

# 一种迁移学习和可变形卷积深度学习的蝴蝶检测算法

李策<sup>1</sup> 张栋<sup>1</sup> 杜少毅<sup>2</sup> 朱子重<sup>1</sup> 贾盛泽<sup>1</sup> 曲延云<sup>3</sup>

**摘要** 针对自然生态蝴蝶多种特征检测的实际需求,以及生态环境下蝴蝶检测效率低、精度差问题,本文提出了一种基于迁移学习和可变形卷积深度神经网络的蝴蝶检测算法(Transfer learning and deformable convolution deep learning network, TDDNET). 该算法首先使用可变形卷积模型重建 ResNet-101 卷积层,强化特征提取网络对蝴蝶特征的学习,并以此结合区域建议网络(Region proposal network, RPN)构建二分类蝴蝶检测网络,以下简称 DNET-base;然后在 DNET-base 的模型上,构建 RPN 网络来指导可变形的敏感位置兴趣区域池化层,以便获得多尺度目标的评分特征图和更准确的位置,再由弱化非极大值抑制(Soft non-maximum suppression, Soft-NMS)精准分类形成 TDDNET 模型. 随后通过模型迁移,将 DNET-base 训练参数迁移至 TDDNET,有效降低数据分布不均造成的训练困难与检测性能差的影响,再由 Fine-tuning 方式快速训练 TDDNET 多分类网络,最终实现了对蝴蝶的精确检测. 所提算法在 854 张蝴蝶测试集上对蝴蝶检测结果的  $mAP_{0.5}$  为 0.9414、 $mAP_{0.7}$  为 0.9235、检出率 DR 为 0.9082 以及分类准确率 ACC 为 0.9370,均高于在同等硬件配置环境下的对比算法. 对比实验表明,所提算法对生态照蝴蝶可实现较高精度的检测.

**关键词** 蝴蝶生态照,可变形卷积,迁移学习,深度卷积神经网络

**引用格式** 李策,张栋,杜少毅,朱子重,贾盛泽,曲延云. 一种迁移学习和可变形卷积深度学习的蝴蝶检测算法. 自动化学报, 2019, 45(9): 1772–1782

**DOI** 10.16383/j.aas.c190104



开放科学(资源服务)标识码(OSID):

## A Butterfly Detection Algorithm Based on Transfer Learning and Deformable Convolution Deep Learning

LI Ce<sup>1</sup> ZHANG Dong<sup>1</sup> DU Shao-Yi<sup>2</sup> ZHU Zi-Zhong<sup>1</sup> JIA Sheng-Ze<sup>1</sup> QU Yan-Yun<sup>3</sup>

**Abstract** Aiming at the demand of butterfly multi-features recognition, and the problems of low precision and efficiency of butterfly detection in ecological environment, a butterfly detection with deformable convolution depth neural network based transfer learning is proposed (TDDNET). Firstly, the ResNet-101 convolutional layer is reconstructed by using the deformable convolutional model, which can reinforce the learning of feature extraction network for butterfly features. At the same time, this algorithm is combined with the region proposal network (RPN) to construct a two-classes detection network named DNET-base. Next, on the DNET-base to build TDDNET, the subnetwork RPN is used to guide the deformable sensitive position RoI pooling layer, which can obtain the scores feature map and the multi-scale object location. Then, we use the Soft-nms to obtain better detection results. Finally, the model after DNET-base training is transferred to the TDDNET, and fine-tuning the TDDNET multi-classification parameters. In testing datasets which have 854 images, the butterfly  $mAP_{0.5}$  of the proposed algorithm is 0.9414,  $mAP_{0.7}$  is 0.9235, the detection rate (DR) is 0.9082 and the classification accuracy (ACC) is 0.9370. The experiments demonstrate that the proposed algorithm outperforms the state-of-the-art model in the same hardware environment. The results show that the proposed algorithm can detect butterflies with high accuracy.

**Key words** Image of living butterfly, deformable convolution, transfer learning, deep convolution neural network

**Citation** Li Ce, Zhang Dong, Du Shao-Yi, Zhu Zi-Zhong, Jia Sheng-Ze, Qu Yan-Yun. A butterfly detection algorithm based on transfer learning and deformable convolution deep learning. *Acta Automatica Sinica*, 2019, 45(9): 1772–1782

收稿日期 2019-02-25 录用日期 2019-06-12  
Manuscript received February 25, 2019; accepted June 12, 2019  
国家重点研发计划重点专项(2017YFA0700800), 国家自然科学基金(61866022, 61876161), 甘肃省基础研究创新群体(1506RJIA031)资助

Supported by National Key Research and Development Program of China(2017YFA0700800), National Natural Science Foundation of China (61866022, 61876161) and Gansu Province Basic Research Innovation Group Project (1506RJIA031)

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 兰州理工大学电气工程与信息工程学院 兰州 730050 2. 西安交

通大学人工智能与机器人研究所 西安 710049 3. 厦门大学信息科学与技术学院计算机科学系 厦门 361005

1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050 2. Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049 3. College of Information Science and Engineering, the Department of Computer Science, Xiamen University, Xiamen 361005

丰富多样,《世界蝴蝶分类名录》<sup>[1]</sup> 记录了世界蝴蝶 17 科、47 亚科、1690 属、15141 种,其中记载中国蝴蝶 12 科、33 亚科、434 属、2153 种. 自 2016 年我国环境保护部启动了蝴蝶多样性观测工作<sup>[2]</sup>, 全国蝴蝶观测数据库的数据量呈现海量增长. 如何对生态蝴蝶及时且准确地检测, 这对昆虫分类学专家是一个很大的挑战. 因此, 自然生态蝴蝶种类检测问题已成为促进蝴蝶相关领域研究与应用的关键问题之一.

随着机器学习发展与应用, 为实现蝴蝶自动、快速、准确地检测与识别创造了有利条件. 2013 年 Kang 等<sup>[3]</sup> 提出了一种基于分支长度相似熵的形状识别方法, 使用 BLS 熵谱 (Branch length similarity, BLS) 作为 BP 神经网络的输入特征训练神经网络来识别蝴蝶. 2014 年 Kaya 等<sup>[4]</sup> 先后尝试了 Gabor 特征、颜色和纹理特征与极限学习机、人工神经网络以及 Logistic 回归等方法相结合, 探讨蝴蝶自动识别方法. 2015 年李凡<sup>[5]</sup> 提出基于蝴蝶形态与纹理分布规律的特征提取与优化方法, 采用改进的 K 最近邻<sup>[6]</sup> (K-nearest neighbor, KNN) 分类算法进行分类, 研究并实现了 50 种蝴蝶的自动分类方法. 近年来, 基于深度学习的蝴蝶目标检测取得了良好的检测结果, 主要原因是卷积神经网络可直接从图像像素级提取具有更加强表征力的特征<sup>[7-8]</sup>. 2016 年 Liu 等<sup>[9]</sup> 使用基于全局对比区域的方法来计算病虫害目标位置的显著性特征图, 再由深度卷积补缀网络 (Deep convolution neural network, DCNN) 对图像特征进行分类, 但此方法未考虑目标与背景相似等实例. 2017 年周爱明等<sup>[10]</sup> 使用 CaffeNet 模型在蝴蝶标本图像和 180 幅生态蝴蝶图像上, 通过训练不同的后验概率支持向量机做分类器来实现蝴蝶分类. 以上算法大多是对标本模式照进行识别, 不能直接应用在生态蝴蝶的检测问题上. 2018 年谢娟英等<sup>[11]</sup> 构建基于 Faster R-CNN<sup>[12]</sup> 的蝴蝶自动检测系统, 采用蝴蝶模式照图像和生态照图像对其模型进行训练, 然后在蝴蝶生态照片中实现对 94 类蝴蝶的自动检测, 并使用 ZF<sup>[13]</sup>、VGG\_CNN\_M\_1024<sup>[14]</sup>、VGG16<sup>[15]</sup> 三种预训练网络做算法自身对比实验, 并且均优于对比算法 YOLO-v2<sup>[16]</sup>、YOLO-v3<sup>[17]</sup> 模型.

