文章编号：1672-3961(2022)06-0115-08 DOI：10.6040/j.issn.1672-3961.0.2022.087

# 面向小样本目标检测任务的自适应特征重构算法



刘丁菠[1,2]，刘学艳[2,3]，于东然[2,4]，杨博[2,3]，李伟[5*]

（1.吉林大学软件学院，吉林 长春 130012；2.吉林大学符号计算与知识工程教育部重点实验室，吉林 长春 130012；3.吉林大学计算机科学与技术学院，吉林 长春 130012；4.吉林大学人工智能学院，吉林 长春 130012；5.吉林大学商学与管理学院，吉林 长春 130012）

**摘要**：针对目标检测任务中样本量不足时新类别检测性能变差的问题，提出面向小样本目标检测任务的自适应特征重构算法。该算法包含两个模块：基础类别特征偏移缓解模块，用于获取预训练阶段基础类别的特征方向；场景特征自适应约束模块，用于根据场景特征与各类别原型特征的相关性确定当前场景对于某些类别的偏好，从而自适应地调整基础类别偏移方向对实例特征的影响。试验结果表明，在 PASCAL VOC 和 MS COCO 数据集上，该模型对于小样本目标检测任务的检测能力均优于对比算法，在保证对于基础类别实例检测能力的基础上，对新类别的检测精度最高可分别提升 12.4%与 2.1%。本研究提出的模型可以保证对于基础类别相关实例的检测能力，并提升新类别实例检测性能。

**关键词**：目标检测；小样本学习；自适应特征重构；场景特征；类别偏好

**中图分类号**：TP391 **文献标志码**：A

**引用格式**：刘丁菠，刘学艳，于东然，等. 面向小样本目标检测任务的自适应特征重构算法[J].山东大学学报(工学版)，2022，52(6)：115-122.

LIU Dingbo，LIU Xueyan，YU Dongran，et al. Adaptive feature reconstruction algorithm for few-shot object detection[J]. Journal of Shandong University（Engineering Science），2022，52(6)：115-122.

# Adaptive feature reconstruction algorithm for few-shot object detection

LIU Dingbo[1,2]，LIU Xueyan[2,3]，YU Dongran[2,4]，YANG Bo[2,3]，LI Wei[5*]

（1. College of Software，Jilin University，Changchun 130012，Jilin，China；2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education，Jilin University，Changchun 130012，Jilin，China；3. College of Computer Science and Technology，Jilin University，Changchun 130012，Jilin，China；4. School of Artificial Intelligence，Jilin University，Changchun 130012，Jilin，China；5. School of Business and Management，Jilin University，Changchun 130012，Jilin，China）

**Abstract**：Aiming at the problem that the detection accuracy of novel categories decreased when the sample size was insufficient in object detection tasks，an adaptive feature reconstruction algorithm was proposed. The algorithm contained two modules，one of which was the base categories′ feature offset′s alleviation module，which obtained the offset direction of the base categories in the pre-training stage. The other was the scene feature adaptive constraint module，which learned the correlation between scene features and the prototype features of each category to determine the preference of the current scene for certain categories，and adaptively adjusted the influence of base categories offset direction on instance features. Experimental results showed that the performance of the proposed algorithm was superior to the comparison algorithms of few-shot object detection tasks on PASCAL VOC and MS COCO datasets. On the basis of ensuring the detection capability of the base categories′ instances，the detection accuracy of novel categories′ instances could be improved by 12.4% and 2.1%，respectively. The proposed model could guarantee the detection ability of instances of base categories and improve the detection performance of instances of novel categories.

**Keywords**：object detection；few-shot learning；adaptive feature reconstruction；scene feature；category preference

---

**收稿日期**：2022-03-04；**网络首发时间**：2022-11-09 16：59：39

**网络首发地址**：https：//kns.cnki.net/kcms/detail/37.1391.T.20221109.0912.006.html

**基金项目**：国家自然科学基金项目(62172185、61876069)；国家重点研发计划项目(2021ZD0112501、2021ZD0112502)；吉林省重点科技研发项目(20180201067GX、20180201044GX)；吉林省自然科学基金项目(20200201036JC)

**第一作者简介**：刘丁菠(1997— )，女，吉林珲春人，硕士研究生，主要研究方向为计算机视觉。E-mail：dbliujlu@126.com

*通信作者简介：李伟（1977— ），女，辽宁朝阳人，讲师，博士，主要研究方向为最优化理论。E-mail：W_li@jlu.edu.cn

## 0 引言

近年来由于卷积神经网络的兴起,目标检测技术已取得巨大发展,并成功应用于智慧交通、工业检测等多种具体场景中[1]。但深度学习相关方法往往需要大量训练样本,当样本不足时,易出现过拟合现象[2]。当前世界中,针对部分新类别,图像样本的获取以及标注都需要较高成本[3]。因此本研究希望找到一种解决办法,在某些类别样本量不足时,保证模型仍可具有较高检测精度。

