层的卷积核大小为 1,第二层的卷积核大小为 3,并且使用间隔大小为 1 的空间卷积以避免使用下采样,进而在相

同计算量的前提下,提供更大的感受野。为了防止梯度弥散,增强模型的非线性能力,减少过拟合,卷积层后均使用了 ReLU 激活函数^[22],如公式(1)所示。为了达到更好的分类效果,数据纠正分支模块的分类层前使用了 ELU 激活函数,ELU 激活函数有

ReLU函数基本上所有的优点,如公式(2)所示,当节点输入为负时,网络仍然有输出值,这样可以避免ReLU函数中某一节点死亡的问题。

$$\text{ReLU} = \max(0, x) \quad (1)$$

$$\text{ELU}(x) = \begin{cases} x, & x > 0 \\ a(e^x - 1), & x \leq 0 \end{cases} \quad (2)$$

为了取得不同能力的分类器,模型在全连接层上进行不加载预训练参数的操作,并且使用标准初始化操作分别对主干网络和设计的分类器的全连接层进行参数初始化,以确保2个分类器在开始训练阶段拥有不同的分类能力,初始化函数是恺明初始化(Kaiming normal initialization)^[23],服从 $N=(0, \text{std})$ 的分布, std的方程如下:

$$\text{std} = \sqrt{\frac{2}{(1 + a^2 + (\text{fan} - \text{out}))}} \quad (3)$$

该初始化方式不需要计算全连接层的输入维度,默认将参数 a 的值设置为1。fan-out也是该层的输出维度。在2个分类器有不同的起点后,进行优化微调模型中所有层的参数,进行更深层的特征提取学习。

1.5 损失函数改进

主干网络使用的损失函数为分类任务中常用的交叉熵(CE loss),该损失函数可以更好地描述2个概率分布之间的距离,公式表示为:

$$L_b = -\sum_i^N y_i' \log(y_i) \quad (4)$$

其中, y' 和 y 分别表示分类器对第 i 个实例的预测结果和真实的结果。

在训练的过程中,由于设计了数据纠正分支模块,对该模块的输出设计了新的损失函数以进行误差计算。在数据纠正分支模块使用以下损失函数,公式表示为:

$$L_c = \lambda L_{\text{MSE}}(y_i', y_i) + L_b \quad (5)$$

其中 λ 是1个固定值,训练设置为0.5, L_{MSE} 表示均

方误差(mean squared error),使用该损失是为了表示生成图片与原图片的差异。在测试阶段,对于每个分类器,结果计算如下:

$$R_i = \arg \max(c^i) \quad (6)$$

式(6)中, i 表示第 i 个分类器, R_i 表示第 i 个分类器的结果, c^i 表示所有实例经过分类器得到预测结果的集合,然而具有特定粒度信息的单个网络的预测是唯一的且在2个网络预测时是互补的,当简单地将2个网络的输出以相等的权重组合在一起时,这会产生更好的性能。组合预测 R_A 可以写为:

$$R_A = \arg \max(c^e + c^b) \quad (7)$$

式(7)中, c^e 表示设计的数据纠正分支模块的分类器, c^b 表示主干网络的分类器,经过组合计算得到更好的分类结果。

1.6 蝴蝶数据集

与华南农业大学植物保护学院蝴蝶专家王敏教授团队合作开展本研究项目,构建蝴蝶数据集,该数据集包含蝴蝶300种,一共53 567张。由不同地方、季节、天气环境下的各种姿态的生态图片和标本图片组成,涵盖了我国境内的大多数蝴蝶类别。其中,蝴蝶的标本照片主要收集于互联网和中国蝶类志等昆虫学专著。生态照片主要来源于数码相机野外实拍和互联网资源,如Google、百度等。在进行数据采集以及初步处理后得到的原始数据集蝴蝶图片中,蝴蝶主体在图像中占比较小,而周围的背景环境占比较大。为了去除复杂的背景,构建了目标检测数据集,且使用目标检测技术在原始图片中提取出蝴蝶图像数据构成实验所用的蝴蝶识别数据集。图5A是最终数据集的部分样本,每一行对应1个类别,可以看出其存在类间差异小(第二行与第三行)和类内差异大(每一行)的特点。经过人工筛选和专家反复确认标签的准确性,最终整理得到的蝴蝶数据集共包含26 045张图片,训练集、验证集、测试集