综上所述, 现有的生态蝴蝶检测任务与相关算法目前仍存在以下三个问题: 1) 算法基本以蝴蝶标本模式照图像进行识别研究, 偏向于单纯的分类任务, 而在生态照图像上的拓展能力较弱; 2) 所使用的数据集中包含的蝴蝶类别偏少, 因此建立的识别模型泛化能力较低; 3) 对蝴蝶的分类一般到科级, 而对亚科到种名的精细区分较为困难.

由于采集的蝴蝶生态照图像常会受到光照和观

察点变化的影响, 多种类蝴蝶目标检测任务则成为了挑战性的视觉检测任务之一. 在自然生态照蝴蝶检测任务数据集<sup>[11]</sup> 上更具挑战性的因素有: 1) 待检测的蝴蝶生态图像分辨率迥异 (最大 7630 像素 ~ 4912 像素与最小 800 像素 ~ 450 像素), 且蝴蝶类别要求细分至种名; 2) 94 类蝴蝶多特征与多尺度, 且形态变化较大; 3) 背景复杂: 蝴蝶与背景相似性高; 4) 蝴蝶重叠、曝光不足进一步导致检测难等. 如图 1 所示为蝴蝶生态照示例.

针对上述问题和挑战, 本文提出了一种基于迁移学习和可变形卷积深度神经网络的蝴蝶检测算法 (TDDNET), 可对蝴蝶自然生态照中不同种名的 94 类蝴蝶实现较高精度检测.

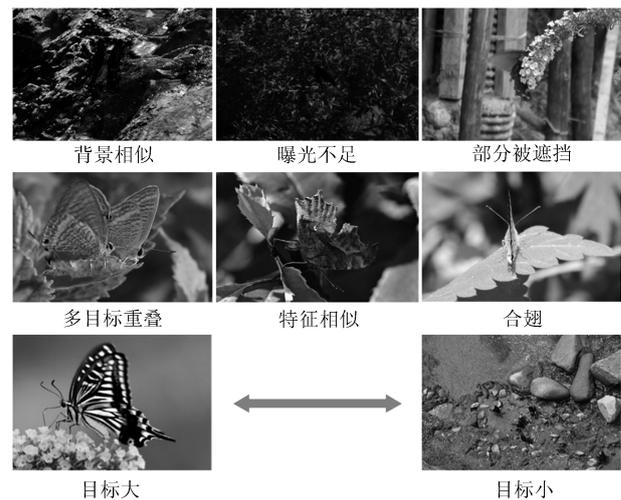


图 1 蝴蝶生态照示例图<sup>[11]</sup>

Fig. 1 Examples of butterfly ecology<sup>[11]</sup>

## 1 本文所提算法

本文提出了一种基于迁移学习和可变形卷积深度神经网络的蝴蝶检测算法 (TDDNET). 主要包括底层特征学习和模型迁移两个阶段, 所提算法框架如图 2 所示, 在第一阶段构建了二分类检测网络 (DNET-base) 强化特征学习能力; 第二阶段则由模型迁移方法优化 TDDNET 的检测性能. 其中, 第一阶段又分为两个部分: 1) 通过可变形卷积模型重建 ResNet-101<sup>[18]</sup> 特征提取层; 2) 结合 RPN<sup>[12]</sup> 网络构建一个二分类检测网络 (DNET-base), 对蝴蝶目标和背景进行分离训练, 增强网络对蝴蝶特征学习能力, 使得网络对蝴蝶的检出率更高. 而第二阶段则可分为三个部分: 1) 以 DNET-base 网络模型为基础, 保持特征提取网络模块的参数不变; 2) 重新构建 RoI<sup>[12]</sup> 池化模块, 以 RPN 网络指导敏感位置区域可变形 RoI 池化过程, 获得多尺度目标的评分特征图和精准位置信息; 3) 通过 Soft-max 和 Soft-NMS<sup>[19]</sup> 进行多分类优化, 最终形成 TDDNET

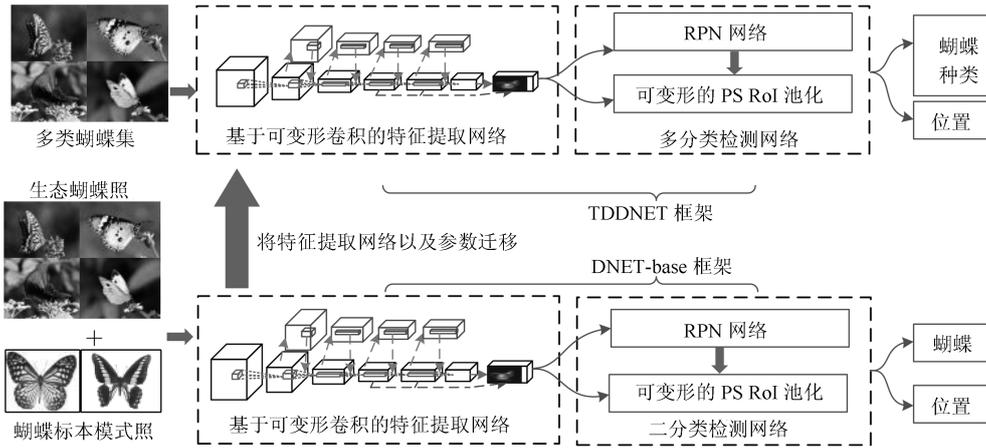


图 2 本文所提算法 TDDNET 的原理框架示意图

Fig. 2 Schematic diagram of TDDNET's principle framework proposed in this paper

模型.

在模型训练过程中: 先将蝴蝶数据归为一类, 通过训练重构的可变形卷积 ResNet-101 使得 DNET-base 模型对蝴蝶特征的提取能力最大化; 而后将 DNET-base 特征提取网络和参数迁移至 TDDNET 模型, 再重新训练 TDDNET 的多分类器. 这种方法可使 TDDNET 网络的收敛性变得更快更稳定. 下面, 将本文所提算法详述如下.

### 1.1 可变形卷积

在蝴蝶的自然生态照中, 蝴蝶多姿多样, 色彩斑斓, 并且部分蝴蝶伪装能力极强. 因此通过大量数据与数据增强使常规卷积网络完全“记忆”生态蝴蝶的多样变化较为困难. 因为常规卷积网络在构建模型变换时被限制于固定的几何结构, 这种局限性决定了卷积单元在输入图像上只能在固定位置上采样, 造成卷积层提取的特征表征能力较弱. 在类似卷积的池化过程也只能在固定的比例下降低特征空间分辨率, 致使特征丢失严重, 进一步导致损失函数的拟合能力弱和网络检测精度较差. 为解决上述问题, 本文所提算法中采用 Dai 等<sup>[20]</sup>提出的可变形卷积模型, 重新构建 ResNet<sup>[18]</sup>网络结构, 以及 RoI 兴趣区域池化模型来提升网络检测性能. 如图 3 所示, 为常规卷积与可变形卷积采样方式示例.

可变形的卷积模型引入了空间几何形变的学习能力, 从而更好地适应空间形变的目标特征提取与目标检测任务. 如图 3(b)~3(d) 所示, 在可变形卷积中将常规的网格拓展为拥有偏置量  $\{\Delta p_n | n = 1, \dots, N\}$  的偏置矩阵 Offset, 其中  $N = |R|$ . 对输入图像中的每一个在  $p_0$  位置上的变形卷积如式 (1) 所示.

