

文章编号:1671-6833(2022)01-0034-07

# 基于 YOLOv3 和注意力机制的野外蝴蝶种类识别

周文进, 李 凡, 薛 峰

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650500)

**摘 要:** 针对目前野外环境下蝴蝶分类粒度细致, 现有模型对其识别效率低、精度差的问题, 以野外蝴蝶图像的种类自动识别为目标, 在自建混合数据集基础上, 对 YOLOv3 模型的主干网络进行了改进, 提出一种内嵌通道注意力 MultiSE1D 识别网络。该网络使用多尺度提取高维特征, 使网络具有多种感受野, 更好地关注了蝴蝶众多子类间、周围环境间存在的局部细微差异; 并使用一维卷积代替压缩激励层, 避免通道特征降维的同时, 有效降低了模型参数量, 提升了模型运行效率。所提算法在自建数据集上进行实验, 结果表明: 模型最终的平均精度均值( $mAP$ ) 达到了 83.2%。改进的识别网络可以有效提升原模型提取蝴蝶图像特征的准确性及细节特征的学习能力, 能为野外蝴蝶数字图像的种类识别问题提供有效的解决方案。

**关键词:** 蝴蝶; 自动识别; YOLOv3; 通道注意力; 多尺度

**中图分类号:** TP391.41

**文献标志码:** A

**doi:** 10.13705/j.issn.1671-6833.2022.01.007

## 0 引言

蝴蝶(锤角亚目), 隶属于昆虫纲第二大目(鳞翅目), 全世界已知鳞翅目昆虫数量多达 20 余万种, 其中蝶类约占十分之一<sup>[1]</sup>。蝴蝶与人类关系密切, 不同种类的蝴蝶品种会对人类生产生活产生有益或有害的影响。由于其对环境变化的敏感, 近年来, 珍稀品种的蝴蝶数量急剧减少, 因此, 探索一种环保、智能的蝴蝶分类鉴定方法具有十分重要的实际意义, 也是保护众多珍稀蝴蝶品种的重要基础。

近年来, 随着人工智能领域的突破与发展, 越来越多的研究者投入到蝴蝶的识别研究中。传统的研究方法通过提取蝴蝶的颜色、纹理、形状等局部特征<sup>[2]</sup>并结合不同类型的分类器, 在部分品种的蝴蝶标本自动识别研究中取得了一些成果。Li 等<sup>[3]</sup>通过多尺度曲率直方图(HoMSC)和图像块的灰度共生矩阵(GLCMoIB)特征提取方法的设计, 结合 KNN 分类器实现了 50 种蝴蝶标本图像的自动识别。Chen 等<sup>[4]</sup>提出了一种新的特征描述方法(高斯卷积角)来描述蝴蝶标本图像中复杂的静脉特征和纹理特征, 在公开数据集上对 10

种蝴蝶进行了有效识别。然而, 蝴蝶翅面的自然形态、纹理特征十分复杂, 很难通过函数表达或统计描述的方式进行精准表达, 同时, 这类算法复杂度大, 当待识别种类数量增加时, 模型运算量急剧增加, 识别效率受到很大影响。

谢娟英等<sup>[5]</sup>制作了中国蝴蝶数据集, 在 Faster R-CNN 基础上对 94 种蝴蝶进行了识别, 其模型的  $mAP$  最低值接近 60%; 之后 Xie 等<sup>[6]</sup>又在其数据集的基础上提出新的数据划分方式与数据扩充技术, 在 RetinaNet 模型上实现了最好  $mAP$  为 79.7% 的识别效果。魏宏彬等<sup>[7]</sup>在 YOLOv3 的基础上, 结合新的边界框回归损失函数 DIoU 实现了蔬菜的自动识别。然而蝴蝶物种分类粒度极为细致, 同一属中种类多且形态差异小, 简单采用针对其他对象识别问题研发的网络模型, 很难取得较好的特征学习效果, 尤其针对亲缘关系较近的蝴蝶种类, 有区分效果的细微局部差异更是无法获取。此外, 现有的基于深度学习模型性能的提升主要通过增加网络层数实现, 这将带来模型参数激增、训练过程冗长、不易收敛等问题。