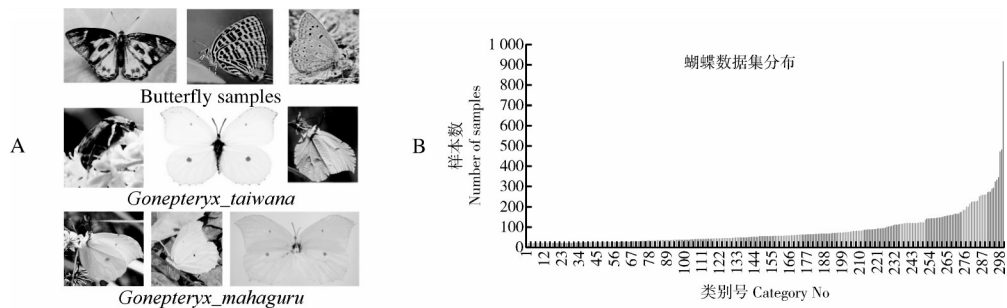


图5 蝴蝶数据集样本示例(A)与样本分布情况(B)

Fig.5 Instances(A) and distribution(B) of butterfly dataset

根据6:1:3的比例将15 627张图片分到训练集,划分到验证集有2 605张,测试集划分到的图片为7 813张。

1.7 IP102昆虫数据集

为了进一步推进昆虫识别研究,Wu等^[18]构建了包含102种害虫的长尾数据集,如图6所示。图6A

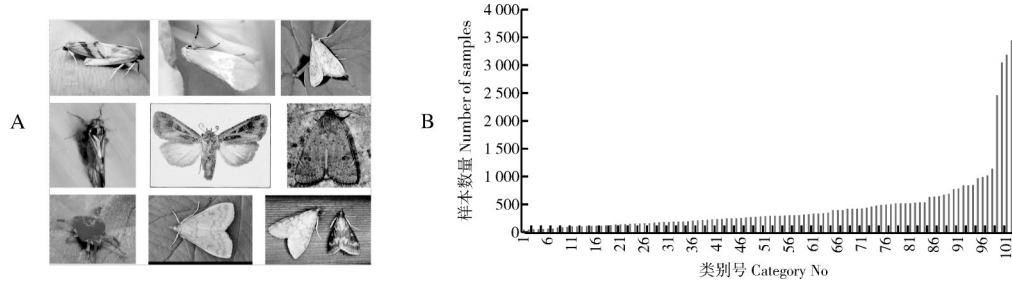


图6 IP102数据集样本示例(A)与样本分布情况(B)

Fig.6 Instances(A) and distribution(B) of IP102

1.8 试验环境与数据预处理

本研究试验均在Ubuntu 16.04系统下完成,基于深度学习框架PyTorch1.11.0和编译器Python 3.7实现,硬件环境:CPU为Intel[®] Xeon(R) E5-2687w CPU[®] 3.10 GHz, 64 G运行内存, GPU为NVIDIA RTX 3090,显存大小为24 GB,加速库为CUDA 11.3。在真实场景采集的蝴蝶和IP102昆虫数据集里,图像中目标的大小、角度、姿势各有不同,为便于模型学习,在此把图片统一缩放或者扩大到224像素×224像素。此外,为优化模型学习效率,本研究采用在ImageNet数据集上的预训练参数。为了避免过拟合,在输入网络训练之前将数据集的所有图像进行随机水平翻转,随机翻转概率为0.5。为了限制处理后的图像数据值处于一定范围内,归一化采用在ImageNet数据集上得到的参数。