$$y(p_n) = \sum_{p_n \in R} \omega(p_n)x(p_0 + p_n + \Delta p_n) \quad (1)$$

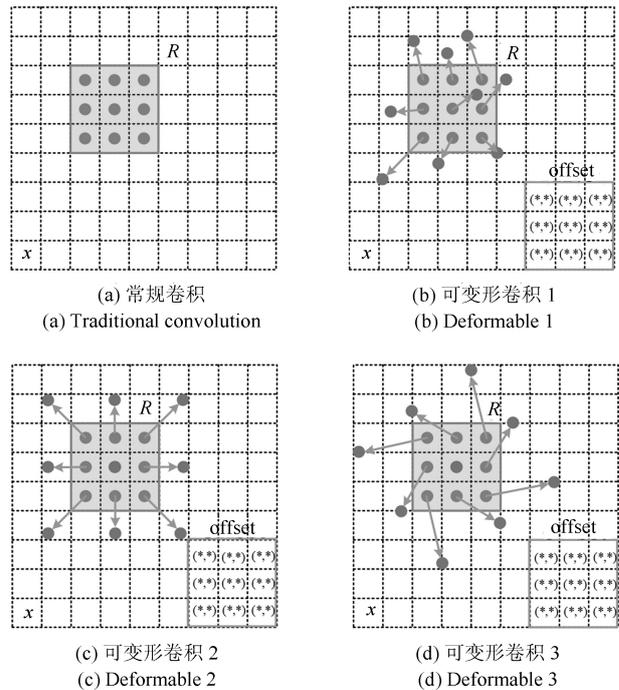


图 3 常规卷积和可变形卷积<sup>[22]</sup>的采样方式示例

Fig. 3 The instances of traditional and deformable convolution<sup>[22]</sup>

其中,  $\Delta p_n$  为偏置量, 且是一个分数,  $\omega(*)$  为采样点权重. 然而, 这样的操作引入了一个新问题, 即需要对不连续的位置变量求导. 借鉴 Jaderberg 等<sup>[21]</sup>的双线性插值的思想来求解. 偏置矩阵通过一个同输入图像一样大小的卷积层学习而来, 如图 4 所示, 即获得的偏置域的大小与输入图像的大小一致, 其中通道维度  $2N$  对应  $N$  个二维的偏置矩阵. 卷积核与现有的卷积一样具有相同的空间解析度和扩展度 (如在图 4 中  $3 \times 3$  的核扩展度为 1). 同样, 将得到的特征图送入下一个卷积层时, 采用一样的可变形

卷积模型提取特征. 在训练时, 同时学习用于产生输出特征的卷积内核和偏置矩阵.

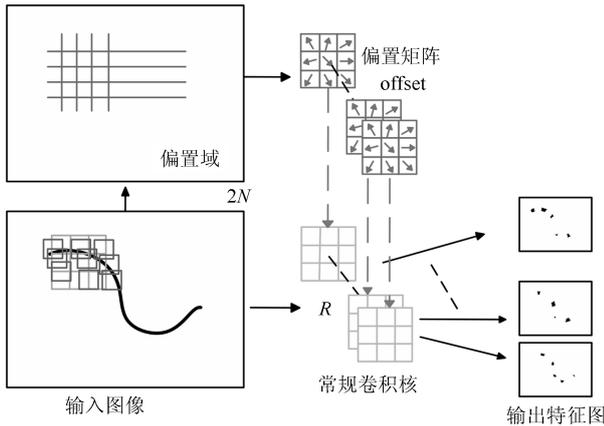


图 4  $3 \times 3$  可变形卷积特征计算过程示例

Fig. 4 An example of deformable convolution feature calculation process ( $3 \times 3$ )

因偏置矩阵使卷积的采样位置可自由变换, 偏置矩阵定义了  $R$  接受域的大小和扩张量, 所以可变形卷积的偏置域指向的采样点对目标趋向性较强, 则输出特征信息就较多. 这种自适应确定蝴蝶形变尺度和蝴蝶位置的方法在检测中是非常有效的. 如图 5 所示, 当可变形卷积效果堆叠时, 其模型的复合变形对特征提取能力的提升影响也较大.

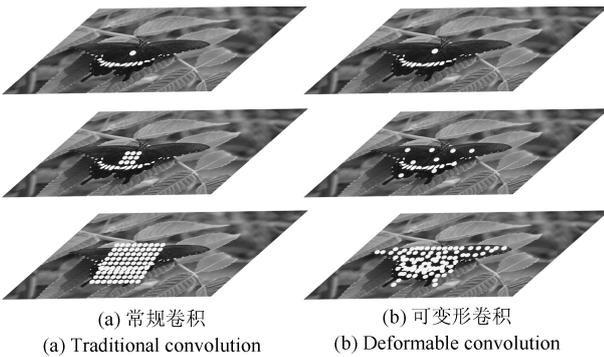


图 5 两种卷积在网络中的计算过程

Fig. 5 The computation of both convolutions in networks

### 1.2 可变形的位置敏感兴趣区域池化

为了获得自然生态图像中蝴蝶的位置和分类, 所提算法采用了位置敏感兴趣区域<sup>[22]</sup> (Position sensitive RoI pooling, PS RoI) 池化来构建 TDDNE 的分类器. 随着网络的加深, 其平移旋转不变性越强, 这个性质对于保证分类模型的鲁棒性具有积极意义. 然而, 在检测问题中, 过度的平移旋转不变性, 又使得网络对目标位置信息的感知能力削弱. 因此, 在蝴蝶检测中考虑到特征提取和对小目标检测边框对齐问题, 采取与可变形卷积一样的结

构策略来构建区域空间可变形的敏感位置 RoI 池化过程. 同时, 取消特征聚集过程中的量化操作, 使用双线性插值的方法获得浮点数的像素坐标, 让整个特征聚集过程转换为一个连续量的操作, 最后进行均值池化. 如图 6 所示, 为可变形的位置敏感 RoI 池化.

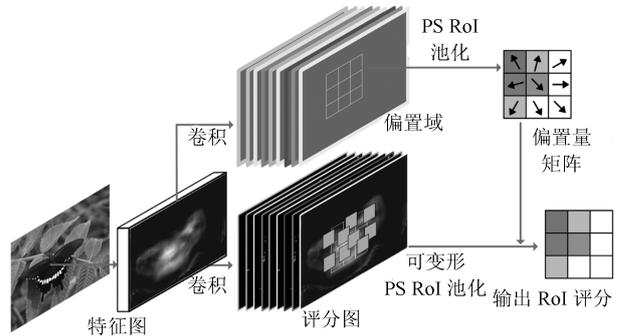


图 6 可变形的位置敏感 RoI 池化示意

Fig. 6 Deformable pooling of position sensitive RoI

首先, 利用卷积产生特征图的偏置域, 由于 RoI 池化将 RoI 区域分成  $k \times k$  块 ( $k$  为可调参数), 因此通过一个全卷积层可产生偏置量矩阵  $\{\Delta p_{ij} | 0 \leq i, j < k\}$ , 它被附加在区域空间池化块 (类似卷积核) 的位置上, 即可得到可变形的 PS RoI 池化模型, 即第  $(i, j)$  块的输出值是通过其对应的评分图  $x_{i,j}$  (替换常规特征图计算中的) 而得到. 因此, 可参照式 (1) 改写可变形的敏感位置 RoI 池化的计算式 (2).

$$y(i, j) = \sum_{p \in bin(i, j)} \frac{x(p_0 + p + \Delta p_{ij})}{n_{ij}} \quad (2)$$

其中,  $\Delta p_{ij}$  仍是一个分数,  $n_{ij}$  是区域块位置上的像素数且这个全卷积层是通过反向传播学习得来. 因为在可变形的 PS RoI 池化后得到固定大小的  $k \times k$  区域块特征, 直接用全连接层归一化即可得到  $k \times k$  个偏置域  $\Delta \hat{p}_{ij}$ . 但是这些偏置域并不能直接使用, 因为 RoI 区域大小不一致, 而且输入特征图的宽  $w$  和高  $h$  也不一致, 故采用一个增益  $\gamma = 0.1$  加以矫正, 与  $(w, h)$  点乘可得真值  $\Delta p_{ij} = \gamma \Delta \hat{p}_{ij} \cdot (w, h)$ .

### 1.3 构建蝴蝶检测网络 TDDNET 框架

**阶段 1 (DNET-base).** 考虑到数据集中蝴蝶种类数量分布不均等性和蝴蝶种类形态之间的相似性, 本文先设计了一种基于可变形卷积模型的二分类检测网络 (简称 DNET-base), 强化网络对特征的学习能力. 如图 8 中的第一阶段参数与二分类过程所示, 所建网络包括两个部分:

1) 通过可变形卷积模型重建 ResNet-101 特征提取过程, 即把 ResNet-101 的全连接层和均值池化

层都移除,重新构建 ResNet-101 结构剩余层 Conv2 (Res2c)、Conv3 (Res3b3)、Conv4 (Res4b 22) 和 Conv5 (Res5a、Res5b、Res5c) 的卷积层重构为可变形的卷积网络层. 具体构建方法如图 7 所示,通过对输入特征图进行全卷积获得相同维度的偏置域层 Offset,在 Offset 的偏置量指引作用下,对输入特征进行可变形卷积操作获得输出特征.