综上, 本文针对野外蝴蝶图像自动识别性能

收稿日期: 2021-05-06; 修订日期: 2021-07-20

基金项目: 云南省重大科技专项计划项目(202002AD080001)

通信作者: 李凡(1986—), 男, 云南昆明人, 昆明理工大学副教授, 博士, 主要从事计算机视觉方面研究, E-mail: 478263823@qq.com。

的提升问题,通过构建一个包含 200 种蝴蝶的混合数据集(其中包含野外拍摄的蝴蝶图像及室内拍摄的蝴蝶标本图像)的方式解决现有研究中高质量样本数据缺乏的情况,并在此基础上通过改进 YOLOv3 主干网络的方式,得到内嵌通道注意力 MultiSE1D 的识别网络,通过多尺度的方式提取高维特征,使网络具有多种感受野,并通过一维卷积代替压缩激励层的方式,在避免通道特征降维的同时,能有效降低模型参数量,提升模型总体效率。

1 实验对象及方法

1.1 数据集构成

本文构建的蝴蝶数据集收集自美国标本蝴蝶网站 (<http://www.butterfliesofamerica.com>), 蝴蝶分类层级为科、亚科、族、属、种,包含 4 科、96 属、200 种蝴蝶,每种蝴蝶均包含野外拍摄的图像和标本图像,共计 5 374 张图像。

数据集包含野外拍摄的自然状态蝴蝶图像 3 612 张,部分图像如图 1 所示。自然状态下的蝴蝶拟态性强,容易与周边的植物等环境发生混淆。这类蝴蝶图像数据会造成特征提取时涉及的蝴蝶主体信息太少,因此本文在为数据集选取野外图像时摒弃了只有一张图像的种类。野外蝴蝶种类图像数量最少的有 2 张,最多的有 117 张,大多数蝴蝶种类图像在 50 张以内。

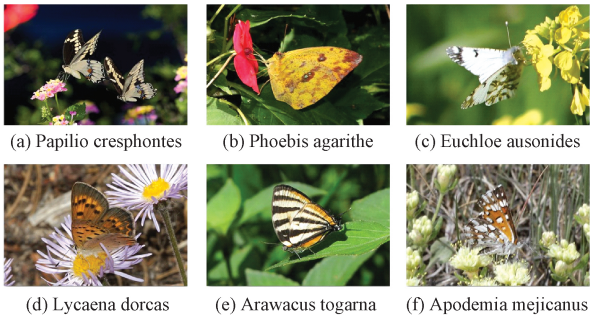


图 1 6 种野外蝴蝶图像

Figure 1 6 kinds of wild butterfly images

蝴蝶数据集中的标本图像共有 1 762 张,部分标本图像如图 2 所示,图像呈现的是蝴蝶的正面与背面,背景统一为白色,纹理颜色都比较清晰,且每张标本图像只包含一只蝴蝶,与野外图像相比,标本图像包含了蝴蝶身体部位的全部信息。为了在实验中不制造额外的变量,本文在为数据集选取标本图像时摒弃了破损、褪色严重的标本。标本蝴蝶种类图像中,绝大多数蝴蝶种类有 10 张图像,每个种类至少有正背面 2 张图像。

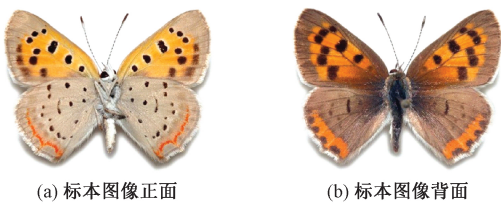


图 2 标本图像中蝴蝶的正面与背面

Figure 2 Front and back of butterfly in specimen image

1.2 数据集标注与划分

1.2.1 数据集标注

本文使用 Labellmg 标注工具对数据集图像进行位置标注,标注方式如图 3 所示,种类标注采用二名法(学名由属名与种名组成),如: *Papilio indra*。标注野外图像蝴蝶位置时尽量使标注区域包含触角,为了不引入太多不必要的背景,会摒弃过长的触角;标注标本图像蝴蝶位置时,只标注蝴蝶的主体位置,摒弃多余的白色背景,一方面与野外蝴蝶位置标注保持一致,另一方面提高蝴蝶主体特征提取效果。

