



多阶段特征重分布算法的小样本目标检测

刘露露, 贺占庄, 马钟*, 刘彬, 王莉

(西安微电子技术研究所 西安 710065)

【摘要】深度神经网络在目标检测任务上需要训练大量的标签数据, 然而在许多实际应用场景中标签数据难以获取。针对这一问题, 提出了一种面向小样本目标检测的多阶段特征重分布算法 (MSFR)。该算法通过对特征向量进行重分布变换, 解决了小样本任务下源域数据和目标域数据分布不一致的问题; 通过多阶段学习策略将源域知识逐步迁移到小样本目标任务中, 进一步提高知识迁移效率。在 VOC 数据集上的大量实验表明, 与现有小样本目标检测算法相比, 该算法在不同任务上的精度最高提升了 9.06%。该算法在大幅提高小样本目标域类别检测性能的同时, 较大限度地保持了对源域类别的检测精度, 具有较大的实用价值。

关键词 深度神经网络; 特征向量重分布; 小样本目标检测; 多阶段训练
中图分类号 TP391 文献标志码 A doi:10.12178/1001-0548.2022016

Multi-Stage Feature Redistribution for Few-Shot Object Detection

LIU Lulu, HE Zhanzhuang, MA Zhong*, LIU Bin, and WANG Li

(Xi'an Microelectronics Technology Institute Xi'an 710065)

Abstract Deep neural networks (DNN) in object detecting tasks have witnessed significant progress in the past years. However, it relies on intensive training data with accurate bounding box annotations for a remarkable performance. Once the labelled data are hard to catch, the generalization ability of DNN is far from satisfactory. We propose a few-shot object detecting method based on a multi-stage training strategy within feature redistribution (MSFR). Based on the analyses of the distribution of source domain dataset and target domain dataset in few-shot tasks, a feature redistribution algorithm is proposed to make the feature distribution meet Gaussian distribution or quasi-Gaussian distribution. It solves the inconsistency distribution of the source domain dataset and the target domain dataset. Then, a multi-stage training algorithm is proposed, which improves the efficiency of transferring the source domain knowledge to the target domain task when only a small amount of labeled data for training in each class. Thus, our proposed method significantly improves the detection performance of few-shot target domain categories while maximizing the detection accuracy of the source domain categories. The experimental results on VOC datasets show that the proposed algorithm achieves a precision improvement of up to 9.06% on different tasks, compared with existing few-shot object detection approaches.

Key words deep neural networks; feature redistribution; few-shot object detection; multi-stage training algorithm

近年来, 深度神经网络已广泛应用于多个计算机视觉任务, 如图像识别、目标检测等。深度神经网络的成功依赖于大量有标签的数据进行训练, 然而大量标签数据的获取在许多实际应用场景, 如军事、医疗等是非常昂贵、甚至不可能的。因此, 如何利用少量标签数据实现对目标任务的学习, 即小样本学习 (few-shot learning, FSL), 已成为目前深度学习的一个热点及难点。

根据网络学习方式不同, 目前小样本学习算法主要有以下几类: 数据增广、迁移学习、元学习和度量学习。基于数据增广的小样本学习算法利用数据增广策略以扩充样本数据, 主要包括在图像层进行增广和在图像特征层进行增广两种方式^[1-4]。基于数据增广的方法通过增加数据量的方式缓解样本不足的问题, 通常与其他小样本学习算法结合起来应用, 因此本文不做特别研究。基于迁移学习的小样

收稿日期: 2022-01-07; 修回日期: 2022-08-31

基金项目: 国家科技重大专项 (2017-V-0014-0066)

作者简介: 刘露露 (1991-), 男, 博士生, 主要从事深度学习、小样本学习等方面的研究。

*通信作者: 马钟, E-mail: mazhong@mail.com

本学习算法, 通过在大量训练数据下学习相关域的知识, 获取一个性能较好的初始化模型, 并将学习到的知识快速迁移到小样本任务中, 如 IFT^[5], TransMatch^[6] 及 LST^[7] 等。自文献 [8] 提出 MAML 算法后, 元学习算法就已广泛应用于小样本任务中^[9-12]。基于元学习的小样本学习算法是一种与模型无关的方法, 因而可以应用于任何采用基于梯度下降算法优化的模型中。元学习是一种特殊的迁移学习, 算法针对小样本任务对大量训练数据不断进行重采样, 以学习任务相关的元知识, 然后利用元学习的思想设计网络结构及训练模式, 将学习到的元知识快速泛化到小样本任务中, 如 NTM^[10]、MANN^[11] 及 Reptile^[12] 等。基于度量学习的小样本学习算法则是根据图像特征间的相似性程度实现对图像的分类, 算法通常结合元学习的思想利用大量数据训练一个特征提取网络, 然后对特征采取某种相似性度量方式来获取不同的度量表示, 如孪生网络^[13]、匹配网络^[14]、原型网络^[15] 及关系网络^[16] 等。

