

# 基于聚合重构数据增强和多维特征知识转移的小样本目标检测

吴家骏,王 一,朱松豪

(南京邮电大学 自动化学院,南京 210023)

E-mail: zhush@njupt.edu.cn

**摘要:**近年来,目标检测获得了广泛关注和研究,并取得许多成果.然而,要想获得一个性能优越的检测模型,需要大量标签样本进行训练.与之形成鲜明对比的是,人类仅需借助少量示例,就能快速学习新知识.为缩小两者间的差距,小样本目标检测得到越来越多关注.小样本目标检测方法旨在通过数量有限的标签样本,实现新类知识学习,且在此过程中,不会灾难性遗忘先前学习的基类知识,进而提升新类检测性能.然而,现有的小样本目标检测方法存在以下问题:1) 过度关注模型精度而忽略了模型效率;2) 只关注模型分类性能而忽略了模型定位性能.为解决这些问题,本文提出了一种新颖的基于聚合重构数据增强和多维特征知识转移的小样本目标检测方法.具体而言,首先提出聚合重构数据增强策略,通过从生成图像中提取目标对象,进行缩放后聚合在随机选择的基类样本中,从而在增加数据多样性、缓解数据稀缺的同时,增强模型对于不同数据集的泛化能力.然后,进行类间语义特征知识转移,实现分类器权值理想初始化,提高模型收敛速度;并显式建模类间定位特征知识,提高模型定位能力.实验结果表明,本文方法在小样本目标检测任务中表现良好,与现有方法相比具有一定的竞争力.

**关键词:**小样本目标检测;微调;多维特征知识转移;聚合重构数据增强

中图分类号: TP311

文献标识码: A

文章编号: 1000-1220(2026)05-1190-08

## Few-shot Object Detection Via Aggregation-reconstruction Data Augmentation and Multidimensional Feature Knowledge Transfer

WU Jiajun, WANG Yi, ZHU Songhao

(College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

**Abstract:** In recent years, object detection has received widespread attention and research and has achieved many results. However, to obtain a high-performance detection model, a large number of labeled samples are required for training. In sharp contrast, humans can quickly learn new knowledge with only few examples. To narrow the gap between these two, few-shot object detection has received increasing attention. The few-shot object detection method aims to achieve new class knowledge through a limited number of annotated samples, without catastrophically forgetting previously learned base class knowledge, thereby improving the performance of new class detection. However, existing few-shot object detection methods have the following problems: 1) Excessive focus on model accuracy while neglecting model efficiency; 2) Only focusing on model classification performance while neglecting model localization performance. To address these issues, this paper proposes a novel few-shot object detection method based on aggregation-reconstruction data augmentation and multidimensional feature knowledge transfer. Specifically, an aggregation reconstruction data augmentation strategy is proposed, which extracts specific objects from generated images, scales them, and aggregates them into randomly selected base class samples. This enhances the model's generalization ability to different datasets while increasing data diversity and alleviating data scarcity. Then, a semantic feature knowledge transfer strategy is proposed to achieve ideal initialization of classifier weights and improve model convergence speed, and a localization feature knowledge transfer strategy is proposed to improve the model's localization ability. The experimental results demonstrate that the proposed method performs well in few-shot object detection tasks and has certain competitiveness compared to existing methods.

**Keywords:** few-shot object detection; fine-tuning; multidimensional feature knowledge transfer; aggregation-reconstruction data enhancement

## 0 引言

近年来,深度学习的快速发展极大提高了目标检测性能.训练一个性能优越的目标检测模型需要大量标签的样本集,然而这样的条件在很多情况下是无法满足的.具体而言,传统

目标检测方法存在以下问题:1) 获取大量标签的样本集,需花费大量的人力、物力和财力;2) 学习样本特征的过程较为复杂,这是因为该过程需要同时进行目标定位和目标分类;3) 在样本稀少情况下,样本特征难以准确表征,从而导致模型检测性能不佳.因此,小样本目标检测成为计算机视觉领

域中非常重要的一个研究方向。

小样本目标检测旨在借助人对于新事物的认知机理,通过少量标签样本实现新类样本的识别及定位,使得模型在样本稀疏的情况下也能具有良好的泛化性能。小样本目标检测取得快速发展,并广泛应用于医学图像病理分析、工业产品质量检测、作物生长状况检测等领域。

目前,大多数小样本目标检测方法主要基于两阶段微调策略,这是因为两阶段微调策略可有效缓解训练过程中样本数量不平衡问题的同时,防止新类样本的过拟合及基类知识的灾难性遗忘,从而有效提高检测精度。

小样本目标检测中,样本的稀缺性以及假阳性会导致模型无法准确获取样本的特征知识,从而导致模型分类及定位性能下降。图像生成模型能够构建高质量样本,但在小样本目标检测任务中图像生成策略仍然处于探索阶段。其中,本文主要关注以下两点:1)如何生成任务相关的样本;2)如何从生成样本中提取高质量样本。因此,本文提出一种用于数据增强的聚合重构策略。首先,利用图像生成模型获取符合特征分布的样本;然后,对生成图像中的目标主体进行最小化区域框裁剪;最后,将获得的裁剪对象进行随机缩放,并聚合至随机选择的基类样本中构成新的样本。

为提高模型效率,本文提出语义特征知识转移策略。语义特征知识转移策略旨在对分类器设置理想的初始化参数,极大优化模型训练过程,提高模型收敛速度。将基类特征知识映射至新类,并利用这些特征知识对模型分类器参数进行初始化。