面向数据样本分布不均衡的应用场景,研究者提出针对长尾问题与小样本问题的方法。长尾问题指数据的少量类别具有大量样本,但大量类别只包含少量样本的情况[4];小样本问题与长尾问题相似,但更侧重于评估模型在样本量不同的任务下的表现性能[5]。解决长尾任务的方案主要有重采样[6]、重加权[7]等,例如 Lin 等[7]提出的 FocalLoss 方法,以损失的形式降低大量简单负样本所占权重,使模型更多关注于少数分类困难的样本。关于小样本任务的工作早期主要是基于元学习的方法[8-10],该方法的主要思想是让模型学会学习,每次从基础类别中抽取不同任务训练模型,使模型最终能够泛化到由新类别组成的任务上。除元学习方法外,在小样本目标检测任务中,基于两阶段微调方式的模型结构往往更加简单,并且也可取得较好检测效果[11-13],例如 Wang 等[11]提出的两阶段微调模型(two-stage fine-tuning approach, TFA)在预训练阶段利用基础类别的大量实例对模型进行训练,使得模型能够对图像中类似于目标的物体具有较好的定位能力和特征生成能力;在微调阶段冻结候选框提取网络与特征生成网络,仅重新训练模型的分类层与回归层,使模型获得检测新类别的能力。

尽管上述工作已取得一些进展,但小样本目标检测任务仍然具有挑战性。由于基础类别中具有大量训练样本,在预训练阶段模型会学习到大量关于基础类别的知识[14],这导致在微调阶段,面对新类别实例时,模型将提取出更多基础类别相关特征。这种对于基础类别特征的偏移,导致模型不能重视新类别特征,影响实例在基础类别与新类别中的分类倾向,因而在微调阶段需要尽量消除偏移影响。但是由于这种偏移由基础类别相关特征组成,对于基础类别而言,如果对其直接消除,模型将无法提取出足够具有判别性的特征,削弱了模型对基础类别相关实例的表示能力,会使其检测性能下降。

针对上述挑战,本研究提出面向小样本目标检测问题的自适应特征重构算法。

本研究旨在引入场景特征[15],约束基础类别特征偏移方向,使其在保证基础类别检测性能的前提下,纠正因特征偏移引起的新类别实例表示问题。为了消除预训练过程与微调过程的鸿沟,在预训练过程中,本研究以实例特征融合的形式保存基础类别中具有代表性的特征,并获取基础类别特征偏移方向。在微调过程中,将图像整体特征作为场景特征,学习其与各类别特征的相关性关系,确定当前场景对于类别的偏好,以此重构实例特征,将重构后的实例特征作为基准特征,指导对于预训练过程中获取的基础类别特征偏移方向的约束力度,对实例特征进行自适应重构。本研究的主要工作如下。

(1)为缓解特征偏移对新类别实例检测的影响,本研究构建了基础类别特征偏移缓解模块。具体地,在预训练阶段,以实例特征融合方法保存基础类别实例中具有代表性的特征,以此获取基础类别的特征偏移方向。在微调阶段,以固定参数缓解由特征偏移带来的影响。

(2) 为了进一步实现不同图像实例对于特征重构力度的不同需求,本研究构建了基于场景特征的自适应约束模型,在基础类别特征偏移缓解模型的基础上引入自适应机制。具体地,利用场景特征为每个实例独立计算偏移缓解权重,进而对各个实例的特征重构力度进行自适应调整。

(3) 在 PASCAL VOC 和 MS COCO 两个经典数据集上,针对小样本目标检测任务,对本模型进行验证。试验结果表明,本模型在以上数据集中可以取得较好效果。在多种任务中,本模型能够保证对于基础类别相关实例具有良好检测效果,同时普遍提高小样本目标检测精度。

## 1 方法

本研究提出了面向小样本目标检测任务的自适应特征重构算法,模型结构如图 1 所示。本章将具体介绍小样本目标检测问题的定义和该方法包含的两个模块:基础类别特征偏移缓解模块和场景特征自适应约束模块。

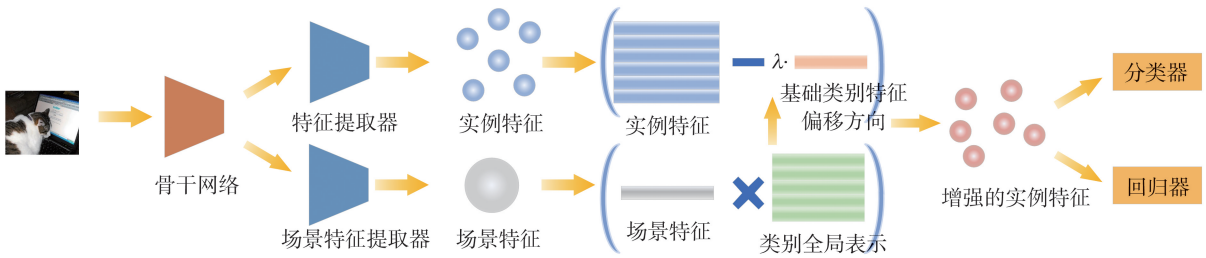



图 1　模型结构图

Fig.1　The structure of the proposed model

### 1.1　小样本目标检测问题定义

基于两阶段微调的小样本目标检测模型整体训练过程包含预训练与微调两阶段：预训练阶段数据集

$$\boldsymbol{D}_{\text{base}}=\{(\boldsymbol{x},\boldsymbol{y})\mid\boldsymbol{x}\in\boldsymbol{X},\boldsymbol{y}\in\boldsymbol{Y}\},$$

式中：$\boldsymbol{x}$ 为实例视觉特征；$\boldsymbol{y}=(c,\boldsymbol{l})$ 为标签信息，其中 $c\in\boldsymbol{C}_{\text{base}}$ 为实例 $\boldsymbol{x}$ 所属类别，$\boldsymbol{l}$ 为坐标位置信息；$\boldsymbol{X}$ 为实例集合；$\boldsymbol{Y}$ 为标签信息集合。微调阶段数据集

$$\boldsymbol{D}_{\text{train}}=\boldsymbol{D}_{\text{novel}}+\boldsymbol{D}'_{\text{base}},$$