1.9 试验设置

模型训练采用随机梯度下降法(stochastic gradient descent, SGD)训练,初始学习率设为0.002,动量为0.9,权重衰减为0.000 05,在蝴蝶数据集和IP102昆虫长尾数据集上均迭代90次。由于不同数据集样本的差异性,在蝴蝶数据集上的批处理大小为32,在第20轮将学习率调整为0.001,在第30轮将学习率调整为0.000 5,在第50轮时将学习率调整为0.000 1。在IP102昆虫长尾数据集上批处理大小为16,每30轮学习率以0.1的衰减率衰减。

1.10 评估指标

为了验证模型的性能,使用准确率(accuracy, A)

的图片来自9个不同类别,IP102数据集也存在类间差异小的问题(最后2张图的黄色虫子)。IP102数据集一共包含75 222张图片,根据原论文提供的划分文档将45 095张图片分到训练集,划分到验证集有7 508张,测试集划分到的图片为22 619张,训练集、验证集、测试集的比例为6:1:3。

作为评价标准,并且准确率从测试集中得到。为得到更真实的性能评估,测试集与训练集图片相互独立,准确率(accuracy, A)如公式(8)所示。

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + T_N} \quad (8)$$

其中, T_P 为真实的正样本数量, F_P 为虚假的正样本数量, F_N 为虚假的负样本数量, T_N 是真实的负样本数量。

为了进一步比较模型之间的性能,再引入精确率、召回率、 F_1 分数这三个评价指标。其中精确率(precision, P)如公式(9)和(10)所示。

$$P_k = \frac{T_{P_k}}{T_{P_k} + F_{P_k}} \quad (9)$$

$$P = \frac{1}{n} \sum_k^n P_k \quad (10)$$

在公式(9)中, T_{P_k} 表示第 k 个类别的真实的正样本数量, F_{P_k} 表示第 k 个类别的虚假的正样本数量,公式(10)通过先计算出每个类别的精确率,再对它们取平均获得多分类的精确率。召回率(recall, R)如公式(11)和(12)所示。

$$R_k = \frac{T_{P_k}}{T_{P_k} + F_{N_k}} \quad (11)$$

$$R = \frac{1}{n} \sum_k^n R_k \quad (12)$$

在公式(10)中, F_{N_k} 表示第 k 个类别的虚假的负样本数量,先计算出每个类别的召回率,再对它们取平均获得多分类的召回率。 F_1 分数的计算公式如(13)所示。

$$F_1 = \left(\frac{1}{n} \sum_k^n \frac{2 \cdot P_k \cdot R_k}{P_k + R_k} \right)^2 \quad (13)$$

通过公式(13)计算得到最终的 F_1 分数,其是精确率和召回率的调和平均数。

2 结果与分析

2.1 蝴蝶数据集试验结果

将本研究提出的SSCM模型与DRC^[13]、BBN^[24]、RIDE^[25]等3个目前主流的长尾识别模型在蝴蝶数据集上进行试验对比,训练的准确率曲线如图7所示。选择的基线方法是标准的CNN模型

ResNet-50。模型训练需要的迭代次数不同,其中DRC、BBN、RIDE等3个长尾模型迭代100次,单网络CNN模型ResNet-50和SSCM模型迭代90次,所有模型均训练到结果曲线趋向平稳,然后取最后的90次迭代进行对比分析。试验结果表明,SSCM模型一开始就拥有比较高的识别性能,而且相对于长尾模型,SSCM模型在不改变数据集分布的情况下增加了数据量,由此得到的结果曲线更加平稳,在对比DRC和RIDE的曲线时更加明显。