为提高模型定位性能,本文提出定位特征知识转移策略。定位特征知识转移策略旨在优化回归器参数,提高模型的定位性能,从而提升模型分类性能。计算基类与新类间的定位特征知识相似性,并利用获取的定位特征知识相似性作为标准化权重聚合所有边界框回归变量。

大量实验结果表明,本文所提方法可有效提高小样本目标检测精度。本文的主要贡献如下。

1)提出了一种基于聚合重构数据增强和多维特征知识转移的小样本目标检测方法,进一步提高模型性能。

2)提出了一种用于数据增强的聚合重构策略,通过提取生成图像中的目标对象,进行缩放后聚合在随机选择的基类样本中构成新样本,有助于模型捕获样本的特征细节。

3)提出了语义特征知识转移策略,利用基类样本学习的特征知识对分类器权值进行初始化,有助于提高模型效率,加快模型收敛速度。

4)提出了定位特征知识转移策略,将学习的类间定位特征知识进行显式建模,进一步挖掘类间差异性,提高模型定位能力。

## 1 相关工作

### 1.1 目标检测

目标检测作为计算机视觉领域的基础任务,同时也极具挑战性。它的核心目的是从图像中搜索感兴趣目标,并对感兴趣目标进行精准定位与准确分类。早期的目标检测方法依赖包括尺度不变特征变换、方向梯度直方图、Haar 特征等在内的

手工设计特征。随着深度学习的蓬勃发展,采用大量标签样本数据进行模型训练,实现目标对象特征表示自动提取,从而极大提高检测性能。基于深度学习的模型,将特征提取、目标分类、位置回归等环节整合,实现了端到端的训练与预测流程,极大优化了目标检测的实时性。目前,目标检测模型主要包含两种框架,即基于 RCNN 系列的两阶段检测框架<sup>[1]</sup>和基于 YOLO 系列的一阶段检测框架<sup>[2]</sup>。此外,文献[3]在保持原有方案的基础上,对感兴趣建议区域进行微调,然后对目标检测分类器进行微调;文献[4]使用小卷积滤波器预测边界框类别置信度和偏移量,并从不同尺度特征图中生成不同尺度的预测边界框,进一步提高模型性能。

### 1.2 小样本目标检测

一个性能优越的目标检测模型通常需要大量的标签样本进行训练,然而,这样的标签样本需要花费大量的人力物力和财力,现实中有时很难获取。因此,近年来小样本目标检测获得广泛关注,其旨在利用有限样本,实现对新知识学习。当前的小样本目标检测方法主要分为以下两类:1)基于元学习的方法;2)基于迁移学习的方法。元学习方法旨在从不同元任务中获取知识,迁移学习方法旨在将基类样本知识转移至新类样本。迁移学习方法分为预训练和微调两个阶段,在预训练过程中使用大量标签基类样本进行训练,接着使用基类和新类构成的平衡集对模型进行微调,且在微调过程中冻结网络参数,仅对分类器进行参数调整。

除以上两类方法外,研究人员也提出其它小样本目标检测方法。文献[5]通过类间语义信息,解决由样本稀缺和类别混淆导致的模型性能退化问题;文献[6]提出一种三阶段微调方法,通过引入额外的可塑性分类器微调阶段,缓解模型泛化能力不足问题;文献[7]通过增强底层任务相关特征表征,提升模型学习与任务相关知识的能力;文献[8]通过考虑前景与背景间的相关性增强合成效果,提高模型检测精度;文献[9]提出一种基于原型软标签的学习框架,有效地从测试数据中挖掘隐藏的新类实例,缓解新类样本稀缺问题;文献[10]提出一种基于元学习的多模态聚合特征模块,利用丰富的文本语义信息,缓解域间差异导致的模型性能下降问题。

### 1.3 数据增强

新类样本的稀缺性是导致小样本目标检测模型性能不佳的一个重要因素,因此,通过引入数据增强策略缓解新类样本稀缺问题,使得模型学习更多的类知识,提高模型对不同类的区分能力。目标检测任务中数据增强方法主要分为以下两类:1)包括颜色变换、高斯噪声等在内的不改变标签框的增强方式,如文献[11]利用强化学习和循环神经网络(Recurrent Neural Network, RNN)控制器进行数据增强,文献[12]利用贝叶斯方法进行数据增强;2)包括剪切变换、旋转变换、镜像变换等在内的改变标签框的增强方式,如 Cutout<sup>[13]</sup>、Mix-up<sup>[14]</sup>、Mosaic<sup>[15]</sup> 和 CutMix<sup>[16]</sup>。

数据增强对于小样本目标检测性能有着很大影响。一方面,通过生成更多图像,有效缓解新类样本稀缺;另一方面,通过样本多样性,学习更多特征知识细节,从而提高模型性能。因此,本文提出一种新的数据增强策略,该策略从生成图像中提取目标对象,并进行随机缩放后聚合至随机选择的基类样本图像,提高样本多样性的同时,提升模型检测精度。

## 2 本文方法

### 2.1 概述

本文方法包含以下两个阶段. 第1阶段是预训练阶段, 通过使用大量标签的基类样本对模型进行预训练. 第2阶段是微调阶段, 采用提出的聚合重构数据增强策略缓解新类样本数据稀缺, 并采用特征转移策略建立类别间的关联性, 降低计算复杂度, 提高模型效率以及分类准确性.