式中：子数据集 $\boldsymbol{D}_{\text{novel}}$ 由新类别 $\boldsymbol{C}_{\text{novel}}$ 所有实例组成，在小样本问题中，每个新类别都只有 $t$ 个实例可见并参与训练；$\boldsymbol{D}'_{\text{base}}$ 由从每个基础类别中各自采样出的 $t$ 个实例组成。

### 1.2　基础类别特征偏移缓解模块

相对于新类别而言，基础类别具有大量训练样本。在预训练过程结束后，特征提取器倾向于提取更多与基础类别相关的特征。因此在微调阶段可能会忽视新类别之中更具有判别性的特征。纠正由基础类别实例特征引起的特征偏移问题，可以更好地检测出新类别相关实例。

在预训练过程中，首先通过实例特征融合保存基础类别中真实值所对应的具有代表性的特征。具体地，在目标检测器第一阶段结束后，获取各个建议区域的位置信息，一张图像可以获得 $n$ 个建议区域。基于这些建议区域的位置信息，使用特征提取器提取出各个建议区域的视觉特征 $\boldsymbol{x}_{ij}$，针对其中正样本进行进一步特征融合，获取基础类别特征偏移方向

$$z=\frac{1}{N\times n}\sum_{i\in(1,N)}\sum_{j\in(1,n)}\boldsymbol{x}_{ij},\tag{1}$$

式中：$N$ 为训练集中图像数量，$n$ 为当前图像中正样本实例的数量，$i$ 为图像编号，$j$ 为图像中正样本实例编号，$\boldsymbol{x}_{ij}$ 为对象实例的视觉特征，$z\in\mathbf{R}^{1\times d}$ 为模型对于基础类别特征的倾向性程度。特征提取的具体流程如图 2 所示。

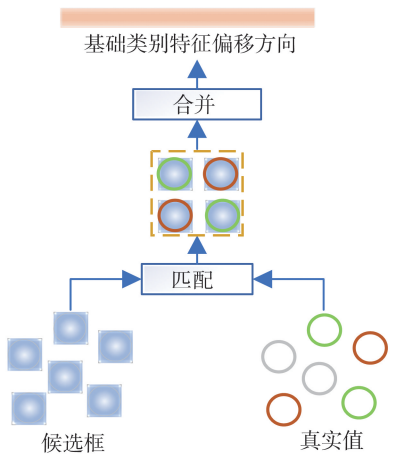



图 2　基础类别特征方向提取

Fig.2　Base feature offset direction extraction

模型在预训练阶段获取基础类别特征偏移方向之后，在微调阶段去除偏移方向的影响，具体为

$$\boldsymbol{x}'_{ij}=\boldsymbol{x}_{ij}-\lambda z,\tag{2}$$

式中：$\lambda$ 为基础类别特征偏移方向约束权重，$\lambda$ 越大表示对于特征偏移有越大的约束；进行特征偏移约束之后，得到了重构后的实例特征 $\boldsymbol{x}'_{ij}$。此特征缓解了面向基础类别特征方向的偏移问题，使分类器更加关注于新类别特异的特征。

### 1.3　场景特征自适应约束模块

1.2 节主要对基础类别特征偏移方向提取过程进行分析，对其进行约束可以降低实例数量不平衡对新类别检测性能的影响。上述方案中权重 $\lambda$ 为人为定义，虽然基础类别特征偏移方向对新类别检测有阻碍作用，但它却是基础类别实例特征的组成部分，以固定权重参数进行偏移约束可能会导致基础类别相关实例检测性能下降。

场景特征为图像全局信息[15]，因其蕴含更多的语义知识，可用于辅助学习实例与类别特征向量之间的相关性关系，进一步为实例类别的划分提供帮助。因此，本节引入场景特征自适应约束模块，为基础类别特征偏移约束权重进行自适应调整，进一步满足不同图像实例对于特征重构力度的不同需求。

图像经过特征金字塔网络（feature pyramid net-

work, FPN)后,可以获取 5 层全局特征,每层特征分别指示了图像在不同尺度下的信息。采用插值思想将 5 层特征统一到相同维度,维度的选择采用折中的形式并以第 2 层特征的尺度为准。特征插值部分

$$\boldsymbol{f}_{ik}=\text{interpolate}\left(\boldsymbol{f}_{ik},\boldsymbol{v}=\left(\frac{w_k}{w_2},\frac{h_k}{h_2}\right)\right),\tag{3}$$

式中:变量 $\boldsymbol{v}$ 指示了将特征金字塔每一层特征进行插值之后的尺度变化,$w$ 与 $h$ 为特征图的宽和高,$\boldsymbol{f}_{ik}$ 为特征金字塔中第 $k$ 层特征。进行特征插值后,分别对每层特征进行处理并取平均,从而得到场景特征

$$\boldsymbol{s}_i=\text{mean}(fc(\text{relu}(\text{conv}(\boldsymbol{f}_{ik})))),\tag{4}$$