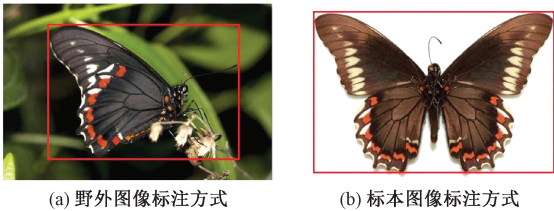


图 3 蝴蝶图像中蝴蝶主体标注方式

Figure 3 Annotation method of butterfly body in butterfly image

1.2.2 数据集划分

本文将蝴蝶数据集中野外图像按照训练集、测试集各 80%、20% 的比例进行划分。根据训练集的不同制作了 3 种数据集:数据集 1 中训练集包含 2 806 张野外图像,0 张标本图像;数据集 2 中训练集包含 0 张野外图像,1 762 张标本图像;数据集 3 中训练集包含 2 806 张野外图像与 1 762 张标本图像。本文的目标是野外蝴蝶的识别,因此上述 3 种数据集的测试集均为相同的野外蝴蝶图像,有 806 张图像。选择上述划分数据集的方式是因为通过分析野外图像与标本图像的特点,发现野外图像中的蝴蝶由于姿态、拍摄角度、遮挡、聚集等生物习性缺失了很多蝴蝶主体信息,而且有的种类野外图像资源非常少,特别是 10 张样本以内的,更加剧了蝴蝶主体信息的缺失,这样提取出来的特征不足以识别测试集中的样本,而且蝴蝶物种粒度极为细致,一个属下的种相似度非常高,甚至可能无法识别,而标本蝴蝶图

像正反面提供了完整的蝴蝶主体信息,可以弥补野外蝴蝶图像缺失的特征。

1.3 模型网络结构

本文采用YOLOv3<sup>[8]</sup>作为所有提出的网络模型的基线架构,模型网络结构如图4所示,ConvBlock1~ConvBlock5为各阶段卷积块,括号内给出了每个卷积块的特征图边长、通道数,Conv2D为二维卷积,Pool表示最大池化。Backbone(主干网络)选用轻量高效的特征提取网络Tiny-darknet,特征提取网络共7个卷积层,相邻卷积层间夹杂最大池化层进行下采样,整个网络结构中没有全连接层,并在主干网络中加入了改进的基于SENet<sup>[9]</sup>通道注意力网络MultiSE1D。Neck(检测颈)为两级的特征金字塔结构。Detection head(检测头)接收Neck中输出的两种尺度的特征图对不同大小的目标进行检测:第一个尺度特征图下采样比较高,达到了32倍,适合于检测图像中等和大目标;另一个尺度特征图通过拼接得到,下采样倍数为16,此处的特

征图感受视野一般,适用于较小目标检测。每种尺寸特征图的通道数 $C$ 计算如下:

$$C = 3(N + 5)。$$
 (1)

式中:3表示该尺度下每个网格中先验框的数量;5表示预测框的5个参数,分别为坐标、宽高、置信度; $N$ 表示要预测的种类个数,在本文中 $N$ 为200。

将改进SENet后的注意力网络(MultiSE1D)放置在主干网络的第7个卷积后,使得两个尺度的特征图都能共享到注意力网络的参数,同时高维特征经过注意力网络后的特征维度并没有发生改变,使其可以十分方便地移植到任何主干网络中。

1.4 MultiSE1D 通道注意力网络

本文嵌入主干网络的MultiSE1D注意力网络如图5所示,Block1~Block6为各阶段卷积块,括号内给出了每个卷积块的特征图边长、通道数,Conv1D为一维卷积。卷积块Block1经过两种尺度的卷积提取得到2个卷积块Block2、Block3,为