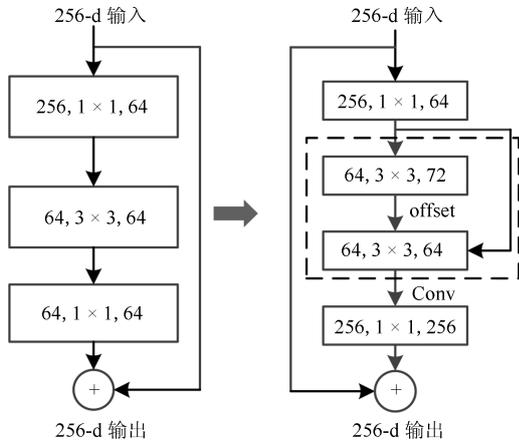


图 7 构建 ResNet 单元为可变形 ResNet 结构

Fig. 7 Construct the ResNet unit as a deformable ResNet structure RoI

2) 结合 RPN 网络构建二分类检测网络,通过 RPN 网络与 PS RoI 池化层后,可获得分类得分图和目标边框. 因为是二分类模型,即类别只有蝴蝶和背景. 通过 Soft-max 和 Soft-NMS<sup>[19]</sup> 在  $1000 \times 2$  的特征信息中获得准确的蝴蝶类别和位置信息. 即将数据集中的全部蝴蝶种类归为一类 “Butterfly”. 然后,通过 DNET-base 网络对蝴蝶目标和背景进行分离训练. 这种方法可有效避免部分种类数据匮乏与种类数据量不均衡(数据集中部分蝴蝶种类只

有 1 个样本,最多 92 个样本)造成的训练困难问题,可有效强化“第一阶段”网络对蝴蝶特征的提取能力.

**阶段 2 (TDDNET).** 如图 8 所示,借鉴迁移学习的思想,针对性地训练多分类器 (94 类 + 1 背景),即对第一阶段参数进行迁移,将其作为第二阶段中提取特征的基层网络参数,将其获得的聚集特征传递至第二阶段的多分类模型中进行训练,以此提高分类精确度. 首先,基于 DNET-base 构建 TDDNET 框架的特征提取网络. 然后,结合可变形卷积模型,构建以 RPN 网络指导敏感位置区域可变形 RoI 池化层部分,以此获得多尺度目标的评分特征图和精准位置信息,最后通过 Soft-max 和 Soft-NMS 进行多分类优化,形成完整的 TDDNET 模型. 并且在所提算法中使用在线难示例挖掘 (Online hard example mining, OHEM) 算法<sup>[23]</sup> 优化训练 PS RoI 的卷积检测算子,由此可获得最小的训练损失和较高的均值平均精度 (Mean average precision, mAP). 因为 OHEM 取消了人为设置的参数,同时放宽了正负样本的约束,以零阈值作为负样本下界,并取消正负样本比例. 其计算方法为 OHEM 对 RPN 提供的候选区域计算损失并排序,挑选出损失最大的目标区域为难样例再加入网络训练. 在所提算法中使用 Soft-NMS 方法提取目标边框. 非极大值抑制 NMS 算法,可获取目标的最佳坐标,并移除目标的重复边界框,如式 (3) 所示.

$$s_i = \begin{cases} s_i, & iou < N_t \\ 0, & iou \geq N_t \end{cases} \quad (3)$$

其中,  $s_i$  为评分,  $N_t$  为抑制阈值. 由于 NMS 采用置信度最高的检测方法,因其相邻目标检测框置信度强制为 0,所以对于区域重叠较大的目标

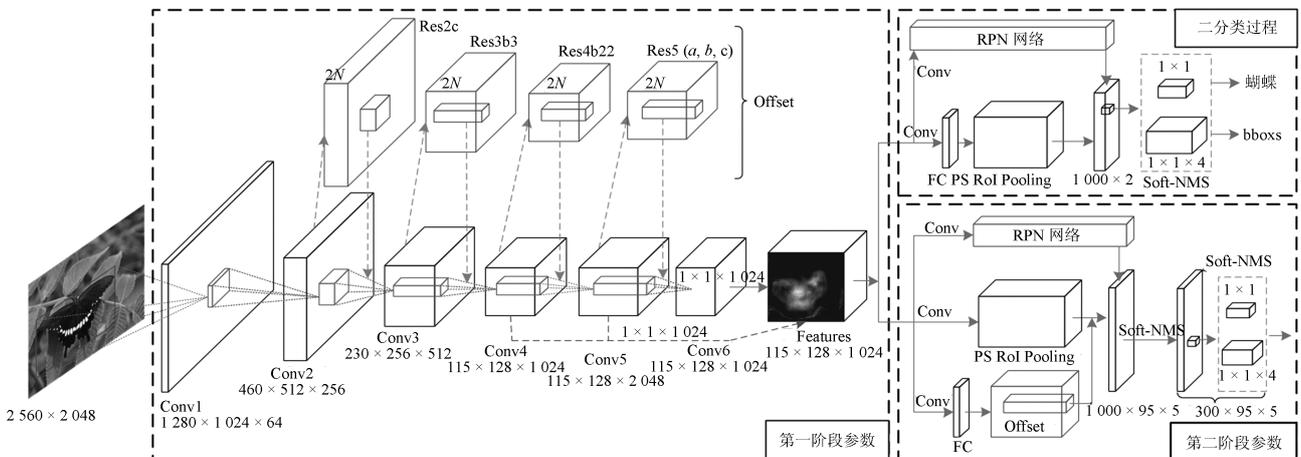


图 8 本文所提算法的网络模型与参数说明 (TDDNET)

Fig. 8 Network model and parameter description of the algorithm proposed in this paper (TDDNET)

(如图 1) 会出现漏检, 从而导致算法的检出率降低. Soft-NMS<sup>[19]</sup> 中将 NMS 算法进行改进, 使得  $s_i \leftarrow s_i f(iou(M, b_i))$ , 由此可使用线性加权的方式改写 NMS 算法函数, 如式 (4) 所示.

$$s_i = \begin{cases} s_i, & iou < N_t \\ s_i(1 - iou(M, b_i)), & iou \geq N_t \end{cases} \quad (4)$$

其中,  $iou(M, b_i)$  为最大评分的边界框  $M$  与待处理边界框  $b_i$  的交并比. 在 TDDNET 框架中使用的损失函数与 R-FCN 和 Faster R-CNN 中的一样, 采用多目标检测损失函数, 即同时考虑分类损失和位置损失. 在可变形的 PSRoI 池化后会得到  $k^2$  个区域块, 对每一个区域块都有  $c+1$  (为  $c$  类 +1 背景) 维的分类预测向量, 由此产生分类得分特征图, 如式 (5) 和 (6) 所示.

$$r_c(\Theta) = \sum_{(x,y) \in bin(i,j)} \frac{m_{i,j,c}(\Phi|\Theta)}{n} \quad (5)$$

$$\Phi = (x + x_0 + \Delta x_{bin(i,j)}, y + y_0 + \Delta y_{bin(i,j)}) \quad (6)$$

其中,  $m_{i,j,c}$  为  $k^2(c+1)$  个得分特征图之一,  $n$  为区域块中的像素数量,  $(x_0, y_0)$  表示 RoI 区域块的左上角位置.  $\Delta x_{bin(i,j)}$ ,  $\Delta y_{bin(i,j)}$  为第  $(i, j)$  区域块的偏置量,  $\Theta$  为 TDDNET 的训练参数. 所提算法中使用 Soft-max 来响应分类, 如式 (7) 所示.

$$S_c(\Theta) = \frac{e^{r_c(\Theta)}}{\sum_{c'=0}^c e^{r_{c'}(\Theta)}} \quad (7)$$

由此, 可通过交叉熵损失和 Soft-L1<sup>[10, 22]</sup> 边框回归定义 TDDNET 中的损失函数, 如式 (8) 所示.

$$L(S, b_{(x,y,w,h)}) = L_{cls}(S, S_{c^*}) + \lambda[c^* > 0]L_{reg}(b, b^*) \quad (8)$$

其中,  $b_{(x,y,w,h)}$  为预测位置,  $b^*$  为 Ground-Truth 目标位置标注值,  $c^*$  为类别真值标签, 如果  $c^* = 0$  表示为背景,  $L_{cls}(S, S_{c^*}) = -\ln_{cls}(S|S_{c^*})$  表示交叉熵损失函数,  $L_{reg}$  表示 Soft-L1 边框回归损失函数.