式中:$\boldsymbol{s}_i\in\mathbf{R}^{1\times d}$ 为第 $i$ 张图像的场景特征;conv(·)为卷积操作;为了使场景特征稀疏同时避免出现梯度消失问题,使用 relu(·)作为激活函数;在进行激活操作后使用全连接层 $fc$(·)将特征映射到特定维度 $d$,最后对特征金字塔中各部分取均值,得到一维场景特征 $\boldsymbol{s}_i$。计算场景关于各个类别的偏好

$$\boldsymbol{w}_i=\boldsymbol{s}_i\times\boldsymbol{t}^{\mathrm{T}},\tag{5}$$

式中:$\boldsymbol{w}_i\in\mathbf{R}^{1\times m}$ 为场景 $\boldsymbol{s}_i$ 关于各个类别的偏好,其中 $m$ 为类别数量;$\boldsymbol{t}\in\mathbf{R}^{m\times d}$ 为各类别的全局表示,本研究选取类别特征向量作为类别的全局表示;基于场景偏好与实例偏好获得融合两者信息的实例特征向量,具体公式为

$$\boldsymbol{r}_{ij}=\boldsymbol{h}_{ij}\times(\boldsymbol{w}_i\times\boldsymbol{c}),\tag{6}$$

式中:$\boldsymbol{r}_{ij}\in\mathbf{R}^{1\times d}$ 为融合场景偏好与实例偏好的实例特征向量,用以指示与基础类别偏移方向 $\boldsymbol{z}$ 的相似度,相似度高则表示实例靠近基础类别,反之表示实例靠近新类别;$\boldsymbol{c}\in\mathbf{R}^{m\times d}$ 为类别权重矩阵;⊗为向量元素与矩阵对应位置加权相乘;$\boldsymbol{h}_{ij}\in\mathbf{R}^{1\times m}$ 为实例 $j$ 关于类别的偏好。流程图如图 3 所示。

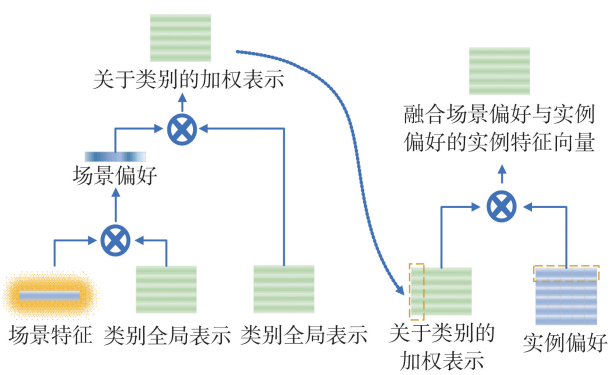



图 3 场景特征融合

Fig.3 Scene feature fusion

为平衡特征偏移方向对基础类别相关实例与新类别相关实例的影响,特征偏移权重的变化趋势与相似度趋势相反,基于融合场景信息的实例特征向量 $\boldsymbol{r}_{ij}$ 与基础类别偏移方向 $\boldsymbol{z}$ 的相似度得到融合场景信息的权重

$$\omega_{ij}=1-\text{sim}(\boldsymbol{r}_{ij},\boldsymbol{z}),\tag{7}$$

式中:sim(·)为相似度计算,本研究采用余弦相似度作为计算方式。将 $\omega_{ij}$ 代替公式(2)中的衰减权重 $\lambda$,进行目标特征自适应重构,得到重构后的实例特征

$$\boldsymbol{x}'_{ij}=\boldsymbol{x}_{ij}-\omega_{ij}\times\boldsymbol{z},\tag{8}$$

式中:$\boldsymbol{x}'_{ij}$ 为对实例 $j$ 进行自适应特征重构后得到的特征。利用该方法对不同实例进行自适应特征重构,会使得模型更加关注当前场景中更可能出现的类别。

### 1.4 损失函数

在对实例特征进行重构之后,通过计算实例特征与类别表示之间的余弦相似度,得出当前实例所属类别

$$\boldsymbol{p}_{ij}=\beta\times\frac{\boldsymbol{x}'^{\mathrm{T}}_{ij}\times\boldsymbol{c}}{\left|\left|\boldsymbol{x}'_{ij}\right|\right|\times\left|\left|\boldsymbol{c}\right|\right|},\tag{9}$$

式中:$\boldsymbol{p}_{ij}\in\mathbf{R}^{1\times d}$ 为图像 $i$ 中实例 $j$ 的预测值,指示了实例 $j$ 属于某类别的概率;$\beta$ 为余弦相似度参数,本研究将其设置为 20。

由于在微调阶段,RPN 网络处于冻结状态,因此该阶段损失函数[16]

$$L=L_{\text{reg}}+L_{\text{cls}},\tag{10}$$

式中:$L_{\text{reg}}$ 为回归损失,使用平均绝对误差作为回归损失函数;$L_{\text{cls}}$ 为分类损失,使用交叉熵作为分类损失函数,具体为

$$L_{\text{cls}}=-\frac{1}{N\times n}\sum_{i=1}^{N}\sum_{j=1}^{n}\boldsymbol{y}_{ij}\log(\boldsymbol{p}_{ij})。\tag{11}$$

### 1.5 学习策略

本研究遵循文献[11]中的训练模式,将整个训练过程分为预训练阶段和模型微调阶段。在预训练阶段,根据基础类别中大量训练样本训练模型,并获取基础类别特征偏移方向。在模型微调阶段,根据从基础类别中采样的部分数据与新类别的所有可见数据微调模型,并冻结骨干网络、区域提取网络以及特征提取器,同时在训练过程中引入场景特征自适应约束模块进行特征自适应重构。由于微调过程中需要训练的参数变少,因此,相比于预训练阶段,微调阶段的迭代次数更少。