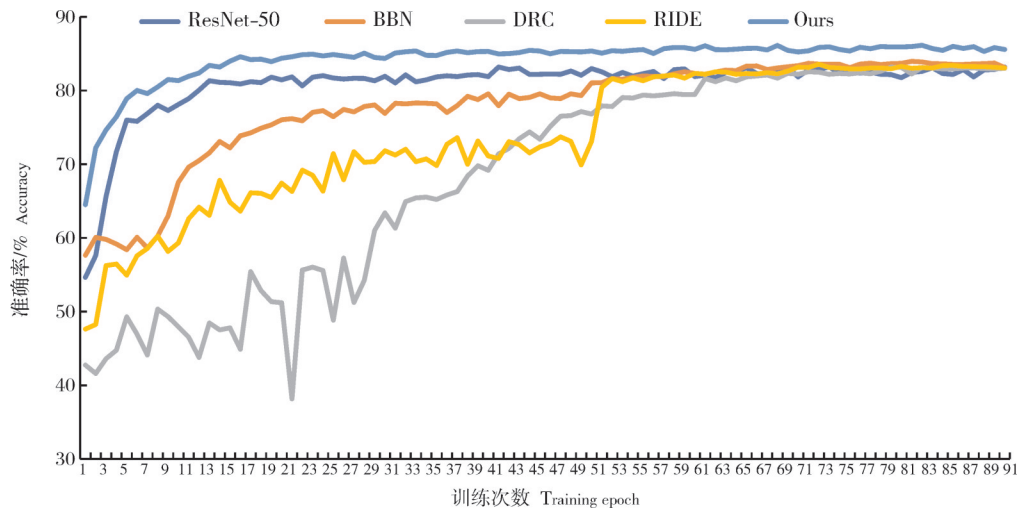


图7 蝴蝶数据集上的训练曲线

Fig.7 The training curve of models in the butterfly dataset

表1为各种模型在第一准确率指标上的试验结果。由表1可知,在建立的300种蝴蝶数据集上,本研究建立的SSCM模型相比标准的REResNet-50以及其他的长尾识别方法的识别性能更好。对比单网络的ResNet-50,SSCM模型的识别准确率提升了3.3个百分点。长尾分布下的识别模型DRC和BBN都是通过类别再平衡方式提升整体数据集的识别准确率。后续的试验结果显示,在蝴蝶数据集上这2个模型会不同程度地削弱头部类数据的性能,从而导致整体数据集的识别精度下降,2个模型的识别准确率分别为83.1%、83.96%,对比之下,较SSCM模型的识别准确率分别低3和2.14个百分点。通过召回率、精确率、 F_1 分数等评估指标的比较,在蝴蝶数据集上,SSCM拥有更好的模型稳定性。这是由于SSCM模型通过数据合成的方式,在不改变数据集总体分布的情况下,有效地提升了小样本数据的特征量,而且没有损失头部性能,更好地提升了总体性能。对比RIDE方法,由于蝴蝶数据集和IP102昆虫数据集带有细粒度识别的特点,而SSCM模型通过

图片切割模块增加了网络的有效输入,数据纠正分支模块学习了与主干网络不相一致的特征,一定程度纠正了数据间的误差,增强了网络的分类能力,SSCM模型的识别准确率较RIDE提升了2.71个百分点。

模型在数据量不足的尾部类别上性能不佳是长尾识别的一个严重的问题,但是模型在尾部数据性能不佳很难被诸如准确率的这种标准的评价指标所反映出来。为了更直观地显示出模型的改进效果,将超过100张图像的类别划分为头部类,低于30张图像的类别划分为尾部类,所有模型的头部和尾部对应的准确率见表2。由表2可知,SSCM模型在尾部数据获得了最好的性能提升,达到了68.34%,并且在头部数据的性能也有提升,比RIDE方法提升了2.85个百分点。结果表明,SSCM在没有牺牲模型在头部数据的性能的情况下,提升了对尾部数据识别的性能,进而提高了总体的识别性能。

表 1 不同模型在蝴蝶数据集上的性能比较

Table 1 Comparison of performance metrics of models on the butterfly dataset

模型 Model	网络主干 Backbone	年份 Year	第一准确率/% Top-1 accuracy	召回率/% Recall	精确率/% Precision	F ₁ 分数 F ₁ -score
—	ResNet-50	2016	82.80	76.1	81.4	0.755
DRC	ResNeXt-50	2019	83.10	76.8	82.3	0.759
BBN	ResNet-50	2020	83.96	77.4	81.6	0.775
RIDE	ResNeXt-50	2021	83.39	76.4	81.5	0.761
This study	ResNet-50	2022	86.10	81.3	84.6	0.818