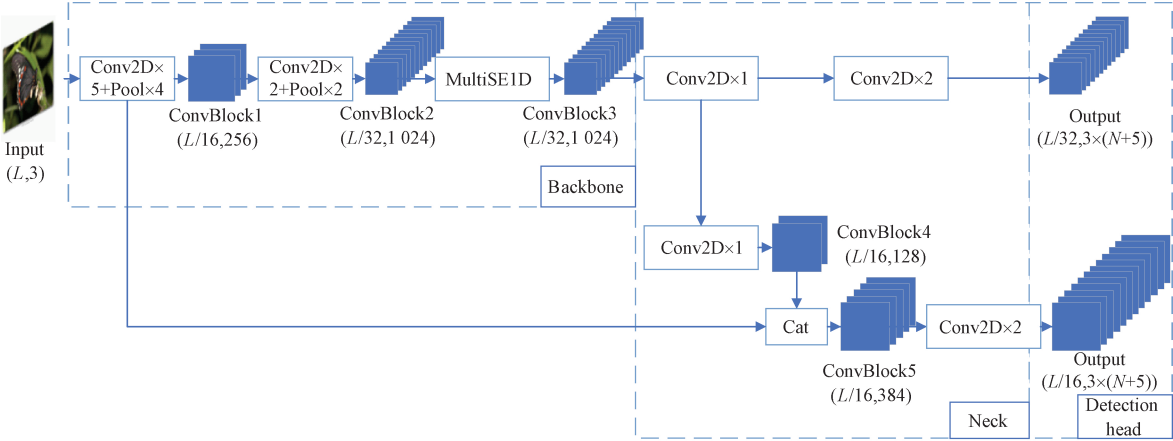


图4 模型网络结构

Figure 4 Structure diagram of model network

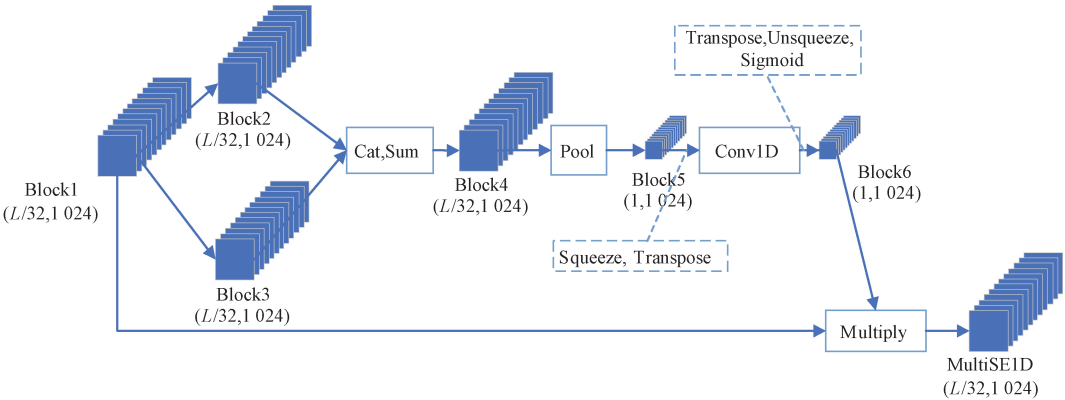


图5 MultiSE1D 网络结构图

Figure 5 Structure diagram of MultiSE1D network

后面生成的通道注意力权重具有两种感受野信息提供依据,Block4 由 2 个卷积块相加得到,具有两种感受野,可以更好地反映图像中局部信息与不同尺度的主体信息。Block4 经池化后每个通道特征图边长变为  $1 \times 1$ ,由于一维卷积输入为三维,并且本文是对 Block5 通道作一维卷积,所以在 Block5 进入一维卷积之前要进行降维与转置操作,并在一维卷积之后进行转置与升维恢复原来形式,再通过 Sigmoid 函数得到通道注意力权重 Block6,在这个过程中没有使用全连接层,并且 Block5 的通道进行一维卷积时没有发生降维,始终为 1 024。网络最后得到的 MultiSE1D 与原始 Block1 的特征图相比获得了局部跨通道联系。