小样本学习的研究在近年来取得了不错的进展, 然而仍存在一些挑战: 1) 目前针对小样本学习的研究主要集中于图像分类任务, 对小样本目标检测任务的研究较少; 2) 小样本目标检测任务较之小样本图像分类任务有更大的挑战。现有小样本目标检测算法大部分借助于小样本图像分类算法, 如元学习^[17-20]、迁移学习^[21-24] 等来实现对目标的检测。然而这两种算法在小样本目标检测任务上均存在一些问题: 基于元学习的目标检测算法优化难度较大, 同时舍弃了对源域数据的检测能力; 基于迁移学习的算法需要源域数据和目标域数据有一定的相似性, 即小样本目标数据与大量训练数据有类似的分布, 相似性越高, 迁移效率越好。此外, 基于元学习和迁移学习的算法在面对小样本任务时均未考虑源域数据与小样本目标域数据间的分布差异, 知识迁移效率较差, 因而检测精度较低。

因此, 为了更好地将先验知识迁移到目标任务中, 提升小样本目标检测任务的检测性能, 同时对源域类别仍有较好的检测性能, 本文提出了一种多阶段特征向量重分布的小样本目标检测算法。针对特征分布不一致的问题, 本文提出了一种重分布变换算法, 通过对特征向量进行变换, 使得源域特征和小样本目标域特征有相同或相似的分布; 为了进一步提高域知识的迁移效率, 提出了一种多阶段逐

步微调的策略, 将知识逐步迁移到目标域中, 从而实现对小样本目标任务的高精度检测。

1 本文算法总体结构

1.1 小样本目标检测定义

小样本任务旨在从源域数据集 D_b 中学习域相关的知识, 并利用少量目标域数据集 D_s 将知识迁移到目标域任务中。 D_s 为目标任务的训练集, D_q 为目标任务的测试集, 用于衡量算法对小样本任务的泛化能力。因此, 小样本任务的数据集 D 包含3个部分: D_b , D_s 和 D_q 。其中, D_s 和 D_q 共同组成小样本目标任务的目标域数据集 D_n 。 D_b 、 D_s 和 D_q 定义为:

$$\begin{aligned} D_n &= D_s \cup D_q & D &= D_b \cup D_n & C &= C_b \cup C_n \\ D &= \{(x, y) | x \in X, y \in Y\} \\ y &= \{(c_i, l_i) | i = 1, 2, \dots, M, l_i \in C\} \end{aligned} \quad (1)$$

式中, C 表示数据集 D 的全部目标类别; C_b 为源域目标类别; C_n 为目标域待检测类别; x 为输入图像; y 表示每个候选区域 c_i 及其对应的目标类别 l_i 。若目标域训练数据集 D_s 中待检测类别 C_n 包含有 N 个类别, 每个类别有 K 张图像, 则将该任务记为 N -way K -shot任务。小样本目标检测任务中, 通常取 $N = 5$, $K = \{1, 2, 3, 5, 10\}$ 。

为了更加准确地测试算法性能, 减小数据泄露的风险, D_b 、 D_s 和 D_q 应满足:

$$D_b \cap D_n = \emptyset \quad D_s \cap D_q = \emptyset \quad (2)$$

1.2 方法概述

本文提出一种面向小样本目标检测任务的域知识迁移模型, 模型对候选区域特征进行重分布变换以使得源域和目标域有相似的特征分布, 利用源域数据集 D_b 和目标域数据集 D_n 进行多阶段学习, 实现小样本下域知识的高效迁移。本文选取广泛使用的目标检测网络 Faster RCNN^[25] 作为基础网络结构。

本文提出的算法模型 MSFR 结构如图 1 所示。第一阶段利用大量源域数据 D_b 学习域相关的知识, 需对网络的所有模块进行学习。后续阶段训练旨在将学习到的源域知识逐步迁移到目标域中, 因而在知识迁移的过程中需冻结相关模块的参数: 其中, 灰色部分和短线点部分表示在不同训练阶段所需冻结参数的模块; 长短线部分表示该阶段需额外使用的数据, 分类模块使用余弦分类器。具体学习过程将在下面详细介绍。

原 Faster RCNN 网络的候选区域分类模块 cls 使用全连接层分类器来实现, 然而全连接层分类器旨在寻找类别间的最佳分类面, 当训练样本较少时难以实现。为了更好地适用于小样本任务, 本文提出使用余弦相似性度量分类器代替原网络中的全连接层分类器。设目标类别数为 c , 则候选区域分类器需将候选区域分为 $c+1$ 类 (包含背景目标), 分类器 cls 的权重矩阵 W 可以写为 $[w_1, w_2, w_3, \dots, w_{c+1}]$, 其中 w_j 表示类别 j 的权重向量。余弦分类器的计算如下:

$$d_{i,j} = \frac{w_j f_{ROI}(x_i)}{\|w_j\| \times \|f_{ROI}(x_i)\|} \quad (3)$$

式中, x 表示输入图像; x_i 表示输入图像的第 i 个候选区域; j 表示目标类别; $f_{ROI}(x_i)$ 表示第 i 个候选区域的特征; $d_{i,j}$ 表示候选区域 x_i 与目标类别 j 的余弦相似性距离。