表 2 不同模型在蝴蝶数据集的头部与尾部性能比较

Table 2 Comparison of performance of models in head and tail data of butterfly dataset

模型 Model	网络主干 Backbone	年份 Year	头部类准确率/% Many-shot accuracy	尾部类准确率/% Few-shot accuracy
—	ResNet-50	2016	85.79	62.94
DRC	ResNeXt-50	2019	86.15	63.38
BBN	ResNet-50	2020	86.14	64.11
RIDE	ResNeXt-50	2021	86.25	65.49
This study	ResNet-50	2022	88.96	68.34

2.2 消融实验

使用 SSCM 在 300 种蝴蝶数据集上进行消融实验分析,结果如表 3 所示。表 3 中,基线模型是单网络 ResNet-50,“+”表示数据纠正分支模块接在该 Layer 层的输出位置,“—”表示保持原网络结构。在网络其他部分结构不变的前提下,在深度不同的网络层接入数据纠正分支模块,分析接入不同特征层与不同数量分类器对最终识别性能的影响。由表 3 可知,接入第一层的特征输出后反而获得更低的性能,数据纠正模块在接入第二层的输出后,性能开始超越了原来的单主干网络。在接入 Layer3 网络层的输出时,获得了最佳的模型性能。

表 3 蝴蝶数据集消融实验结果

Table 3 Ablation experiment comparison results on butterfly dataset

基准 Baseline	Layer1	Layer2	Layer3	准确率/% Accuracy
+	—	—	—	82.8
+	+	—	—	71.5
+	—	+	—	84.5
+	—	—	+	86.1

为了更直观地解释这个结果,将主干网络的每个 Layer 层的输出均转化为特征图,并且每层随机抽选 3 张特征图,如图 8 所示。由图 8 可知,特征的复杂度是逐层增加的,前面的网络层只学习到图片的基础特征,得到的特征图像也与原图片更加接近,随着网络层的加深,深层网络层可以学习更高级更局部

的特征,更好指导分类。将设计的数据纠正分支模块接在前面的网络层会导致学习基础特征的网络层被迫去学习更高级的特征,进而大大降低了网络的性能。接在 Layer3 后面的数据纠正分支模块由于具有原网络不同的卷积层,在不影响前面网络层学习特征的情况下,通过加入均方损失纠正分割模块产生的新数据与原数据集的误差,使得该模块的分类器与原网络的分类器具有不同的分类能力,从而获得了更好的性能提升。

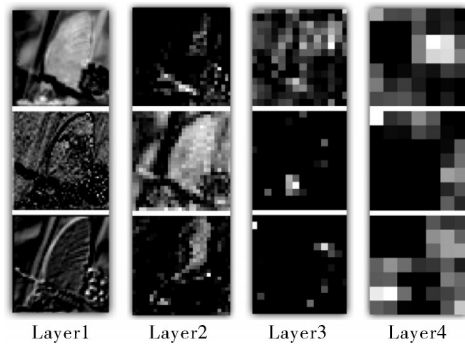


图 8 主干网络进行特征提取后的每层特征图
Fig.8 Feature maps for each layer of the feature extraction backbone

2.3 昆虫数据集试验结果

为进一步检验 SSCM 模型在呈现长尾分布的数据集上的分类性能,在同样具有长尾分布的公开 IP102 昆虫数据集上,与 DRC、BBN、RIDE 3 个主流的长尾识别模型进行对比。另外,SSCM 模型还与 Bollis 等^[26]在此昆虫数据集的试验结果进行了对比

(表4)。由表4可知, EfficientNet-B0达到了所有单网络模型最好的效果(59.80%), MobileNet-V2和EfficientNet-B0是比较有代表性的轻量级网络, 参数量大小约为 20×10^6 , 相比SSCM模型的参数量(100×10^6), 虽然模型较小, 但性能同样存在着大幅下降, SSCM模型的识别准确率分别低了18.15和11.35个百分点。在同样加载预训练参数的条件下, IP102的ResNet-50准确率为49.4%, 改进后的模型在此数据集上提升了21.75个百分点, 性能提升非常大。RIDE和BBN识别准确率分别是67.79%和68.13%, 但SSCM模型的识别准确率对比这2种方法分别提升了3.36和3.02个百分点。