## 2 实验结果与分析

为了验证所提算法的有效性, 在 2018 年第三届中国数据挖掘竞赛上提供的蝴蝶生态照数据集<sup>[11]</sup>上, 与现阶段一些主流目标检测算法做对比实验. 对比算法分别为 Faster R-CNN<sup>[12]</sup>、FPN<sup>[24]</sup>、R-FCN<sup>[22]</sup>、SSD<sup>[25]</sup>、YOLO-v3<sup>[17]</sup>, 其中还对比了由可变形卷积模型重建的 Faster R-CNN、R-FCN、FPN 网络模型的变体. 通过定性和定量的实验对比, 验证

所提算法在生态照上的蝴蝶目标检测效果较好. 所提算法与对比实验评测平台信息为: 1) CPU 为 Intel Core i7 6700; 2) 采用英伟达 GTX 1070 8 GB 显存 GPU; 3) 使用 Ubuntu 16.04 操作系统, 内存 16 GB; 4) 除了 YOLO-v3 网络实验外, 所提算法与对比实验均依赖于 MXNET 开发库框架, 其版本为 0.12.0, OpenCV 版本为 3.4.1.

### 2.1 数据集

所提算法与对比算法使用数据包括: 标准数据集和拓展数据集. 其中, 标准数据集为 2018 年第三届中国数据挖掘竞赛所提供的蝴蝶图像数据集, 其中蝴蝶生态照数据集 721 张共 94 类蝴蝶 (测试集暂未公开), 图像分辨率最大为  $7630 \times 4912$  与最小为  $800 \times 450$ , 且生态蝴蝶标注为种名类别, 如金裳凤蝶 (编号 AAaa0001002), 多姿麝凤蝶 (编号 AAaa0003011). 生态蝴蝶数据集中目标特征多样、尺度变化较大, 些许蝴蝶类的伪装色与背景极为相似, 以及部分图像曝光不足等特点. 数据集中有部分类别, 如西番翠凤蝶与克里翠凤蝶、云豹蛱蝶与伊诺小豹蛱蝶, 其在形态上基本一致, 但在纹理和颜色特征<sup>[26]</sup>上存在一定差异, 即一些细节纹理和颜色特征起到了主导作用, 具有一定细粒度特性<sup>[27]</sup>.

标准数据集中, 每种蝴蝶至少 1 个样本, 最多包含 92 个样本, 呈现出典型的长尾分布; 另外, 蝴蝶标本模式照图像数据集中与 94 类一致的有 480 张. 每种蝴蝶至少 1 个样本, 最多包含 11 个样本. 94 类蝴蝶的整体数据分布如图 9 所示.

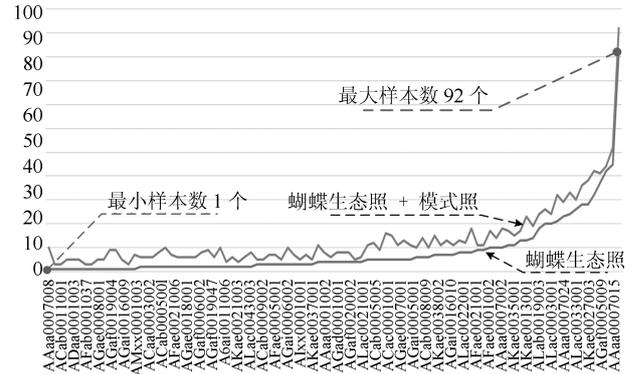


图 9 蝴蝶生态照图像数据集样本分布

Fig. 9 Sample distribution of butterfly image dataset

拓展数据集, 根据标准数据集的统计信息 (如图 9 所示) 进行再次收集, 将少于 10 个样本的蝴蝶种类进行数据扩充, 共扩充 789 张图像均来自于网络上的蝴蝶生态图像. 最后的数据集样本分布如图 10 所示. 在实验中做了简单交叉验证, 即将数据按照各个类的数量进行对半划分, 确保每个类在测试集和训练集中都有近似相等的数据量, 并做两者的

交替实验,其检测结果相差小于1%。因此,为了确保数据充分驱动模型,以及测试数据集公正性,在全部蝴蝶生态照图像数据集中,按照种类样本数量的20%抽取相应图像作为最终测试集,由此将蝴蝶生态照图像数据集划分为训练集1215张,测试集286张。

最后,考虑到数据集中蝴蝶目标平移或旋转不变性,对数据集进行增广,包括水平翻转、旋转±30°,得到蝴蝶训练数据集5085张(生态照1215×3张+模式照480×3张),蝴蝶测试照854张。从不同角度采集数据,对蝴蝶目标检测性能有一定的促进作用。所以,旋转角度可以随意选择,但角度选择不宜过多,以免造成冗余训练,拓展数据集分布情况如图10所示。

### 2.2 主观结果分析

针对生态蝴蝶图像一些特点,对所提算法与对比算法的改进变体在测试数据集上进行主观检测,如图11所示。在图11中实线箭头为误检目标(分类错误)、虚箭头为目标重叠框、菱形为漏检目标。从图11可看出,对于大目标的图像,本文算法和FPN\*均获得了完整目标框,然而FPN\*检出重叠目标,R-FCN\*和Faster R-CNN\*出现目标割裂,这说明蝴蝶形态(展翅正视与合翅侧视)在对比算法存在一定的影响。在对重叠目标中R-FCN\*与Faster RCNN\*均出现误检目标框,因此在特征区分度上较弱于其他算法。在相似目标、弱特征目标以及密集小目标上对比算法均出现了误检、漏检以

及重叠检测现象,本文所提算法仅出现了对部分小目标漏检,整体表现优于对比算法。在蝴蝶目标被遮挡和合翅正视情况下,本文算法可以获得较为完整的目标边界框,Faster RCNN\*出现了目标割裂和漏检情况,其他算法检测到的蝴蝶边界框与标注边界框重合度存在相对较大的差值,且在这两种情况下检测置信度都相对较低。因此,本文算法在大目标与重叠目标上对特征的细微区分表现较好,对背景相似目标与弱特征目标的检测也较为稳定,且少误检和漏检。在图11中,所提算法检测结果与预检测目标(Ground-truth)较相符。

### 2.3 客观结果分析

评价标准采用 $mAP_{0.5}$ 和 $mAP_{0.7}$ ,以及检出率(Detection rate, DR)和精确度(Accuracy, ACC)。其中,DR与ACC来自2018年第三届中国数据挖掘大赛的评价标准。检出率DR为所有覆盖率的平均值,其中覆盖率为交并比(Intersection over union, IoU)的值。精确度ACC为分类正确的数量与生态蝴蝶总数量的比值。

所提算法自身对比实验,包括四种情况:1)所提算法的完整模型验证;2)所提算法中使用NMS算法的验证;3)不采用迁移学习,即摒弃DNET-base模型架构,直接训练DDNET(NMS)模型架构验证,此模型也是我们在第三届数据挖掘大赛上使用的模型;4)所提算法中采用无可变形卷积的ResNet-101网络验证。如表1所示,所提算法检测效果表现较好。同时,对比了可变形卷积网络在不同