#### 1.4.1 多尺度通道注意力网络构建

由于压缩激励网络原始模型<sup>[9]</sup>可以有效地利用图像全局信息,因此本文首先对高维特征的每个独立通道采取了全局平均池化,并通过两个全连接层(压缩激励层)的结合,实现了非线性通道间的相互作用,最后经过一个 Sigmoid 函数生成了通道权重  $X_{SE}$ :

$$X_{SE} = \sigma(FC_1 RELUFC_2(\varphi(X))). \quad (2)$$

式中:  $X \in \mathbf{R}^{\frac{L}{32} \times \frac{L}{32} \times 1024}$ ,  $L$  为网络输入图像边长,在实验中取值 512;  $\varphi(X) = \frac{1}{WH} \sum_{i=1, j=1}^{w, h} X_{ij}$  为全局平均池化(GPA);  $FC_1 RELUFC_2$  为一系列操作,其中  $FC_1$  与  $FC_2$  表示两个全连接层,尺寸分别为  $(C, \frac{C}{r})$ 、 $(\frac{C}{r}, C)$ ,  $r$  为压缩系数,两个全连接层之间加入 RELU 函数防止梯度消失,由于全连接层相比卷积层参数量巨大,为了降低模型的复杂度,采用了降维操作,在本文实验中  $r$  设置为  $16^{[10]}$ 。

原始的通道注意力<sup>[9]</sup>在高维特征中提取通道相关性是通过单一支路用固定的卷积核卷积实现的,这意味着网络在提取特征时只能拥有单一的感受野,实际上可以用  $5 \times 5$ 、 $7 \times 7$  等其他奇数卷积核来提取特征,参考 Li 等<sup>[11]</sup>使用不同的卷积核来获取不同的感受野方法,对通道注意力提取高维特征的方法作出调整,图 5 中 Block2、Block3 分别为使用两种卷积核的卷积提取特征得到的卷积块,这样网络获取了多种感受野,可以更好地反映图像中局部信息与不同尺度的主体信息,然后把多种感受野特征相加在一起作为通道注意力网络的输入,得到多尺度通道特征权重  $X_{MultiSE}$ :

$$X_{MultiSE} = X_{SE} \left( \sum_{i=1}^n f_{Conv2D}(X) \right). \quad (3)$$

式中:  $n$  表示不同的卷积核尺度;  $f_{Conv2D}(\cdot)$  表示二维卷积操作,为了不增加模型复杂度,在实验中直接选取了  $3 \times 3$ 、 $5 \times 5$  的两种奇数卷积核。

#### 1.4.2 非压缩-激励的通道注意力网络构建

通道注意力网络几乎都是通过压缩激励层( $FC_1 RELUFC_2$ )、Sigmoid 函数来获得通道权重,然而网络中存在全连接层会导致模型参数大幅增加。空间注意力(CBAM)<sup>[10]</sup>将全连接层换成了卷积层来减少参数也取得了同样的检测效果。压缩激励层通常为了减少模型复杂度而进行降维,然而其通道特征首先被投影到低维空间,再将其映射回来,使得通道特征与其权重之间的对应是间接的,破坏了二者的直接对应关系,同时全连接层考虑了所有通道间的影响。本文更注重图像局部变化,而一维卷积操作具有局部性,充分考虑了相邻通道间的影响。考虑到以上问题,参考文献[12]方法,本文对多尺度通道注意力作出调整,图 5 中的 Conv1D 与原始通道注意力的区别主要体现在压缩激励层被自适应卷积核的一维卷积代替,因此式(2)中  $X_{SE}$  调整为

$$X_{SE1D} = \sigma(f_{Conv1D}^k(\varphi(X))). \quad (4)$$

$$k = \begin{cases} t, & t \text{ 为奇数;} \\ t + 1, & t \text{ 为偶数。} \end{cases} \quad (5)$$

$$t = 0.5 \log_2 C. \quad (6)$$

式中:  $f_{Conv1D}^k(\cdot)$  表示一维卷积操作;  $C$  为特征图的通道数;  $k$  为自适应卷积核,可以根据特征图的通道数  $C$  适应调整,根据图 5 可知,经过池化后的特征图通道数  $C$  为 1 024,计算可得一维卷积的卷积核  $k$  为 5,可使 Block5 的通道数经过一维卷积后依然保持不变,即  $k$  的自适应改变避免了通道数的降维。