表4 不同方法在IP102昆虫数据集上的分类性能比较

方法 Methods	网络主干 Backbone	年份 Year	第一准确率/% Top-1 accuracy
Bollis	Inception-v4	2020	48.20
IP102	ResNet-50	2019	49.40
DRC	ResNeXt-50	2019	52.21
Bollis	MobileNet-V2	2020	53.00
Bollis	NasNet-A Mob.	2020	53.40
Bollis	EfficientNet-B0	2020	59.80
RIDE	ResNeXt-50	2021	67.79
BBN	ResNet-50	2020	68.13
Ours	ResNet-50	2022	71.15

由表5可知, 在IP102昆虫数据集上SSCM模型相比于其他3个主流的长尾模型不仅获得了更好的识别性能, 而且模型性能在测试集上表现更稳定。

表5 IP102昆虫数据集不同模型的性能指标比较

模型 Model	网络主干 Backbone	年份 Year	第一准确率/% Top-1 accuracy	召回率/% Recall	精确率/% Precision	F_1 分数 F_1 -score
—	ResNet-50	2016	49.40	34.9	36.3	0.334
DRC	ResNeXt-50	2019	52.21	37.6	40.1	0.383
BBN	ResNet-50	2020	67.79	53.1	55.8	0.547
RIDE	ResNeXt-50	2021	68.13	59.8	64.5	0.567
Ours	ResNet-50	2022	71.15	67.6	70.2	0.673

由表6可知, 相比于头部数据的性能, 长尾模型在尾部性能有着更大的提升, SSCM模型比RIDE模型在尾部数据的识别准确率高8.33个百分点。

表6 不同模型在IP102数据集的头部与尾部性能比较

模型 Model	网络主干 Backbone	年份 Year	头部类准确率/% Many-shot accuracy	尾部类准确率/% Few-shot accuracy
—	Resnet-50	2016	61.61	16.90
DRC	Resnext-50	2019	64.35	19.52
BBN	Resnet-50	2020	76.35	34.04
RIDE	Resnext-50	2021	78.78	44.28
Ours	Resnet-50	2022	82.94	52.61

SSCM模型在IP102昆虫数据集上的消融试验结果与其在蝴蝶数据集中识别准确性表现一致, 数据纠正模块均在接入Layer3时获得了最好的性能提升(表7)。

表7 IP102昆虫数据集消融试验结果

基准 Baseline	Layer1	Layer2	Layer3	准确率/% Accuracy
+	—	—	—	66.69
+	+	—	—	53.34
+	—	+	—	69.67
+	—	—	+	71.15

3 讨论

本研究提出了一种SSCM模型以改善长尾识别中尾部类别过拟合问题。由于昆虫识别往往与网络学习到的主体特征有关, 如翅膀的纹理、颜色、形状以及躯干等可以表示该种类与其他种类的区别, 人眼在识别的过程中也常常利用这些特征。为了使网络更有效地学习到这些特征, 目前有2种主流的方法, 一种是基于整张图片的全局特征训练学习, 另一种是基于图片的局部特征对数据进行深一层的学习。受到第2种方法的启发, 本研究设计了图像分割与重组模块。即对输入的图像首先基于图像分割与重组模块对其随机分割并重铸, 进而在不改变数据集分布前提下提升尾部数据集。

具有更深层数的卷积神经网络(convolution neural network, CNN)例如VGG和EfficientNet等可以取得相对更高的平衡的图像数据分类准确率。但是现实生活中特定类别的数据往往是难以收集的, 导致这些类别的数据有限, 而有限的数据需要更高效提取特征的网络。为了满足这个需求, ResNet-50的残差结构解决了深层网络的梯度消失和梯度爆炸问