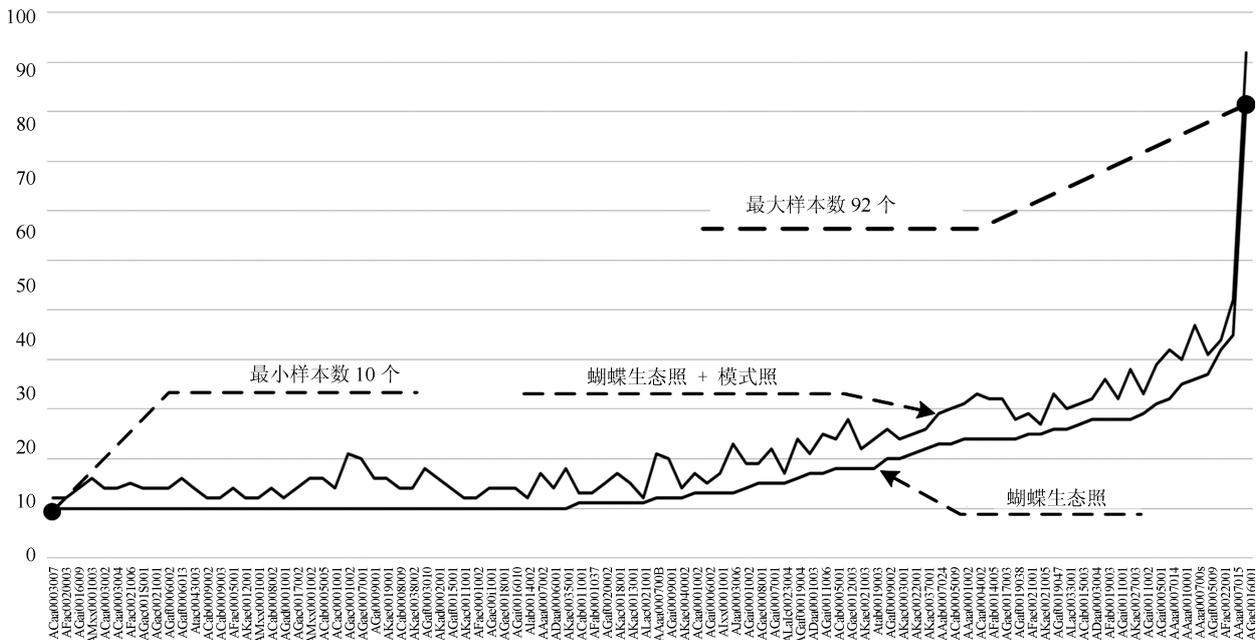


图10 蝴蝶生态照图像拓展数据集样本分布  
Fig.10 Sample distribution of butterfly image dataset

表 1 针对所提算法网络结构自身差异对比  
Table 1 Contrast the differences of the network structure of the proposed algorithm

网络结构差异	mAP0.5	mAP0.7	DR	ACC
TDDNET (Soft-NMS)	<b>0.9415</b>	<b>0.9235</b>	<b>0.9082</b>	<b>0.9370</b>
TDDNET (NMS)	0.9358	0.9208	0.9004	0.9274
DDNET (NMS, 无迁移)	0.9137	0.9009	0.8503	0.9180
TDDNET(无可变形卷积)	0.8827	0.8506	0.8532	0.8728

层时对所提算法的影响, 如表 2 所示.

在表 1 中的实验数据说明, 模型 DDNET (NMS, 无迁移) 的检测效果优于 TDDNET (无可变形卷积), 即加入可变形卷积对网络影响较大. 从测试数据中也表现出 Soft-NMS 的效果较优于 NMS 模型, 并且模型参数的迁移对网络也有积极的影响. 因此, 可变形卷积网络和模型迁移学习方法均有利于提高网络的检测性能.

表 2 针对所提算法中在不同层使用可变形卷积模型的差异  
Table 2 Aiming at the difference of using deformable convolution network in different layers of the proposed algorithm

可变形卷积网络层	mAP0.5	mAP0.7	DR	ACC
TDDNET 完整框架	<b>0.9415</b>	<b>0.9235</b>	<b>0.9082</b>	<b>0.9370</b>
TDDNET 框架 (除 Res2c)	0.9402	0.9174	0.9004	0.9304
Res5 (a, b, c) + PS RoI	0.9258	0.9076	0.8939	0.9186
PS RoI	0.9106	0.8902	0.8899	0.8960
Res5 (a, b, c)	0.8802	0.8609	0.8693	0.8901

在表 2 中实验数据说明, 可变形卷积层对网络的检测性能是非常有利的. 然而随着可变形卷积网络层数的增加, 网络参数也是成倍地增长, 网络耗时也在递增, 需要按照实际问题需求设置可变形卷积网络层.

所提算法与主流检测算法对比实验, 如表 3 所示, 对比算法包括 Faster R-CNN<sup>[12]</sup>、FPN<sup>[24]</sup>、R-FCN<sup>[22]</sup>、SSD<sup>[25]</sup>、YOLO-v3<sup>[17]</sup>, 其中 YOLO-v3 的预训练网络模型分别为 ResNet50 和 DarkNet. 同时, 也对比了由可变形卷积模型构建的 Faster R-CNN、R-FCN、FPN 网络模型的变体, 用 \* 表示. 在表 3 中实验数据说明, 所提算法优于对比算法. 并且, Faster R-CNN、R-FCN、FPN 在经过可变形卷积网络重建后, 相比原来算法检测性能上都有一定改善. FPN\* 与所提算法实验结果数值上非常接近, 然而, FPN\* 算法计算复杂度高于本文算法, 且耗时较长. 从 DR 值和 ACC 的值上也说明, 所提算法对蝴蝶特征的位置敏感性较好, 分类准确性也较高.

表 3 所提算法与其他目标检测算法的实验结果  
Table 3 Experimental results of the proposed algorithm and other target detection algorithms

对比算法	mAP0.5	mAP0.7	DR	ACC
Faster R-CNN <sup>[12]</sup>	0.7879	0.7418	0.8308	0.7845
Faster R-CNN*	0.8207	0.7932	0.8554	0.8144
R-FCN <sup>[22]</sup>	0.8650	0.8405	0.8650	0.8911
R-FCN*	0.8957	0.8594	0.8747	0.9087
FPN <sup>[24]</sup>	0.8926	0.8644	0.8994	0.9057
FPN*	0.9288	<b>0.9261</b>	0.8982	0.9206
SSD <sup>[25]</sup>	0.7794	0.7013	0.8648	0.7564
YOLO-v3 <sup>[17]</sup> (ResNet50)	0.7787	0.7785	0.8751	0.7956
YOLO-v3 <sup>[17]</sup> (DarkNet)	0.7889	0.7822	0.8746	0.8050
TDDNET	<b>0.9415</b>	0.9235	<b>0.9082</b>	<b>0.9370</b>

## 2.4 实验讨论

卷积网络对图像特征提取, 可理解卷积为滤波, 是对特征信息的筛选与叠加, 逐步让目标的特征转向语义特征, 而构建网络让这种多维特征语义化更加有效. 然而, 常规的卷积模型被固定结构所限制, Yu 等<sup>[28]</sup> 尝试扩张卷积的感受野, 可以对目标的轮廓信息保留较好. Zhou 等<sup>[29]</sup> 对卷积核做了旋转操作, 让网络可以得到目标的角度信息. Jeon 等<sup>[30]</sup> 通过星形蔓延方式改变固定的卷积核, 让离散的输入空间变成一个连续的采样空间, 来提升网络对特征提取能力. Jaderberg 等<sup>[21]</sup> 则利用仿射变换的思想提出了空间变换网络, 可对各种形变的数据进行空间变换, 以此提高分类准确率. 可变形卷积模型可以理解为卷积包含了上述模型的基本变换. 可变形卷积让卷积过程变得更加灵活, 对特征提取也变得更准确. 因此, 针对蝴蝶目标特征多样性、形态多变性等特点, 以及目标的细化分类与检测难问题, 在所提算法上有较大的改善, 并且获得了较好的检测效果.