### 2.2 整体框架

小样本目标检测中的数据集包括基类数据集  $D_b = \{(x_i,$

$y_i)\}_{i=1}^I$  和新类数据集  $D_n = \{(x_j, y_j)\}_{j=1}^J$ , 其中  $x_i$  和  $x_j$  表示输入图像,  $y_i \in C_b$  和  $y_j \in C_n$  表示对应的类标签. 值得注意的是, 基类和新类互不相交, 表示为  $C_b \cap C_n = \Phi$ . 通常情况下, 每个基类样本包含足够的边界框标签, 而每个新类只有  $K$  (通常  $K$  小于等于 10) 个有限的样本实例. 小样本目标检测分为两个阶段, 第1阶段使用大量带标签的基类数据  $D_b$  进行模型的预训练, 第2阶段在冻结主干网络的基础上, 使用基类和新类构建的平衡集  $D_{few}$  进行分类器参数微调.

本文提出的基于聚合重构数据增强和特征知识转移的小样本目标检测方法的架构如图1所示. 在微调阶段, 首先利用

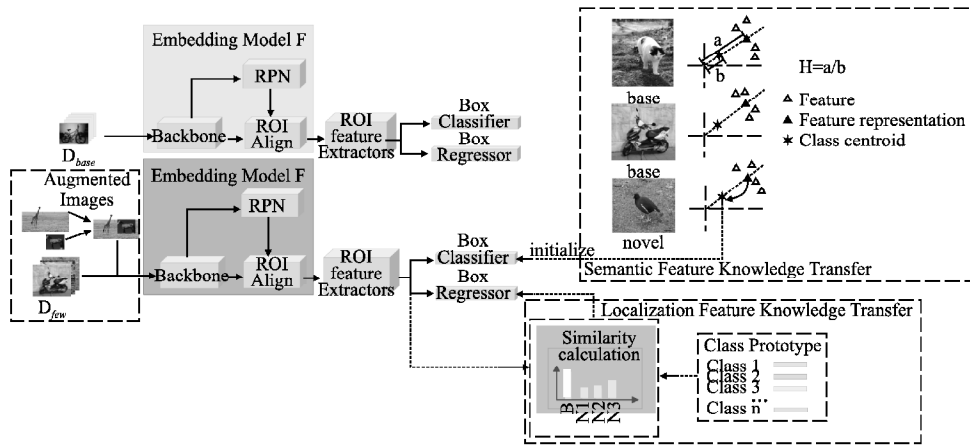


图1 本文所提方法的网络架构示意图

Fig. 1 Framework of the proposed few-shot object detection based on aggregation-reconstruction data augmentation and multidimensional feature knowledge transfer

提出的聚合重构数据增强增加数据多样性; 然后, 利用提出的语义特征知识转移策略, 初始化模型的分器, 提高模型效率; 最后, 利用提出的定位特征知识转移策略, 提高模型定位能力.

### 2.3 聚合重构数据增强

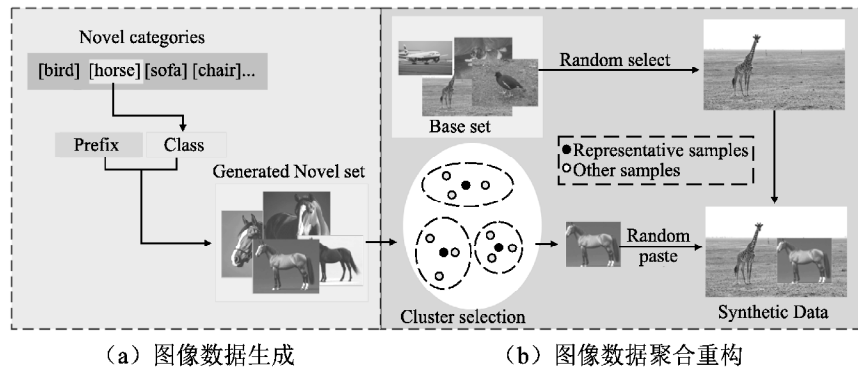
#### • 图像数据生成

目前大多数图像生成器采用类别名称加上前缀的形式, 例如, 可直接输入不带前缀的类别名, 可输入“A/One”作为前缀的类别名, 也可输入“A/One”结合“照片/图片”作为前

缀的类别名. 本文采用“A/One”结合“照片/图片”作为前缀的类别名的图像生成方式.

#### • 图像数据聚合重构

这一步骤旨在获得高质量、多样性且具有代表性的图像数据, 具体实现过程如图2所示. 第1步, 基于对比语言-图像预训练 (Contrastive Language-Image Pre-training, CLIP) [17] 的相似度评分, 筛选出高质量的新类生成图像. 第2步, 利用余弦相似度对生成的新类图像进行聚类, 获取代表性样本. 第3



(a) 图像数据生成

(b) 图像数据聚合重构

图2 聚合重构数据增强的一个示例

Fig. 2 An example of aggregation-reconstruction data augmentation

步, 利用最小封闭对代表性样本中的主对象进行裁剪. 最后, 将裁剪对象进行随机缩放后聚合至随机选择的基类样本中, 重构成新的样本数据. 在原有平衡集中引入重构图像, 构建新

的训练数据集.

### 2.4 语义特征知识转移

为使模型适应新任务, 避免模型负优化, 需对模型分类器

进行重新初始化. 一个理想的初始化状态有利于提高模型收敛速度、减少模型优化时间, 而不当的初始状态则会干扰优化过程、减慢收敛速度. 为缓解上述问题, 本文通过语义特征知识转移对模型分类器进行初始化, 为分类器初始化一个符合特征分布的权值. 由于预训练后的模型包含丰富的先验知识, 因此可以非常高效地预测新类质心.