题,使得网络可以提取更丰富的特征信息,并且可以减轻数据集小带来的过拟合问题。

此外,本研究设计数据纠正分支模块,使用均方误差反馈新数据与原数据的差异,与原网络共同影响前面的特征层,以平衡网络的分类能力,从而加强网络的分类效果。试验结果表明,该模型在蝴蝶数据集和公开的昆虫长尾数据集上准确率分别达到86.1%、71.15%,较对比的方法提升了2.14和3.02个百分点。

训练好的模型作为后端部署在服务器上,通过小程序和App获取到实拍的蝴蝶图片,对图片进行快速识别,可以实时对农业环境里的昆虫进行快速识别,甄别农业害虫,为农业监测提供支持。后续研究中,将增加更多的昆虫类别,并且结合新的网络主干,以进一步探索特征学习与数据集不平衡的联系,进而保证识别的精确度和对识别种类多样性的同时提升。

参考文献 References

- [1] ESPELAND M, BREINHOLT J, WILLMOTT K R, et al. A comprehensive and dated phylogenomic analysis of butterflies[J]. *Current biology*, 2018, 28(5): 770-778.
- [2] KAYA Y, KAYCI L, UYAR M. Automatic identification of butterfly species based on local binary patterns and artificial neural network[J]. *Applied soft computing*, 2015, 28: 132-137.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the acm*, 2017, 60(6): 84-90.
- [4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. arXiv, 2014: 1409.1556 [2022-09-02]. <https://doi.org/10.48550/arXiv.1409.1556>.
- [5] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[DB/OL]. arXiv, 2014: 1409.4842 [2022-09-02]. <https://doi.org/10.48550/arXiv.1409.4842>.
- [6] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [DB/OL]. arXiv, 2019: 1905.02244 [2022-09-02]. <https://doi.org/10.48550/arXiv.1905.02244>.
- [7] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks [DB/OL]. arXiv, 2019: 1905.11946 [2022-09-02] <https://doi.org/10.48550/arXiv.1905.11946>.
- [8] 黄彤镔, 黄河清, 李震, 等. 基于YOLOv5改进模型的柑橘果实识别方法[J]. *华中农业大学学报*, 2022, 41(4): 170-177. HUANG T B, HUANG H Q, LI Z, et al. *Citrus* fruit recognition method based on the improved model of YOLOv5[J]. *Journal of Huazhong Agricultural University*, 2022, 41(4): 170-177 (in Chinese with English abstract).
- [9] 艾正茂, 谭鹤群, 万鹏. 基于机器视觉的日本鲭定向装置设计与试验[J]. *华中农业大学学报*, 2022, 41(2): 208-216. AI Z M, TAN H Q, WAN P. Design and experiment of a mackerel orientation device based on machine vision[J]. *Journal of Huazhong Agricultural University*, 2022, 41(2): 208-216 (in Chinese with English abstract).
- [10] REED W J. The Pareto, Zipf and other power laws[J]. *Economics letters*, 2001, 74(1): 15-19.
- [11] POUYANFAR S, TAO Y D, MOHAN A, et al. Dynamic sampling in convolutional neural networks for imbalanced data classification[C]//2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Mar. 21-23, 2018, Miami, the United States. Miami: IEEE, 2018: 112-117.
- [12] HE H B, GARCIA E A. Learning from imbalanced data [J]. *IEEE transactions on knowledge and data engineering*, 2009, 21(9): 1263-1284.
- [13] KANG B, XIE S, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition [DB/OL]. arXiv, 2019: 1910.09217 [2022-09-02]. <https://doi.org/10.48550/arXiv.1910.09217>.
- [14] ZHANG S Y, LI Z M, YAN S P, et al. Distribution alignment: a unified framework for long-tail visual recognition[DB/OL]. arXiv, 2021: 2103.16370 [2022-09-02]. <https://doi.org/10.48550/arXiv.2103.16370>.
- [15] CUI Y, JIA M L, LIN T Y, et al. Class-balanced loss based on effective number of samples [DB/OL]. arXiv, 2019: 1901.05555 [2022-09-02] <https://doi.org/10.48550/arXiv.1901.05555>.
- [16] DENG L M, WANG Y J, HAN Z Z, et al. Research on insect pest image detection and recognition based on bio-inspired methods[J]. *Biosystems engineering*, 2018, 169: 139-148.
- [17] XIE C J, WANG R J, ZHANG J, et al. Multi-level learning features for automatic classification of field crop pests[J]. *Computers and electronics in agriculture*, 2018, 152: 233-241.
- [18] WU X P, ZHAN C, LAI Y K, et al. IP102: a large-scale benchmark dataset for insect pest recognition [C]//Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 15 - 20, 2019, Long Beach, USA. Long Beach: IEEE, 2019: 8779-8788.
- [19] HE X T, PENG Y X, ZHAO J J. Which and how many regions to gaze: focus discriminative regions for fine-grained visual categorization[J]. *International journal of computer vision*, 2019, 127(9): 1235-1255.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [DB/OL]. arXiv, 2016: 1512.03385 [2022-09-02]. <https://doi.org/10.48550/arXiv.1512.03385>.
- [21] HARIHARAN B, GIRSHICK R. Low-shot visual recognition by shrinking and hallucinating features [DB/OL]. arXiv, 2017: 1606.02819 [2022-09-02]. <https://doi.org/10.48550/arXiv.1606.02819>.
- [22] GLOROT X, BORDE S A, BENGIO Y. Deep sparse rectifier