因为生态蝴蝶图像与标本模式图像属于同源数据, 即存在较大的特征相似性. 由 DNET-base 把蝴蝶模式照的蝴蝶特征与生态图像的蝴蝶特征变换至相同的特征空间, 使其具有同样的分布并训练学习网络参数, 最后将所得参数迁移至 TDDNET, 再通过微调训练迅速地使 TDDNET 在蝴蝶检测任务上获得良好检测性能. 迁移学习还有更出色的算法有待挖掘, 可通过样本迁移、特征迁移、参数或模型迁移、关系迁移等促进深度学习网络对多分类小样本检测任务获得更好的检测性能. 此外, 因为蝴蝶数据<sup>[11]</sup> 的测试集并未公开, 故训练与测试中使用的数据集为 721 张生态照 + 模式照, 以及部分扩充数据.

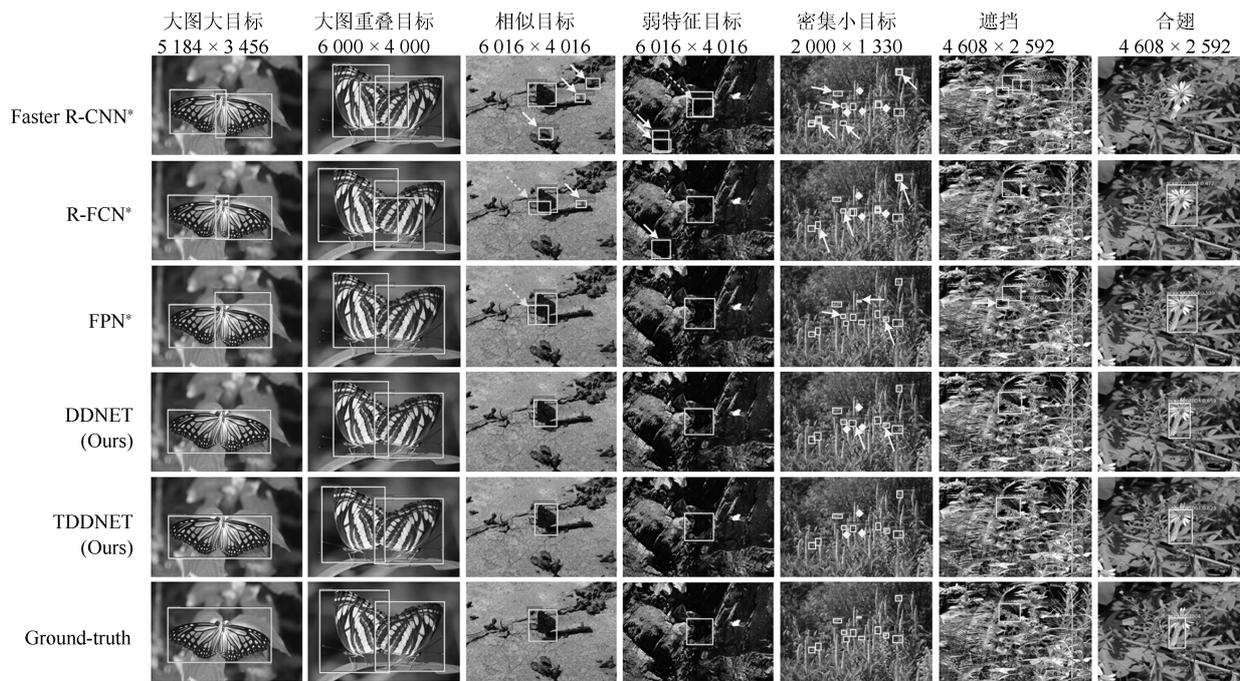


图 11 实验主观结果对比示例

Fig. 11 Contrastive examples of subjective results of experiments

在本文划分的测试集上, 使用 DDNET (NMS, 无迁移) 模型测试结果 DR 为 0.8503 和 ACC 为 0.9180. DDNET 也是我们在竞赛使用的算法模型, 竞赛测试结果 DR 为 0.8368 和 ACC 为 0.8380. 这也说明与 TDDNET 对比, 所提算法在生态蝴蝶检测性能上具有较大的提升.

本文骨干网络模型选择 RCNN 网络. 因为目标检测过程中有很多不确定的因素, 如图像中的目标形状、姿态、数量以及成像时会有光照、遮挡等因素干扰, 目标检测算法主要集中在两个方向: Twostage 算法 (如 RCNN<sup>[11-12, 22, 24, 31]</sup> 系列) 和 Onestage 算法 (如 YOLO<sup>[16-17]</sup>、SSD<sup>[25]</sup> 等). 两者主要区别在于 Twostage 算法需要先生成预选框, 然后进行细粒度目标检测. Onestage 算法会直接提取特征来预测目标分类和位置. 因此, 采用第一种方式的算法偏向于检测精度, 采用第二种方式的算法偏向于检测速度. 数据提供者要求每张图像检测时间在 2 秒内, 故本文算法的骨干网络模型选择 RCNN 网络, 并且所提算法实验检测效率约 2 张/秒至 3 张/秒.

因为在相同的任务上, 不同网络模型对目标学习的偏向性不同, 如检出率、精确性等. 如对生态照中的小目标蝴蝶群体, 利用蝴蝶的群体习性, 故可借鉴李策等<sup>[32]</sup> 的目标语义关联方法实现小目标检测. 也可使用多网络协同检测方式<sup>[33]</sup> 或者网络级联优化方式<sup>[34]</sup>, 通过融合检测结果获得较好的检测效果.

也可使用如 Inception<sup>[35]</sup>、ResNext<sup>[36]</sup> 等深度残差网络继续改进算法模型, 以及借鉴 Mask R-CNN<sup>[31]</sup> 中的 RoI Align 模型等解决“边框对齐问题”, 以此提升目标检测网络的精确度. 此外, 生态蝴蝶照图像中蝴蝶目标尺度变化也是非常大, 可以借鉴 Zhou 等<sup>[37]</sup> 和 Bharat 等<sup>[38]</sup> 的多尺度目标检测算法来改进生态蝴蝶目标检测.

### 3 结论

在分析了部分主流目标检测算法的优势, 以及在生态蝴蝶检测问题上的局限性的基础上, 针对蝴蝶生态照图像中的蝴蝶检测问题, 本文提出了一种基于迁移学习和可变形卷积深度神经网络的蝴蝶检测算法. 所提算法利用可变形卷积模型来增加特征网络的基层卷积层对特征的萃取能力, 结合 RPN 网络, 构建可变形的敏感区域池化模型进一步提升网络的检测精准性, 并利用迁移学习的思想, 在任务中有效地解决了数据样本不平衡与匮乏的问题. 在对比实验中也取得了较好的检测结果. 在未来的工作中将借鉴深度学习思想, 结合迁移学习知识与可变形卷积模型理论继续改进目标检测模型.

### 致谢

感谢 2018 年第三届中国数据挖掘大赛组委会组织的赛事以及提供的蝴蝶数据集.