## 2　试验与结果分析

为评估模型的有效性,在 PASCAL VOC 和 MS COCO 两个经典数据集上,针对小样本目标检测任务,对本研究进行验证,并针对各模块进行消融试验分析。

### 2.1　试验准备

#### 2.1.1　数据准备

本研究所选用的数据集为目标检测任务经典数据集,分别为 PASCAL VOC 数据集[17]与 MS COCO 数据集[18]。为公平比较,本研究与文献[8]采用相同数据划分方式。具体地,对于 VOC 数据集,本研究进行 3 组试验,在每组试验中,随机选取 5 个类别作为新类别,将其余 15 个类别作为基础类别。针对每个新类别,从 VOC07 与 VOC12 的训练集与验证集中采样$t$($t$=1、2、3、5、10)个样本作为可见数据训练模型,称为 $t$-shot 任务,使用 VOC07 的测试集作为测试数据。对于 COCO 数据集,选取与 VOC 数据集不相交的 60 个类别作为基础类别,将其余 20 个类别作为新类别,针对每个新类别,从 COCO 的训练集中采样 $t$($t$=10、30)个样本作为可见数据训练模型,使用验证集的 5000 张图像作为测试数据。

#### 2.1.2　系统配置与训练参数

本研究的系统配置为:操作系统,Ubuntu 16.04;中央处理器(CPU),Intel(R) Xeon(R) CPU E5-2643 v4@ 3.40 GHz;图形处理器(GPU),1080Ti。基于 MMDetection 框架[19],使用以 ResNet-101[20]与 FPN[21]为骨干网络的经典目标检测算法 Faster R-CNN[16]作为基准检测器实现本模型。使用动量随机梯度下降方法对模型进行优化,预训练阶段与微调阶段的学习率均为 0.005,动量为 0.9,权值衰减为 0.000 1。对所有实例进行 12 轮迭代训练,在每轮迭代中,设置批量大小为 2。

### 2.2　评价标准

针对 VOC 数据集,基于查准率(precision,P)与查全率(recall,R),本研究采用平均精度(average precision, AP)与平均精度均值(mean average precision, mAP)作为评价指标,具体为:

$$s_{\mathrm{P}}=s_{\mathrm{TP}}/(s_{\mathrm{TP}}+s_{\mathrm{FP}}),$$

$$s_{\mathrm{R}}=s_{\mathrm{TP}}/(s_{\mathrm{TP}}+s_{\mathrm{FN}}),$$

$$s_{\mathrm{AP}}=\int s_{\mathrm{P}}(s_{\mathrm{R}})\,\mathrm{d}s_{\mathrm{R}},$$

$$s_{\mathrm{mAP}}=\left(\sum_{m} s_{\mathrm{AP}}\right)/m,$$

式中:$s_{\mathrm{P}}$ 为查准率;$s_{\mathrm{R}}$ 为查全率;$s_{\mathrm{TP}}$为真阳性样本数;$s_{\mathrm{FP}}$为假阳性样本数;$s_{\mathrm{FN}}$为假阴性样本数;$s_{\mathrm{AP}}$为平均精度,具体为查准率与查全率所围成的曲线面积,用以衡量模型对于一个类别中所有样本的检测能力;$s_{\mathrm{mAP}}$为平均精度均值,衡量模型对于所有类别的检测能力。

针对 COCO 数据集,本研究采用多种评估指标衡量模型的检测能力。具体地,$s_{\mathrm{AP50}}$为将检测框占真实框面积 50%以上的样本作为真阳性样本时的平均精度,$s_{\mathrm{AP75}}$为将检测框占真实框面积 75%以上的样本作为真阳性样本时的平均精度,$s_{\mathrm{AP_S}}$用于衡量模型对于小尺度样本的检测能力,$s_{\mathrm{AP_M}}$用于衡量模型对于中尺度样本的检测能力,$s_{\mathrm{AP_L}}$用于衡量模型对于大尺度样本的检测能力。

### 2.3　试验结果

#### 2.3.1　VOC 数据集

为评估本模型的有效性,选取具有代表性的经典小样本目标检测算法:小样本目标检测任务特征重加权(few-shot object detection via feature reweighting, FSRW)[8]、MetDet[9]、Meta R-CNN[10]以及应用两阶段微调方式解决小样本目标检测问题的新算法 TFA[11]进行对比,表 1 为本模型与对比算法在 VOC 数据集的 3 组试验中,所有新类别的平均精度均值 $s_{\mathrm{mAP}}$。

表 1　本模型与对比算法在 VOC 数据集的不同分组方式下的平均精度均值

Table 1　The $s_{\mathrm{mAP}}$ of our model and comparison algorithms on VOC datasets in different groups　　%