语义特征知识转移的具体实现过程如下所述. 第 1 步, 将属于类别  $c$  的第  $k$  个基类样本图像标签对  $(x_k^c, y_k^c)$  输入嵌入式模型  $F$  获得特征  $f_k^c = F(x_k^c)$ . 第 2 步, 通过对多个特征进行类平均, 获取具有代表性的特征, 对应的公式如式(1)所示:

$$\mu_c = \frac{1}{n_c \times R} \sum_{r=1}^R \sum_{i=1}^{n_c} (f_i^c)_r \quad (1)$$

其中的  $(\cdot)_r$  表示第  $r$  个增强样本,  $n_c$  表示样本数量,  $R$  表示增强样本数量. 第 3 步, 也是非常重要的一步, 将在基类提取地知识映射至新类. 这是因为语义特征知识转移的优劣程度, 决定了分类器初始化权重是否合适. 由图 1 所示可知, 类别的特征和类别的质心位于不同长度的空间上, 也即某一基类特征表征的平均长度与该类质心的平均长度间存在一个如式(2)所示的比例关系  $H$ :

$$H = \frac{\mathbb{E}_{c \in C_b} \left[ \left\| \frac{1}{n_c \times R} \sum_{r=1}^R \sum_{i=1}^{n_c} (f_i^c)_r \right\|_2 \right]}{\mathbb{E}_{c \in C_b} \left[ \|\partial_c\|_2 \right]} \quad (2)$$

最后, 利用比例关系  $H$  对新类特征表征进行缩放, 得到新类质心  $\partial_{c'} = \mu_{c'}/H$ , 并利用得到的新类质心  $\partial_{c'}$  初始化分类器.

### 2.5 定位知识转移

在关注模型分类性能的同时, 模型的定位性能同样不可忽略, 这是因为有些类别间存在相似的外形及边界特征. 定位特征知识流程如图 3 所示. 模型利用丰富的基类样本学习分类知识的同时, 也能从中学习定位知识, 从而提高新类样本的检测精度.

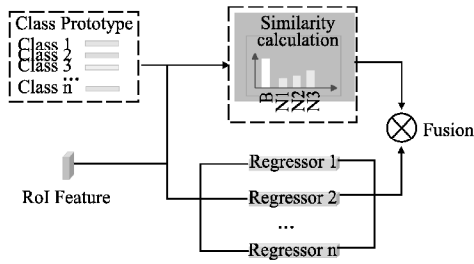


图 3 定位知识转移流程图.

Fig. 3 Flow chart of positioning knowledge transfer

定位知识转移具体实现过程如下所述. 第 1 步, 对利用图像生成器构成的新类样本候选数据集的特征向量进行类平均, 得到类别原型:

$$\eta_{c'} = \frac{1}{n_p} \sum_{i=1}^{n_p} r_i^{c'} \quad (3)$$

其中的  $r_i^{c'}$  表示属于类别  $c'$  的第  $i$  个新类样本的特征向量,  $n_p$  表示第  $i$  个新类样本候选数据集的数量. 第 2 步, 将基类样本特征和新类原型间的余弦距离定义为相似性, 表示为:

$$s_f^{c'} = \beta \cdot \frac{r_f^T \eta_{c'}}{\|r_f\| \|\eta_{c'}\|}, c' \in C_n \quad (4)$$

其中的  $C_n$  表示新类的类别数量,  $\beta$  表示正比例因子,  $r_f$  表示

基类样本的感兴趣区域特征. 第 3 步, 在关注基类与新类间相似性的同时, 也要关注基类内部的相似性. 假设不同基类样本间的距离可用如下所示的常数表示:

$$s_f^c = -\beta, c \in C_b \quad (5)$$

其中的  $C_b$  表示基类的类别数量.

最后, 利用 Softmax 函数对样本  $r$  与所有类的相似性进行归一化:

$$s_{f,c'}^{loc} = \frac{\exp(s_f^{c'})}{\sum_{c \in C_n \cup C_b} \exp(s_f^c)} \quad (6)$$

不仅可以利用边界框真值对目标进行定位, 也可利用相似类的边界框对目标进行定位, 因此将定位特征知识作为标准化权重, 聚集所有类的边界框回归变量进行目标定位. 在给定基类边界区域建议框的情况下, 预测边界框与真值边界框的偏移表示为  $u = (u_x, u_y, u_w, u_h)$ ; 然后, 通过使用平滑的 L1 损失最小化预测值和真值间的误差:

$$L_{loc} = \sum_{c'=1}^C s_{f,c'}^{loc} \cdot \sum_{i \in \{x,y,w,h\}} \text{Smooth}_{L1}(u_i^{c'} - g_i) \quad (7)$$

其中的  $s_{f,c'}^{loc}$  表示定位特征知识,  $g_i$  表示边界框真值,  $u_i^{c'}$  表示来自类别  $c'$  的第  $i$  个新类样本边界框的预测值. 最后, 通过聚集定位特征知识和 L1 损失优化分类器.

### 2.6 损失函数

在预训练阶段, 选择 Faster RCNN 作为骨干网络, 并使用大量基类样本进行训练, 相应的训练损失函数如式(8)所示:

$$L_{pre} = L_{rpn} + L_{cls} + L_{reg} \quad (8)$$

其中的  $L_{rpn}$  表示区分前景背景的损失函数,  $L_{cls}$  表示分类任务所采用的交叉熵损失函数,  $L_{reg}$  表示回归任务所采用的平滑 L1 损失函数.

在微调阶段, 使用定位特征知识转移, 相应的损失可用如公式(9)表示:

$$L = L_{pre} + \lambda_l L_{loc} \quad (9)$$

其中的  $\lambda_l$  平衡不同损失的超参数.