## References

- 1 Shou Jian-Xin, Zhou Yao, Li Yu-Fei. *Systematic butterfly names of the world*. Shaanxi: Shaanxi Science and Technology Press, 2006  
(寿建新, 周尧, 李宇飞. 世界蝴蝶分类名录. 陕西: 陕西科学技术出版社, 2006)
- 2 Ma Fang-Zhou, Xu Hai-Geng, Chen Meng-Meng. Progress in construction of china butterfly diversity observation network (China BON-Butterflies). *Journal of Ecology and Rural Environment*, 2018, **34**(1): 27–36  
(马方舟, 徐海根, 陈萌萌. 全国蝴蝶多样性观测网络 (China BON-Butterflies) 建设进展. 生态与农村环境学报, 2018, **34**(1): 27–36)
- 3 Kang S H, Cho J H, Lee S H. Identification of butterfly based on their shapes when viewed from different angles using an artificial neural network. *Journal of Asia-Pacific Entomology*, 2014, **17**(2): 143–149
- 4 Kaya Y, Kayci L, Uyar M. Automatic identification of butterfly species based on local binary patterns and artificial neural network. *Applied Soft Computing*, 2015, **28**: 132–137
- 5 Li Fan. The research on automatic identification of butterfly species based on the digital image [Master dithesis], Beijing Forestry University, China, 2015  
(李凡. 基于数字图像的蝴蝶种类自动识别研究 [硕士学位论文], 北京林业大学, 中国, 2015)
- 6 Sun Xiao, Pan Ting, Ren Fu-Ji. Facial expression recognition using ROI-KNN deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 883–891  
(孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别. 自动化学报, 2016, **42**(6): 883–891)
- 7 Zhang Hui, Wang Kun-Feng, Wang Fei-Yue. Advances and perspectives on applications of deep learning in visual object detection. *Acta Automatica Sinica*, 2017, **43**(8): 1289–1305  
(张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望. 自动化学报, 2017, **43**(8): 1289–1305)
- 8 Chang Liang, Deng Xiao-Ming, Zhou Ming-Quan Wu Zhong-Ke, Yuan Ye, et al. Convolutional neural networks in image understanding. *Acta Automatica Sinica*, 2016, **42**(9): 1300–1312  
(常亮, 邓小明, 周明全, 武仲科, 袁野, 等. 图像理解中的卷积神经网络. 自动化学报, 2016, **42**(9): 1300–1312)
- 9 Liu Z Y, Gao J F, Yang G G. Localization and classification of paddy field pests using a saliency map and deep convolutional neural network. *Scientific Reports*, 2016, **6**: 204–210
- 10 Zhou Ai-Ming, Ma Peng-Peng, Xi Tian-Yu, Wang Jiang-Ning, Feng Jin, Shao Ze-Zhong, et al. Automatic identification of butterfly specimen images at the family level based on deep learning method. *Acta Entomologica Sinica*, 2017, **60**(11): 1339–1348  
(周爱明, 马鹏鹏, 席天宇, 王江宁, 冯晋, 邵泽中, 等. 基于深度学习的蝴蝶科级标本图像自动识别. 昆虫学报, 2017, **60**(11): 1339–1348)
- 11 Xie Juan-Ying, Hou Qi, Shi Ying-Huan, Lv Peng, Jing Li-Ping, Zhuan Fu-Zhen, et al. The automatic identification of butterfly species. *Journal of Computer Research and Development*, 2018, **55**(8): 1609–1618  
(谢娟英, 侯琦, 史颖欢, 吕鹏, 景丽萍, 庄福振, 等. 蝴蝶种类自动识别研究. 计算机研究与发展, 2018, **55**(8): 1609–1618)
- 12 Ren S Q, He Kai-Ming, Girshick R. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, **39**(6): 1137–1149
- 13 Zeiler M D, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland: Springer Verlag, 2014. 818–833
- 14 Chatfield, Ken. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint*, arXiv: 1405.3531, 2014
- 15 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*, arXiv: 1409.1556, 2015
- 16 Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 6517–6525
- 17 Redmon J, Farhadi A. YOLO-v3: An incremental improvement. *arXiv preprint*, arXiv: 1804.02767, 2018
- 18 He Kai-Ming, Zhang Xiang-Yu, Ren Shao-Ping. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA: IEEE, 2016. 770–778
- 19 Bodla N, Singh B, Chellappa R. Soft-NMS: Improving object detection with one line of code. In: Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017. 5562–5570
- 20 Dai J F, Qi H Z, Xiong Y W. Deformable convolutional networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017. 764–773
- 21 Jadergarg M, Simonyan K, Zisserman A. Spatial transformer networks. In: Proceedings of the 30th Annual Conference on Neural Information Processing Systems, Barcelona, Spain: Curran Associates, Inc. 2016. 2017–2025
- 22 Dai J F, Li Y, He K M. R-FCN: Object detection via region-based fully convolutional networks. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 379–387
- 23 Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA: IEEE, 2016. 761–769
- 24 Lin T Y, Dollar P, Girshick R B. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 936–944
- 25 Liu W, Anguelov D, Erhan D. SSD: Single shot multibox detector. In: Proceedings of the 13th European conference on computer vision, Amsterdam, the Netherlands: Springer Verlag, 2016. 21–37
- 26 Liu Li, Zhao Ling-Jun, Guo Cheng-Yu, Wang Liang, Tang Jun. Texture classification: state-of-the-art methods and prospects. *Acta Automatica Sinica*, 2018, **44**(4): 584–607  
(刘丽, 赵凌君, 郭承玉, 王亮, 汤俊. 图像纹理分类方法研究进展和展望. 自动化学报, 2018, **44**(4): 584–607)
- 27 Luo Jian-Hao, Wu Jian-Xin. A survey on fine-grained image categorization using deep convolutional features. *Acta Automatica Sinica*, 2017, **43**(8): 1306–1318  
(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. 自动化学报, 2017, **43**(8): 1306–1318)
- 28 Yu Fisher, Vladlen Koltun, Thomas Funkhouser. Dilated residual networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 636–644

- 29 Zhou Y Z, Ye Q X, Qiu Q, Jiao J B. Oriented response networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 4961–4970
- 30 Jeon Y, Kim J. Active convolution: Learning the shape of convolution for image classification. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 1846–1854
- 31 He K M, Gkioxari G, Dollar P. Mask r-cnn. In: Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017. 2980–2988
- 32 Li Ce, Zhang Ya-Chao, Lan Tian, Du Shao-Yi. An object detection algorithm with visual perception for high-resolution remote sensing images. *Journal of Xi'an Jiaotong University*, 2018, **6**(52): 9–16  
(李策, 张亚超, 蓝天, 杜少毅. 一种高分辨率遥感图像视觉感知目标检测算法. *西安交通大学学报*, 2018, **6**(52): 9–16)
- 33 Kim S W, Kook H K. Parallel feature pyramid network for object detection. In: Proceedings of the 15th European Conference on Computer Vision, Munich, Germany: Springer Verlag, 2018. 234–250
- 34 Cai Z W, Nuno V. Cascade R-CNN: Delving into high quality object detection. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, USA: IEEE, 2018. 6154–6162
- 35 Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, California, USA: AAAI, 2017. 4278–4284
- 36 Xie S N, Girshick R, Dollar P. Aggregated residual transformations for deep neural networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017. 5987–5995
- 37 Zhou P, Ni B B. Scale-transferrable object detection. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, USA: IEEE, 2018. 528–537
- 38 Bharat Singh, Larry S. Davis. An Analysis of Scale Invariance in Object Detection SNIP. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, USA: IEEE, 2018. 3578–3587



李 策 工学博士, 兰州理工大学电气工程与信息工程学院教授. 主要研究方向为计算机视觉与模式识别, 智能机器人, 图像处理及应用. 本文通信作者.

E-mail: xjtllice@gmail.com

(**LI Ce** Ph.D., professor at the College of Electrical and Information Engineering, Lanzhou University of Technology.

His research interest covers computer vision and pattern recognition, intelligent robot, image processing and application. Corresponding author of this paper.)



张 栋 兰州理工大学电气工程与信息工程学院硕士研究生. 主要研究方向为计算机视觉与图像处理.

E-mail: dongzhangcv@gmail.com

(**ZHANG Dong** Master student at the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interest

covers computer vision and image processing.)



杜少毅 工学博士, 西安交通大学人工智能与机器人研究所教授. 主要研究方向为计算机视觉, 机器学习和模式识别.

E-mail: dushaoyi@mail.xjtu.edu.cn

(**DU Shao-Yi** Ph.D., professor at Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interest covers computer

vision, machine learning and pattern recognition.)



朱子重 兰州理工大学学院硕士研究生. 主要研究方向为计算机视觉与图像处理.

E-mail: zizhongzhu\_cv@163.com

(**ZHU Zi-Zhong** Master student at the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interest

covers computer vision and image processing.)



贾盛泽 兰州理工大学学院硕士研究生. 主要研究方向为计算机视觉与图像处理.

E-mail: jiasz0607@163.com

(**JIA Sheng-Ze** Master student at the College of Electrical and Information Engineering, Lanzhou University of Technology. His research interest

covers computer vision and image processing.)



曲延云 工学博士, 厦门大学信息科学与技术学院教授. 主要研究方向为模式识别, 计算机视觉和机器学习.

E-mail: yyqu@xmu.edu.cn

(**QU Yan-Yun** Ph.D., professor at the College of Information Science and Engineering, Xiamen University. Her research interest covers pattern recog-

nition, computer vision and machine learning.)