| 方法 | 分组 1 | | | | | 分组 2 | | | | | 分组 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t$=1 | $t$=2 | $t$=3 | $t$=5 | $t$=10 | $t$=1 | $t$=2 | $t$=3 | $t$=5 | $t$=10 | $t$=1 | $t$=2 | $t$=3 | $t$=5 | $t$=10 |
| FSRW[8] | 14.8 | 15.5 | 26.7 | 33.9 | 47.2 | 15.7 | 15.3 | 22.7 | 30.1 | 40.5 | 21.3 | 25.6 | 28.4 | 42.8 | 45.9 |
| MetDet[9] | 18.9 | 20.6 | 30.2 | 36.8 | 49.6 | 21.8 | 23.1 | 27.8 | 31.7 | 43.0 | 20.6 | 23.9 | 29.4 | 43.9 | 44.1 |
| Meta R-CNN[10] | 19.9 | 25.5 | 35.0 | 45.7 | 51.5 | 10.4 | 19.4 | 29.6 | 34.8 | **45.4** | 14.3 | 18.2 | 27.5 | 41.2 | 48.1 |
| TFA[11] | **39.8** | 36.1 | 44.7 | **55.7** | 56.0 | **23.5** | **27.5** | 34.1 | **35.1** | 39.1 | 30.8 | 34.8 | 42.8 | **49.5** | 49.8 |
| 本模型 | 28.2 | **39.6** | **47.1** | 52.5 | **58.1** | 19.2 | 25.1 | **34.4** | 34.9 | 37.2 | **33.8** | **43.0** | **44.7** | 47.1 | **50.3** |
| TFA[a][11] | 25.3 | 36.4 | 42.1 | 47.9 | 52.8 | 18.3 | 27.5 | 30.9 | 34.1 | 39.5 | 17.9 | 27.2 | 34.3 | 40.8 | 45.6 |
| 本模型[a] | **27.3** | **40.1** | **45.6** | **52.3** | **58.1** | **19.2** | **27.6** | **32.1** | **36.4** | **42.9** | **30.3** | **38.6** | **41.8** | **45.8** | **49.7** |

[a] 表示按照 TFA 算法的数据划分方式,在微调过程中,随机选取 30 组可见实例作为训练数据运行得到的试验结果。

从表1可以看出，本模型在3个分组中都可以取得较好效果，检测精度最高可提升12.4%。在分组1和分组2的1-shot任务中，本模型没有得到特别满意的结果，这可能由于每个类别只有一个实例时，场景信息的获取过于依赖这一张图像，因此，在这种状况下的情景信息并不是很令人满意，不能很好地指导偏移缓解。表1的后两行表示在微调过程中，随机选取30组样本作为训练数据运行得到的试验结果，模拟模型在可见新类别实例质量未知下的表现结果。试验结果表明模型在不同情况的新类别实例的组合下，可以具有较稳定的提升。

表2为在VOC数据集的分组1中，3-shot任务与10-shot任务下，每个新类别的平均精度与所有新类别的平均精度均值。

表2 本模型与对比算法在VOC数据集的分组1下各新类别试验结果

Table 2 Experimental results of our model and comparison algorithms on VOC set1 for each novel category %

| 任务 | 方法 | 平均精度 | | | | | 平均精度均值 |
|---|---|---|---|---|---|---|---|
| | | bird | bus | cow | mbike | sofa | |
| $t=3$ | FSRW[8] | 26.1 | 19.1 | 40.7 | 20.4 | 27.1 | 26.7 |
| | FRCN+ft-full[10] | 29.1 | 34.1 | 55.9 | 28.6 | 16.1 | 32.8 |
| | Meta R-CNN[10] | 30.1 | 44.6 | 50.8 | 38.8 | 10.7 | 35.0 |
| | 本模型 | **35.7** | **48.5** | **58.6** | **48.7** | **44.0** | **47.1** |
| $t=10$ | FSRW[8] | 30.0 | 62.7 | 43.2 | 60.6 | 39.6 | 47.2 |
| | FRCN+ft-full[10] | 40.1 | 47.8 | 45.5 | 47.5 | 47.0 | 45.6 |
| | Meta R-CNN[10] | **52.5** | 55.9 | 52.7 | 54.6 | 41.6 | 51.5 |
| | 本模型 | 48.0 | **65.8** | **69.1** | **64.6** | **43.1** | **58.1** |

对比现有工作，无论是3-shot还是10-shot任务，本模型都在cow、mbike和bus这些类别中获得了更高的表现性能，在VOC数据集中，这些类别有更相似的基础类别，如sheep、bike和car，而试验结果表明本模型能够比现有工作更好地区分这些类别。

表3为在VOC数据集分组1中，3-shot任务与10-shot任务下，每个基础类别的平均精度与所有基础类别的平均精度均值。

表3 本模型与对比算法在VOC数据集的分组1下各基础类别试验结果

Table 3 Experimental results of our model and comparison algorithms on VOC set1 for each base category %

| 任务 | 方法 | 平均精度 | | | | | | | | | | | | | | | 平均精度均值 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | aero | bike | boat | bottle | car | cat | chair | table | dog | horse | person | plant | sheep | train | tv | |
| $t=3$ | FSRW[8] | 73.6 | 73.1 | 56.7 | 41.6 | 76.1 | 78.7 | 42.6 | 66.8 | 72.0 | 77.7 | 68.5 | 42.0 | 57.1 | 74.7 | 70.7 | 64.8 |
| | FRCN+ft-full[10] | 67.4 | 62.0 | 54.3 | 48.5 | 74.0 | 85.8 | 42.2 | 58.1 | 72.0 | 77.8 | 75.8 | 32.3 | 61.0 | 73.7 | 68.6 | 63.6 |
| | Meta R-CNN[10] | 67.6 | 70.5 | 59.8 | 50.0 | 75.7 | 81.4 | 44.9 | 57.7 | 76.3 | 74.9 | 76.9 | 34.7 | 58.7 | 74.7 | 67.8 | 64.8 |
| | 本模型 | **77.0** | **75.8** | **67.4** | **66.0** | **79.7** | **88.0** | **57.1** | **66.6** | **84.0** | **81.9** | **78.7** | **44.4** | **68.2** | **75.8** | **74.8** | **72.4** |
| $t=10$ | YLS[8] | 65.3 | 73.5 | 54.7 | 39.5 | 75.7 | 81.1 | 35.3 | 62.5 | 72.8 | 78.8 | 68.6 | 41.5 | 59.2 | 76.2 | 69.2 | 63.6 |
| | FRCN+ft-full[10] | 65.7 | 69.2 | 52.6 | 46.5 | 74.6 | 73.6 | 40.7 | 55.0 | 69.3 | 73.5 | 73.2 | 33.8 | 56.5 | 69.8 | 65.1 | 61.3 |
| | Meta R-CNN[10] | 68.1 | 73.9 | 59.8 | 54.2 | 80.1 | 82.9 | 48.8 | 62.8 | 80.1 | 81.4 | 77.2 | 37.2 | 65.7 | 75.8 | 70.6 | 67.9 |
| | 本模型 | **79.1** | **83.0** | **71.6** | **66.9** | **86.0** | **87.5** | **59.2** | **64.2** | **84.4** | **84.3** | **79.5** | **49.7** | **78.9** | **83.6** | **77.6** | **75.7** |