## 2 实验

本文使用 Pytorch 构建识别网络,在构建的 3 类数据集上实验。实验中的硬件条件为 CPU: Inter CORE i5 9th Gen, 8 GB RAM; GPU: NVIDIA GeForce GTX 1050 (3G)。为加快模型收敛速度,实验中使用在 ImageNet 数据集上预训练过的权重并在本文数据集上进行微调,总共训练 300 个 epoch, batch-size 设为 4,其他超参数使用默认值。模型初始学习率设置为 0.01,学习率的调整策略采用余弦退火衰减<sup>[13]</sup>,学习率随着 epoch 的增加而递减,刚开始学习率下降比较缓慢,当训练到一



半时学习率下降速度变快,最终接近 0.000 5,整个学习率变化类似于余弦曲线。

### 2.1 评价指标

本文采用交并比 (*IoU*) 作为蝴蝶的定位指标,在本文中取 *IoU* 的阈值为 0.5,真实框与其对应网格所有的先验框都产生 *IoU*,取最大值为正例,正例产生置信度损失、检测框损失、类别损失,小于 *IoU* 阈值的预测框作为负例,只产生置信度损失;大于 *IoU* 阈值的预测框(正例除外)将作为忽略样例,不产生任何损失。本文采用平均精度均值 *mAP* 来评价野外蝴蝶的分类情况。*mAP* 是一种衡量模型在所有类别上平均精度的指标,计算如下:

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}。 \tag{7}$$

式中: *N* 为测试集中的野外蝴蝶种数,在本文中 *N* 为 200; *AP* 为平均精度,为 *PR* 曲线的积分,即精度与召回率曲线下方的面积,可由式(8)~(10)计算

$$AP = \int_0^1 P(R) dR; \tag{8}$$

$$P = \frac{TP}{TP + FP}; \tag{9}$$

$$R = \frac{TP}{TP + FN}。 \tag{10}$$

式中: *P* 为 *PR* 曲线的纵坐标; *R* 为 *PR* 曲线的横坐标; *TP* 为被正确划分为正例的个数; *FP* 为被错误划分为正例的个数; *FN* 为被错误划分为负例的个数。

### 2.2 实验结果与分析

为了从本文提出的 3 个数据集中确定最佳数据集,使用原始的 YOLOv3 模型(实验中称为 Baseline)分别在 3 种数据集上进行实验,实验结果如表 1 所示,在数据集 1 上的 *mAP* 达到了 67.1%,在数据集 2 上的 *mAP* 只有 18.1%,最后在数据集 3 上进行实验得到了 80.7%的 *mAP*,比在数据集 1 的结果提升了 13.6 百分点,提升幅度接近于在数据集 2 上的 *mAP*,说明在训练集中加入标本图像可以很好地补充野外蝴蝶图像各种不可避免的特征损失,为野外蝴蝶的分类补充了关键因素。

确定好最佳数据集后,为 Baseline 添加通道注意力以进一步提升模型分类能力,实验结果如表 2 所示。表 2 中 SENet 为文献[9]原始的压缩

表 1 Baseline 在不同数据集上的性能

Table 1 Performance of Baseline on different datasets

数据集	训练集情况	<i>mAP</i> /%
数据集 1	只包含野外图像	67.1
数据集 2	只包含标本图像	18.1
数据集 3	野外与标本图像混合	80.7

表 2 SENet、MultiSE、MultiSE1D 分别加入 Baseline 在数据集上的性能

Table 2 Performance of SENet, MultiSE and MultiSE1D on dataset 3 after adding Baseline

模型	数据集	参数量/ $10^6$	<i>mAP</i> /%
Baseline +SENet	数据集 1	9.26	69.4
Baseline +SENet	数据集 3	9.26	79.5
Baseline +MultiSE	数据集 3	46.3	82.2
Baseline +MultiSE1D	数据集 3	44.7	83.2

激励网络,添加到 Baseline 后在数据集 1 上 *mAP* 提升了 2.3 百分点,然而在数据集 3 上的 *mAP* 没有提升,反降了 1.2 百分点,说明原始的通道注意力并不能很好地适应含有标本图像的数据集。

分析数据集 3 中同一物种的野外图像与标本图像,野外照中蝴蝶主体大致占据整张图像的 1/3,而标本照中蝴蝶主体基本占据了整张图像,这样训练集中就有两种截然不同的蝴蝶主体尺寸,SENet 中提取高维特征直接对每个通道进行全局平均池化,这样显然不符合数据集 3 的情况,没有多尺度的特征提取与之对应。MultiSE 为采用多尺度提取特征的注意力网络,在数据集 3 上的 *mAP* 达到了 82.2%,相较 SENet 提升了 2.7 百分点,说明多尺度提取的方法可以更好地提取图像特征,更适合标本与野外蝴蝶图像混合的数据集。