- neural networks [C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), October 15-18, 2021. Palo Alto, CA, USA. Palo Alto: [s.n.], 2021:315-323.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [DB/OL]. arXiv, 2015: 1502.01852 [2022-09-02]. <https://doi.org/10.48550/arXiv.1502.01852>.
- [24] ZHOU B Y, CUI Q, WEI X S, et al. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition [DB/OL]. arXiv, 2020: 1912.02413 [2022-09-02]. <https://doi.org/10.48550/arXiv.1912.02413>.
- [25] WANG X, LIAN L, MIAO Z, et al. Long-tailed recognition by routing diverse distribution-aware experts [DB/OL]. arXiv, 2020: 2010.01809 [2022-09-02]. <https://doi.org/10.48550/arXiv.2010.01809>.
- [26] BOLLIS E, PEDRINI H, AVILA S. Weakly supervised learning guided by activation mapping applied to a novel *Citrus* pest benchmark [DB/OL]. arXiv, 2020: 2004.11252 [2022-09-02]. <https://doi.org/10.48550/arXiv.2004.11252>.

Synthetic samples combined model-based recognition of long-tailed target

CAI Runji, JIANG Fangyong, ZHENG Taotao, LIU Donglin, XU Chudong
*College of Electronic Engineering (College of AI), South China Agricultural
University, Guangzhou 510642, China*

Abstract Insects are the most diverse animal group in nature. Some species are difficult to collect, which makes datasets often highly heterogeneous with long-tailed distributions. This article proposed a convolution recognition network model based on synthetic samples combined model (SSCM) to solve the problem that the uneven distribution of insect datasets leads to the poor recognition performance of recognition models in tail categories with less data. The model contains three modules including image segmentation and shuffle module, backbone network module and data fix branch module. Through the image segmentation and shuffle module, the training image was segmented and shuffled to obtain new training data and added to the training set. ResNet-50 was used as the network backbone to extract features of image. At the same time, the data fix branch module combined the mean square error and cross-entropy to calculate the error between the synthetic samples and the original image to reduce the adverse effect of the synthetic samples on the tail data. A butterfly dataset containing a total of 26 045 images of 300 species was constructed to evaluate the performance of the model proposed. The results showed that the accuracy of SSCM model was 3, 2.14 and 2.71 percentages higher than that of DRC, BBN and RIDE in the butterfly dataset, respectively. In addition, the validity of the SSCM in the public IP102 insect dataset was verified on the public insect dataset IP102. The results showed that the accuracy of SSCM model was 18.94, 3.02 and 3.36 percentages higher than that of DRC, BBN and RIDE, respectively.

Keywords insect recognition; synthetic samples; image segmentation; agricultural monitoring; long-tailed recognition; ecological environment monitoring; biodiversity protection; pest control

(责任编辑:陆文昌)