## 3 实验结果

### 3.1 数据集

选择两个广泛应用于评估小样本目标检测模型性能的公共数据集: PASCAL VOC<sup>[18]</sup> 和 MS COCO<sup>[19]</sup>. 为公平比较, PASCAL VOC 数据集中的所有类别被划分为 15 个基类和 5 个新类. 对于 MS COCO 数据集, 将与 PASCAL VOC 数据集重叠的 20 个类别划分为新类, 其它类别划分为基类. 每个类别分别选择 1 个、2 个、3 个、5 个和 10 个样本. 选择  $nAP$ 、 $nAP50$  和  $nAP75$  作为模型性能评价指标, 其中  $nAP$  表示新类在不同 IoU 阈值下的平均精度,  $nAP50$  和  $nAP75$  分别表示新类在 IoU 阈值为 0.5 和 0.75 时的精度.

### 3.2 实验细节

选择文献[20]中的 Faster-RCNN 作为本文所提小样本模型的骨干网络, 并选择 ResNet-101 作为在 ImageNet 上预训练的特征提取器. 本文方法的训练策略与所选的基线方法一致, 共进行 20000 次迭代, 使用动量为 0.9 的随机梯度下降优化器进行训练, 每次处理图像数量为 16, 权重衰减系数为  $1 \times e^{-4}$ , 增强样本数量  $R$  设为 10, 正比例因子  $\beta$  设为 5, 损失函数超参

$\lambda_i$  数设为 0.1. 预训练阶段,模型在两个数据集的初始学习率均设为 0.02;微调阶段,模型在 PASCAL VOC 数据集的初始学习率设为 0.02,在 MS COCO 数据集的初始学习率设为 0.01,且两个数据集均在微调阶段的 200 次迭代后进行定位特征知识转移.

### 3.3 实验结果

PASCAL VOC 的结果.表 1 给出本文方法分别在 PAS-

表 1 不同方法在 PASCAL VOC 数据集的以  $nAP50$  为评价指标的检测结果

Table 1 Detection results of different methods on PASCAL VOC with the evaluation metric of  $nAP50$

Method	Group 1					Group 2					Group 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
FSCE <sup>[8]</sup>	44.2	43.8	51.4	61.9	63.4	23.7	30.6	38.4	38.4	48.5	37.2	41.9	47.5	54.6	58.5
TFA w/cos <sup>[20]</sup>	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
MPSR <sup>[21]</sup>	41.7	43.1	51.4	55.2	61.8	22.4	29.3	39.2	39.9	47.8	35.6	40.6	42.3	48.0	49.7
SRR-FSOD <sup>[22]</sup>	47.8	50.5	51.3	55.2	56.8	32.5	35.3	39.1	40.8	43.8	40.1	41.5	44.3	46.9	46.4
UP-FSOD <sup>[23]</sup>	43.8	47.8	50.3	55.4	61.7	31.2	30.5	41.2	42.2	48.3	35.5	39.7	43.9	50.6	53.3
MLAD <sup>[24]</sup>	48.7	54.9	55.9	61.8	66.5	39.1	43.5	47.5	50.1	54.7	41.9	46.1	48.2	55.6	56.5
FSNA <sup>[25]</sup>	43.8	47.4	50.8	57.4	60.3	23.9	32.3	37.9	40.2	41.8	34.0	40.7	45.5	52.3	54.0
D&R <sup>[26]</sup>	41.0	51.7	55.7	61.8	65.4	30.7	39.0	42.5	46.6	51.7	37.9	47.1	51.7	56.8	59.5
Our Method	<b>49.1</b>	<b>55.6</b>	<b>57.1</b>	<b>62.9</b>	<b>67.1</b>	<b>38.9</b>	<b>42.1</b>	<b>48.3</b>	<b>51.7</b>	<b>56.1</b>	<b>44.1</b>	<b>47.5</b>	<b>52.7</b>	<b>59.5</b>	<b>61.5</b>

了样本特征细节;提出的语义特征知识转移策略,通过将基类知识转移至新类,建立类别间关联性;提出的定位特征知识转移策略,通过定位特征知识,显式建立类别间相关性.同时,由表 2 所示实验结果可知,本文方法在对新类检测性能达到最

表 2 不同方法在 PASCAL VOC 数据集组 1 上的基类和新类的检测结果

Table 2 Detection results of different methods on base class and new class in the first group of PASCAL VOC

Method	Group 1			
	3		10	
	$bAP$	$nAP$	$bAP$	$nAP$
FSCE <sup>[8]</sup>	74.2	50.1	76.6	61.7
TFA w/cos <sup>[20]</sup>	79.1	44.7	78.4	56.0
MPSR <sup>[21]</sup>	67.7	50.8	71.2	61.1
SRR-FSOD <sup>[22]</sup>	78.2	51.3	78.2	56.8
UP-FSOD <sup>[23]</sup>	66.3	50.3	69.7	61.7
FSDeView <sup>[27]</sup>	65.9	42.2	69.1	57.4
FRCN + ft-full <sup>[20]</sup>	66.1	29.0	66.0	45.5
Meta R-CNN <sup>[28]</sup>	64.8	35.0	67.9	51.5
Model trained on base class	81.0	-	81.0	-
Our Method	78.5	57.1	78.9	67.1

优的情况下,基类知识遗忘程度最小.在 PASCAL VOC 数据集的实验结果验证了本文所提的聚合重构数据增强策略和特征知识转移策略对于提高小样本目标检测性能的提升至关重要,MS COCO 数据集采用相同的超参数设置.