由于本模型利用场景信息同时考虑了新类别和基础类别相关实例，所以对于基础类别的检测性能也可以达到令人满意的效果。特别是与Meta R-CNN模型相比，各基础类别检测精度在3-shot任务时可提高7.6%，在10-shot任务时可提高7.8%。

### 2.3.2 COCO数据集

表4为本模型与对比算法在COCO数据集中，关于新类别的试验结果。

表4 本模型与对比算法在COCO数据集试验结果

Table 4 Experimental results of our model and comparison algorithms on COCO dataset %

| 任务 | 方法 | $s_{AP}$ | $s_{AP50}$ | $s_{AP75}$ | $s_{AP_S}$ | $s_{AP_M}$ | $s_{AP_L}$ |
|---|---|---|---|---|---|---|---|
| $t=10$ | FSRW[8] | 5.6 | 12.3 | 4.6 | 0.9 | 3.5 | 10.5 |
| | MetDet[9] | 7.1 | 14.6 | 6.1 | 1.0 | 4.1 | 12.2 |
| | Meta R-cnn[10] | 8.7 | 19.1 | 6.6 | 2.3 | 7.7 | 14.0 |
| | TFA[11] | 9.1 | 17.1 | 8.8 | | | |
| | 本模型 | **9.9** | **19.2** | **9.1** | **4.4** | **8.9** | **16.0** |

表4(续)

| 任务 | 方法 | $s_{AP}$ | $s_{AP50}$ | $s_{AP75}$ | $s_{AP_S}$ | $s_{AP_M}$ | $s_{AP_L}$ |
|---|---|---|---|---|---|---|---|
| | FSRW[8] | 9.1 | 19.0 | 7.6 | 0.8 | 4.9 | 16.8 |
| | MetDet[9] | 11.3 | 21.7 | 8.1 | 1.1 | 6.2 | 17.3 |
| $t=30$ | Meta R-cnn[10] | 12.4 | 25.3 | 10.8 | 2.8 | 11.6 | 19.0 |
| | TFA[11] | 12.1 | 22.0 | 12.0 | | | |
| | 本模型 | **14.1** | **26.5** | **13.3** | **6.1** | **13.9** | **20.9** |

从表 4 可以看出，在 10-shot 任务下 AP50 得分比 TFA 增加 2.1%，在 30-shot 任务下增加 4.5%。结果表明，在复杂度更高的 COCO 数据集上也可以达到较好效果。

### 2.4　消融试验

为验证各个组件对于模型性能的影响，对 VOC 数据集分组 1 进行了消融试验分析。

#### 2.4.1　基础类别特征偏移影响消除分析

为验证基础类别偏移方向对于新类别的影响，针对基础类别偏移影响消除模块以及其参数进行了分析，试验结果如表 5 所示。

表 5　基线方法与本模型在不同基础类别特征偏移影响消除力度下的平均精度均值

Table 5　The $s_{mAP}$ of baseline and our model on different base categories′ feature offset's alleviation degree　%

| 方法 | $t=2$ | $t=3$ | $t=10$ |
|---|---|---|---|
| 基线方法($\lambda=0$) | 37.84 | 44.67 | 57.09 |
| 本模型($\lambda=0.5$) | 38.49 | 45.20 | 57.21 |
| 本模型($\lambda=1.0$) | 38.29 | 45.62 | 57.86 |

从表 5 可以看出，在原始的基线方法基础上增加基础类别特征偏移影响消除模块之后，模型性能有所提升。这可能是因为移除了基础类别特征偏移方向之后，使模型特征的提取避免过多偏向具有大量实例的基础类别，在一定程度上验证了基础类别特征偏移方向对新类别检测的负面影响，以及对偏移特征进行消减对模型具有提升作用。当消减权重参数设置为 0.5 和 1.0 时，在不同任务中对于模型具有不同程度的提升，因此对偏移权重进行自适应约束使之在不同情况下具有不同的消减权重是有必要的。

#### 2.4.2　场景特征自适应约束影响分析

为验证场景特征自适应约束模块对于模型的影响，针对其进行了分析，试验结果如表 6 所示。

表 6　场景特征自适应约束模块的平均精度均值

Table 6　The $s_{mAP}$ of scene feature adaptive constraint module　%

| 方法 | $t=2$ | $t=3$ | $t=10$ |
|---|---|---|---|
| 基线方法($\lambda=0$) | 37.84 | 44.67 | 57.09 |
| 本模型(自适应约束 $\lambda$) | **39.60** | **47.09** | **58.09** |