然而性能的提升却带来模型参数的增加,如表 2 参数量所示,MultiSE 比 SENet 参数增加了近 4 倍,为了减少模型参数,同时避免注意力网络中的压缩激励层的降维,调整 MultiSE 得到没有全连接层同时不需要降维的 MultiSE1D,在数据集 3 上的 *mAP* 达到了 83.2%,较调整前提升了 1 百分点,参数量减少了  $1.6 \times 10^6$ ,说明一维卷积和避免降维的方法是有效的。

为了进一步验证本文注意力网络 MultiSE1D 的有效性,在 Baseline 中加入其他注意力网络:SESADRN<sup>[14]</sup>、GCNet<sup>[15]</sup>,在本文数据集 3 上的实验结果如表 3 所示,加入 SESADRN、GCNet 之后模型的 *mAP* 出现不同程度的小幅下降,MultiSE1D 的 *mAP* 比二者都高了近 3 百分点。由

于采用的数据集不同,本文只使用两种注意力网络的核心方法,其他实验因素无法保证一致,表 3 中实验数据只说明其方法在本文数据集上的一般性,MultiSE1D 更适用于标本与野外图像混合的数据集。

表 3 SESADRN、GCNet、MultiSE1D  
加入 Baseline 在数据集 3 上的性能

Table 3 Performance of SESADRN, GCNet  
and MultiSE1D on dataset 3 after adding Baseline

模型	mAP/%	参数量/10 <sup>6</sup>
Baseline	80.7	9.13
Baseline +SESADRN <sup>[14]</sup>	80.2	9.26
Baseline +GCNet <sup>[15]</sup>	80.4	17.5
Baseline +MultiSE1D	83.2	44.7

最后分别选取目标检测网络 Faster R-CNN<sup>[16]</sup>、SSD<sup>[17]</sup>、YOLOv3<sup>[8]</sup>、EfficientDet<sup>[18]</sup>、YOLOv4<sup>[19]</sup> 在数据集 3 上进行实验,实验时控制上述网络输入图像尺寸、学习率、*epoch*、*batch-size* 与本文实验保持一致,实验的硬件环境也一致,均没有使用数据在线扩增策略,不同点在于网络模型使用的主干网络不同,不同网络的 Neck 与 Detection head 也不同,实验结果以及每个检测模型使用的主干网络如表 4 所示。结果表明,本文内嵌通道注意力 MultiSE1D 的识别网络对野外蝴蝶的识别是有效的。

表 4 不同目标检测网络模型在数据集 3 上的性能

Table 4 Performance of different target detection  
network models on dataset 3

模型	mAP/%
Faster R-CNN(VGG16) <sup>[16]</sup>	79.4
SSD(ResNet50) <sup>[17]</sup>	82.8
YOLOv3(Tiny-darknet) <sup>[8]</sup>	80.7
EfficientDet(D1) <sup>[18]</sup>	82.0
YOLOv4(Darknet53) <sup>[19]</sup>	81.8
本文模型	83.2

### 3 结论

针对野外蝴蝶数字图像分类粒度细致、同属间分类特征差异小、与周边环境容易混淆等特点,本文提出了一种改进 YOLOv3 主干网络的方案,实现了内嵌通道注意力的 MultiSE1D 识别网络模型。该模型通过多尺度提取特征使网络具有多种感受野,能更好地学习图像中各种尺度的蝴蝶主体特征及局部分类特征;通过将全连接层替换为一维卷积的方式,有效避免了通道特征的降维,在获得通道特征与通道权重直接联系的同时,有效

降低了模型参数量。该模型通过在本文自建混合数据集上的训练及参数优化,获得了分辨亲缘关系较近的不同蝴蝶种类野外图像的能力。从本文对比实验结果可以看出,该方法可实现 200 种蝴蝶的野外图像自动识别,取得了 *mAP* 最高为 83.2% 的识别效果。相较于现有方法,本文改进的识别网络可以有效提升原模型提取蝴蝶图像特征的准确性及细节特征的学习能力,能为野外蝴蝶数字图像的种类识别问题提供有效的解决方案。

### 参考文献:

[1] 寿建新,周尧,李宇飞.世界蝴蝶分类名录[M].西安:陕西科学技术出版社,2006.