MS COCO 上的结果.表 3 给出本文所提方法在 MS COCO 数据集的以  $nAP$  和  $nAP75$  为性能评价指标的实验结果.与基线方法和其它小样本目标检测方法相比,本文所提方法具有显著的性能改进,其中在 10 个样本情况下的  $nAP$  和  $nAP75$  分别达到 14.7 和 15.2,比基线在 10 个样本情况下的  $nAP$  和  $nAP75$  分别提高了 4.7 和 5.9. MS COCO 数据集的实验结果表明,聚合重构数据增强策略和特征知识转移策略结合有助于模型获得更好的小样本目标检测效果.也即,数据增

CAL VOC 数据集的 3 个分组上的以  $nAP50$  为性能评价指标的实验结果,且由表 1 所示结果可得出以下结论:1)与基线方法<sup>[20]</sup>相比,本文方法具有显著的性能改进;2)与其它先进方法相比,本文方法同样表现出显著的小样本目标检测性能,特别在组 1 中,本文方法在 1 个样本和 2 个样本中的性能分别超过基线方法 9.3% 和 19.5%;3)以上对比实验结果证明,本文提出的聚合重构数据增强策略增加了样本的多样性,优化

强和特征知识转移相辅相成,可在更具挑战性的 MS COCO 数

表 3 不同方法在 COCO 数据集的实验结果

Table 3 Experimental results of different methods on COCO

Method	10-shot		30-shot	
	$AP$	$AP75$	$AP$	$AP75$
FSCE <sup>[8]</sup>	11.9	10.5	15.8	14.7
TFA <sup>[20]</sup>	10.0	9.3	13.7	13.4
MPSR <sup>[21]</sup>	9.8	9.7	14.1	14.2
AirDet <sup>[29]</sup>	13.0	12.4	-	-
TeSNet <sup>[30]</sup>	14.1	14.0	16.5	17.4
UNP <sup>[31]</sup>	12.3	11.5	15.3	14.8
Ours	14.7	15.2	16.7	17.5

据集上取得优异检测性能.

### 3.4 消融实验

语义特征知识转移消融实验.表 4 给出了在 PASCAL VOC 数据集上的消融实验结果,以验证语义特征知识转移对于模型检测性能的影响,其中的“w”和“wo”表示在训练过程中是否使用了语义特征知识转移.实验结果表明,迭代过程中的语义特征知识转移可显著提高模型性能,这验证了本文的以下假设:语义特征知识转移有助于提升模型精度.

定位特征知识转移消融实验.表 5 给出了在 PASCAL VOC 数据集上的消融实验结果,以验证定位特征知识转移策略对于模型检测性能的影响,其中的“w”和“wo”表示在训练过程中是否使用定位特征知识转移策略.实验结果表明,在训练过程中使用定位特征知识转移策略有利于将基类学习的知识映射至新类,优化模型参数并提升模型性能,这验证了本文的以下假设:定位特征知识转移在提高模型定位性能的同时,也能提升模型的检测精度.

聚合重构数据增强消融实验.表 6 给出了在 PASCAL VOC 数据集上的消融实验结果,以验证聚合重构数据增强对于模型检测性能的影响,其中的“w”和“wo”表示在训练过程中是否使用聚合重构数据增强.实验结果表明,与基线相比,

本文提出的聚合重构数据增强可显著提高模型性能,这验证了本文假设,即聚合重构数据增强有助于模型捕获不同类间

表 4 在 PASCAL VOC 数据集上的消融实验结果

Table 4 Ablation results of the semantic feature knowledge transfer strategy on PASCAL VOC

Method	Group 1					Group 2					Group 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
TFA w/cos	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
TFA w/cos + ST	45.5	52.9	53.5	58.8	63.2	34.5	36.5	44.1	48.1	51.7	40.6	44.5	48.7	55.1	57.2
Improvement	+5.7	+16.8	+8.8	+3.1	+6.8	+11.0	+9.6	+10.0	+13.0	+12.6	+9.8	+9.7	+5.9	+5.6	+7.4

注:“ST”代表语义特征知识转移策略。

表 5 在 PASCAL VOC 数据集上的消融实验结果

Table 5 Ablation results of the localization feature knowledge transfer strategy on PASCAL VOC

Method	Group 1					Group 2					Group 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
TFA w/cos + ST(wo LT)	45.5	52.9	53.5	58.8	63.2	34.5	36.5	44.1	48.1	51.7	40.6	44.5	48.7	55.1	57.2
TFA w/cos + ST(w LT)	46.9	53.1	54.5	59.2	63.9	35.9	37.1	45.2	49.0	52.9	41.1	45.2	49.6	56.5	58.1
Improvement	+1.4	+0.2	+1.0	+0.4	+0.7	+1.4	+0.6	+1.1	+0.9	+1.2	+0.5	+0.7	+0.9	+1.4	+0.9

注:“LT”代表定位特征知识转移。

表 6 在 PASCAL VOC 数据集上的聚合重构数据增强实验结果

Table 6 Ablation results of the aggregation-reconstruction data augmentation strategy on PASCAL VOC

Method	Group 1					Group 2					Group 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
+SL(wo Aug)	46.9	53.1	54.5	59.2	63.9	35.9	37.1	45.2	49.0	52.9	41.1	45.2	49.6	56.5	58.1
+SL(w Aug)	49.1	55.6	57.1	62.9	67.1	38.9	42.1	48.3	51.7	56.1	44.1	47.5	52.7	59.5	61.5
Improvement	+2.2	+3.5	+2.6	+3.7	+3.2	+3.0	+5.0	+3.1	+2.7	+3.2	+3.0	+2.3	+3.1	+3.0	+3.4