从表 6 可以看出，相对于不加任何约束的基线方法，检测效果有所提升，验证了场景信息对于目标检测模型的分类任务具有一定的提升作用。场景特征作为图像整体特征蕴含了丰富的类别模式，例如室内场景与户外场景通常包含不同的类别语义信息，不同的场景会对不同的类别有所偏好，通过场景对于类别的偏好，建模其与实例之间的关联，从而为每个实例都自适应地进行基础类别特征偏移方向消减。

### 2.5　可视化

为进一步检验本模型的有效性，对检测效果进行可视化分析，图 4 和图 5 分别为基线方法与本模型在 VOC 数据集分组 1 的 10-shot 任务的检测性能，可以看出考虑到特征偏移和场景特征监督下的重建，本模型可以纠正预测误差的实例并使结果更具有针对性。



(a) 鸟　　(b) 公交车　　(c) 牛

图 4　基线方法检测结果的可视化

Fig.4　Detection results of baseline method



(a) 鸟　　(b) 公交车　　(c) 牛

图 5　本模型检测结果的可视化

Fig.5　Detection results of the proposed method

## 3　结语

本研究提出面向小样本目标检测问题的自适应特征重构算法，旨在引入场景特征，对基础类别特征偏移方向进行约束，使在保证基础类别检测性能前提下，纠正由特征偏移引起的新类别实例的表示问题。为进一步验证本模型的检测能力，分别在 VOC 和 COCO 两个数据集上完成了试验验证，并分析了试验结果，在小样本目标检测任务的性能均

优于对比算法，检测精度最高可分别提升12.4%与2.1%，试验结果表明，本模型在上述两个数据集都可以取得良好效果，能够在保证基础类别检测性能的前提下，提升对于新类别相关实例的检测能力。

**参考文献：**


[1] 谢富，朱定局.深度学习目标检测方法综述[J].计算机系统应用，2022，31(2)：1-12.
XIE Fu，ZHU Dingju. Survey on deep learning object detection[J]. Computer Systems & Applications，2022，31(2)：1-12.

[2] HU H，BAI S，LI A，et al. Dense relation distillation with context-aware aggregation for few-shot object detection[C]// Proceedings of CVPR-21. Nashville，USA：IEEE，2021：10185-10194.

[3] TAN M，PANG R，LE Q V. Efficientdet：scalable and efficient object detection[C]//Proceedings of CVPR-20. Seattle，USA：IEEE，2020：10781-10790.

[4] ZHANG Y，KANG B，HOOI B，et al. Deep long-tailed learning：a survey[EB/OL]. (2021-10-09)[2022-02-25]. https://arxiv.org/pdf/2110.04596.pdf.

[5] KÖHLER M，EISENBACH M，GROSS H M. Few-shot object detection：a survey[EB/OL]. (2021-12-22)[2022-02-25].https://arxiv.org/pdf/2112.11699.pdf.

[6] BUDA M，MAKI A，MAZUROWSKI M A. A systematic study of the class imbalance problem in convolutional neural networks[J]. Neural Networks，2018，106：249-259.

[7] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，42(2)：318-327.

[8] KANG B，LIU Z，WANG X，et al. Few-shot object detection via feature reweighting[C]// Proceedings of ICCV-19. Seoul，Korea：IEEE，2019：8420-8429.

[9] WANG Y X，RAMANAN D，HEBERT M. Meta-learning to detect rare objects[C]// Proceedings of ICCV-19. Seoul，Korea：IEEE，2019：9925-9934.

[10] YAN X，CHEN Z，XU A，et al. Meta r-cnn：towards general solver for instance-level low-shot learning[C]// Proceedings of ICCV-19. Seoul，Korea：IEEE，2019：9577-9586.

[11] WANG X，HUANG T E，DARRELL T，et al. Frustratingly simple few-shot object detection[C]// Proceedings of ICML-20. Online：ACM，2022：9919-9928.

[12] WU J，LIU S，HUANG D，et al. Multi-scale positive sample refinement for few-shot object detection[C]// Proceedings of ECCV-20. Online：Springer，2020：456-472.

[13] SUN B，LI B，CAI S，et al. FSCE：few-shot object detection via contrastive proposal encoding[C]// Proceedings of CVPR-21. Nashville，USA：IEEE，2021：7352-7362.

[14] TANG K，HUANG J，ZHANG H. Long-tailed classification by keeping the good and removing the bad momentum causal effect[J]. Advances in Neural Information Processing Systems，2020，33：1513-1524.

[15] XU H，JIANG C，LIANG X，et al. Reasoning-rcnn：Unifying adaptive global reasoning into large-scale object detection[C]//Proceedings of CVPR-19. Long Beach，USA：IEEE，2019：6419-6428.

[16] REN S，HE K，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] EVERINGHAM M，ZISSERMAN A，WILLIAMS C K I，et al. The PASCAL visual object classes (VOC) challenge[J]. International Journal of Computer Vision，2010，88(2)：303-338.

[18] LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：common objects in context[C]//Proceedings of ECCV-14. Zurich，Switzerland：Springer，2014：740-755.

[19] CHEN K，WANG J，PANG J，et al. MMDetection：open mmlab detection toolbox and benchmark[EB/OL]. (2019-06-17)[2022-02-25]. https://arxiv.org/pdf/1906.07155.pdf.

[20] HEK，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]//Proceedings of CVPR-16. Las Vegas，USA：IEEE，2016：770-778.

[21] LIN T Y，DOLLÁR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]//Proceedings of CVPR-17. Honolulu，USA：IEEE，2017：2117-2125.


（编辑：郭少华）