[2] WANG J N,JI L Q,LIANG A P,et al.The identification of butterfly families using content-based image retrieval[J].Biosystems engineering,2012,111(1):24-32.

[3] LI F, XIONG Y. Automatic identification of butterfly species based on HoMSC and GLCMoIB[J].The visual computer,2018,34(11):1525-1533.

[4] CHEN X, WANG B, GAO Y S. Gaussian convolution angles: invariant vein and texture descriptors for butterfly species identification [C]//2020 25th International Conference on Pattern Recognition (ICPR).Piscataway:IEEE,2021:5798-5803.

[5] 谢娟英,侯琦,史颖欢,等.蝴蝶种类自动识别研究[J].计算机研究与发展,2018,55(8):1609-1618.

[6] XIE J Y,LU Y Y,WU Z Z,et al.Investigations of butterfly species identification from images in natural environments [J]. International journal of machine learning and cybernetics,2021,12(8):2431-2442.

[7] 魏宏彬,张端金,杜广明,等.基于改进型 YOLOv3 的蔬菜识别算法[J].郑州大学学报(工学版),2020,41(2):7-12,31.

[8] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2021-05-01]. <https://arxiv.org/abs/1804.02767>.

[9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2018:7132-7141.

[10] WOO S,PARK J,LEE J Y,et al.CBAM:convolutional block attention module [C]//15th European Conference on Computer Vision(ECCV 2018).Cham: Springer,2018:3-19.

[11] 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).Piscataway:

IEEE,2019;510-519.

[ 12 ] 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 ). Piscataway: IEEE, 2020; 11531-11539.

[ 13 ] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks [ C ]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition ( CVPR ). Piscataway: IEEE, 2019; 558-567.

[ 14 ] XIN D J, CHEN Y W, LI J J. Fine-grained butterfly classification in ecological images using squeeze-and-excitation and spatial attention modules [ J ]. Applied sciences, 2020, 10( 5 ): 1681.

[ 15 ] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond [ C ]//2019 IEEE/CVF International Conference on Computer Vision Workshop ( ICCVW ). Piscataway: IEEE, 2019; 1971-1980.

[ 16 ] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [ J ]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39( 6 ): 1137-1149.

[ 17 ] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[ C ]//ECCV 2016: European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[ 18 ] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[ C ]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition ( CVPR ). Piscataway: IEEE, 2020: 10778-10787.

[ 19 ] BOCHKOVSKIY A, WANG C Y, LIAO H Y MARK. YOLOv4: optimal speed and accuracy of object detection[ EB/OL ]. ( 2020-04-23 ) [ 2021-05-01 ]. <https://arxiv.org/abs/2004.10934>.

Identification of Butterfly Species in the Wild Based on YOLOv3  
and Attention Mechanism

ZHOU Wenjin, LI Fan, XUE Feng

( Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China )

**Abstract:** To solve the problems of fine granularity of butterfly classification, low recognition efficiency and poor accuracy of existing models in the field environment, aim at the automatic recognition of butterfly species in the field, an embedded channel attention MultSE1D recognition network was proposed to improve the backbone network of YOLOv3 model on the basis of self-built hybrid data set. The network used multiscale to extract high-dimensional features, so that the network had a variety of receptive fields, and paid more attention to the local subtle differences between many subclasses of butterflies and the surrounding environment. One dimensional convolution was used to replace the compressed excitation layer to avoid dimensionality reduction of channel features, to reduce the model parameters and to improve the operation efficiency of the model. According to the above method, the final mean average precision (*mAP*) of the model was 83. 2%. The results showed that the improved recognition network could effectively improve the accuracy of the original model to extract the butterfly image features and the learning ability of the detail features, and could provide an effective solution to the problem of identifying the species of butterfly digital images in the wild.

**Keywords:** butterflies; automatic identification; YOLOv3; channel attention; multiscale