注:“SL”代表多维特征知识转移

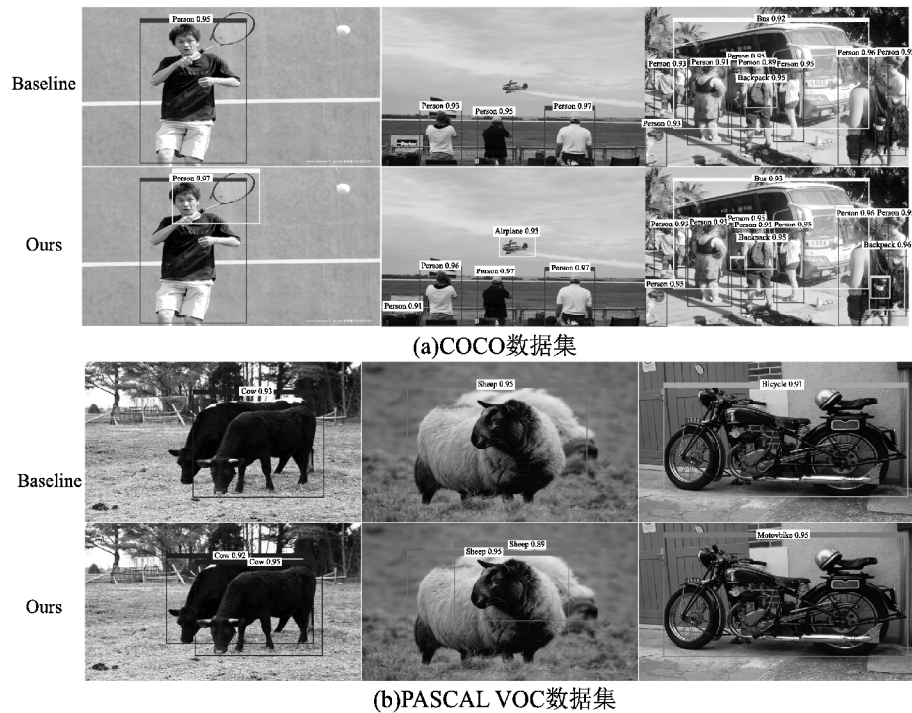


图 4 不同方法的可视化结果,(a)和(b)的上图为基线的检测结果,下图为本文方法的检测结果

Fig. 4 Visualization results of different methods, the upper figures of (a) and (b) are the detection results of the baseline, and the lower figures are the detection results of the proposed method

样本的特征细节,进一步学习类间关系,增加样本的多样性。

### 3.5 计算复杂度

计算复杂度是深度学习中一个非常重要的指标,在模型

的部署和应用过程中,它具有重要的参考价值,特别是在小样本目标检测中,模型效率显得尤为重要.模型的计算复杂度主要体现在浮点运算量、参数量、推理时间上.由表 7 所示实验

结果可知,相较于其它方法,本文所提模型的推理速度更快,参数量更少,这是因为语义特征知识转移为模型分类器设置

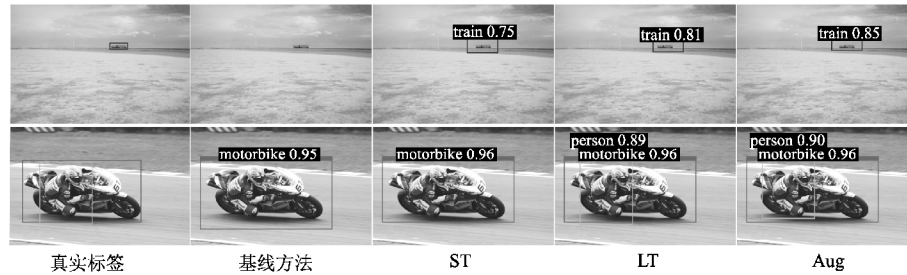


图5 消融实验可视化结果

Fig. 5 Visualization results of the ablation experiment

了理想的初始权重,从而加速了模型的收敛过程、降低了计算复杂度,这与本文的设计思路相一致.

表7 不同方法的计算复杂度

Table 7 Computational complexity of different methods

Method	Parameters	FLOPs	Time	
			Train [ h ]	Infer [ s ]
MPSR	82.12M	479.9G	1.503	0.086
FSCE	60.64M	252.3G	1.849	0.074
TFA w/cos	60.64M	252.3G	8.860	0.073
Ours	60.37M	252.2G	0.673	0.071

### 3.6 可视化结果

可视化实验结果能够直观展示所提方法的性能.图4给出 COCO 数据集上的3组实验对比结果,其中上图展示了使用基线方法获得的实验结果,而下图展示了使用本文方法获得的实验结果.红色矩形框和其它颜色矩形框分别表示基类和新类的检测结果.图5给出了消融实验可视化结果.可视化实验结果表明,本文方法可有效提高模型的准确性并增强模型的定位性能.

## 4 总结

小样本目标检测模型的性能受许多因素的影响,例如样本的稀缺性以及样本特征知识的提取方法等.因此,本文提出了一种新颖的小样本目标检测方法,该方法利用聚合重构数据增强策略,将新类生成图像主体聚合至基类样本图像中构成新样本,提高样本多样性;利用语义特征知识转移策略,将先前学习到的大量基类知识对分类器的新类权值进行初始化,优化模型收敛过程;利用定位特征知识转移策略,将基类与新类的定位特征知识显式建模,提高模型定位能力的同时提升模型检测精度.实验结果表明,聚合重构数据增强及类间特征知识转移有助于提升模型的效率及目标检测性能.虽然大量实验结果验证了本文提出的小样本目标检测方法的有效性,但本文方法并非解决小样本目标检测的最佳方法,例如,在进行图像裁剪粘贴时,可能会对基类样本中的关键信息遮挡.因此,未来工作将研究如何避免这个关键性问题.

### References:

- [1] Ross B Girshick. Fast R-CNN[C]//IEEE Conference on Computer Vision, 2015:1440-1448.
- [2] Joseph Redmon, Santosh Kumar Divvala, Ross B Greshick, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [3] Shaoqing Ren, Kaiming He, Ross B Girshick, 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.
- [4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, et al. SSD: single shot multibox detector [C]//European Conference on Computer Vision, 2016:21-37.
- [5] Zheng Wang, Yingjie Gao, Qingjie Liu, et al. Semantic enhanced few-shot object detection [J]. Computing Research Repository, 2024, 32(6):1-8.
- [6] Yuanhao Yin, Ping Yin. Stability plasticity decoupled fine-tuning for few-shot end-to-end object detection [J]. Computing Research Repository, 2024, 32(1):1-6.
- [7] Jie Mei, Mingyuan Jiu, Hichem Sahbi, et al. Few-shot object detection with sparse context transformers [J]. Computing Research Repository, 2024, 32(2):1-10.
- [8] Anh Khoa Nguyen Vu, Quoc Truong Truong, Vinh Tiep Nguyen, et al. Multi-perspective data augmentation for few-shot object detection [J]. Computing Research Repository, 2025, 33(2):1-19.
- [9] Yingjie Gao, Yanan Zhang, Ziyue Huang, et al. PS-TTL: prototype-based soft-labels and test-time learning for few-shot object detection [C]//ACM Conference on Multimedia, 2024:8691-8700.
- [10] Zeyu Shanguan, Daniel Seita, Mohammad Rostami. Cross-domain few-shot object detection with multi-modal textual enrichment [J]. Computing Research Repository, 2025, 33(2):1-41.
- [11] Ekin D Cubuk, Barret Zoph, Dandelion Mane, et al. Autoaugment: learning augmentation policies from data [J]. Computing Research Repository, 2018, 26(5):1-14.
- [12] Sungbin Lim, Ildoo Kim, Taesup Kim, et al. Fast autoaugment [C]//Annual Conference on Neural Information Processing Systems, 2019:6662-6672.
- [13] Terrance DeVries, Graham W Taylor. Improved regularization of convolutional neural networks with cutout [J]. Computing Research Repository, 2017, 25(8):1-8.
- [14] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, et al. MixUp: beyond empirical risk minimization [C]//International Conference on Learning Representations, 2018:1-13.
- [15] Alexey Bochkovskiy, Chienyao Wang, Hongyuan Mark Liao. Yolov4: optimal speed and accuracy of object detection [J]. Computing Research Repository, 2020, 26(4):1-17.

- [16] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, et al. Cutmix: regularization strategy to train strong classifiers with localizable features [C]//IEEE Conference on Computer Vision, 2019:6023-6032.
- [17] Alec Radford, Jong Wook Kim, Chris Hallacy, et al. Learning transferable visual models from natural language supervision [C]//International Conference on Machine Learning, 2021:8748-8763.
- [18] Mark Everingham, Luc Van Gool, Christopher K I Williams, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2):303-338.
- [19] Tsungyi Lin, Michael Maire, Serge J Belongie, et al. Microsoft COCO: common objects in context [C]//European Conference on Computer Vision, 2014:740-755.
- [20] Xin Wang, Thomas E Huang, Joseph Gonzalez, et al. Frustratingly simple few-shot object detection [C]//International Conference on Machine Learning, 2020:9919-9928.
- [21] Jiayi Wu, Songtao Liu, Di Huang, et al. Multi-scale positive sample refinement for few-shot object detection [C]//European Conference on Computer Vision, 2020:456-472.
- [22] Chenchen Zhu, Fangyi Chen, Uzair Ahmed, et al. Semantic relation reasoning for shot-stable few-shot object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2021:8782-8791.
- [23] Aming Wu, Yahong Han, Linchao Zhu, et al. Universal-prototype enhancing for few-shot object detection [C]//IEEE Conference on Computer Vision, 2021:9547-9556.
- [24] Songhao Zhu, Yi Wang. Multi-level similarity transfer and adaptive fusion data augmentation for few-shot object detection [J]. Visual Communication and Image Representation, 2024, 105(1):104-112.
- [25] Jinxiang Zhu, Qi Wang, Xinyu Dong, et al. FSNA: few-shot object detection via neighborhood information adaption and all attention [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(8):7121-7134.
- [26] Jiangmeng Li, Yanan Zhang, Wenwen Qiang, et al. Disentangle and remerge: interventional knowledge distillation for few-shot object detection from a conditional causal perspective [C]//Annual Conference on Neural Information Processing Systems, 2023:1323-1333.
- [27] Yang Xiao, Vincent Lepetit, Renaud Marlet. Few-shot object detection and viewpoint estimation for objects in the wild [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3):3090-3106.
- [28] Xiaopeng Yan, Ziliang Chen, Anni Xu, et al. Meta R-CNN: towards general solver for instance-level low-shot learning [C]//IEEE Conference on Computer Vision, 2019:9576-9585.
- [29] Bowen Li, Chen Wang, Pranay Reddy, et al. AirDet: few-shot detection without fine-tuning for autonomous exploration [C]//European Conference on Computer Vision, 2022:427-444.
- [30] Xiaowei Zhao, Xianglong Liu, Yuqing Ma, et al. Temporal specialization network for few-shot object detection [J]. IEEE Transactions on Multimedia, 2023, 25(1):8267-8278.
- [31] Bawei Yan, Chunbo Lang, Gong Cheng, et al. Understanding negative proposals in generic few-shot object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(7):5818-5829.