为了便于表述, 本文将模型 θ 记为 $\theta = \theta_{bc} + \theta_{RPN} + \theta_{cls} + \theta_{Reg}$, 其中, θ_{bc} 表示特征提取模块, θ_{RPN} 表示候选区域提取模块, θ_{cls} 和 θ_{Reg} 表示分类模块及回归模块。

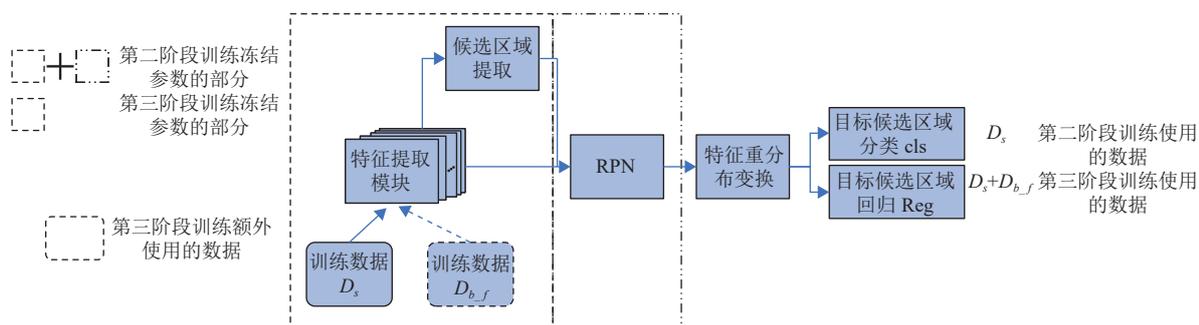


图 1 本文算法模型结构图

2 特征向量重分布

根据中心极限定理, 在许多实际应用场景中, 当统计数据足够多的情况下, 许多统计变量都可近似用高斯分布来表示。同时, 目前的深度学习算法都遵循数据独立同分布这一基本假设, 然而, 在面对小样本任务时, 难以估计出样本真正的分布。同时, 在知识迁移过程中, 若源域数据与目标域数据的分布相同或相似, 迁移效率会得到有效提升; 若二者的分布差异较大, 则迁移效果较差。现有小样本目标检测算法均是大量源域数据上学习到的知识迁移到小样本目标域任务中, 但无法保证源域数据与目标域数据有相同的数据分布。因此, 本文提出一种特征向量重分布算法, 将特征向量分布变换为高斯分布或类高斯分布, 然后进行后续的处理。

特征重分布变换函数需满足以下条件: 1) 该变换函数为一个严格递增的函数, 即在变换后仍保持原数据的大小关系, 只是数据的相对大小发生了改变。2) 该变换函数是一个连续函数, 使得原始数据中比较集中的数据在变换后仍比较集中。3) 该变换函数是一个可导函数。

因此本文提出的重分布变换方式如式 (4) 所示, 设原特征向量为 x , 重分布后的特征向量

为 y :

$$y = BC(x) = \begin{cases} \frac{(x+c)^\lambda - 1}{\lambda g} & \lambda \neq 0 \\ \frac{\ln(x+c)}{g} & \lambda = 0 \end{cases} \quad (4)$$

式中, \bar{x} 为向量 x 的平均值; $g = \bar{x}^{\lambda+1}$; c 为一个常数, 用以保证输入为一个恒大于零的数; $\lambda \in [-2, 2]$ 为一个超参, 用于调整变换后的分布。

该变换的基本原理是: 若向量 x 的各元素与向量均值 \bar{x} 间的差异较小, 则减小元素间的相对距离; 若向量 x 的各元素与向量均值 \bar{x} 间的差异较大, 则增大元素间的相对距离; 通过改变向量间元素的相对距离来改变其分布。参数 λ 直接作用于变换后元素间的相对距离, 因此 λ 对变换后的分布有着直接影响。

为了选取合适的 λ 以得到最好的检测精度, 本文对 $\lambda \in [-2, 2]$ 进行遍历, 步长为 0.1。

3 多阶段微调训练

Faster RCNN 网络主要包括 θ_{bc} 、 θ_{RPN} 、 θ_{cls} 及 θ_{Reg} 。其中, θ_{bc} 用于提取图像特征; θ_{RPN} 用于提取包含目标的候选区域, 该模块仅对候选框中是否包含目标进行判断, 并不关心候选框包含何种目标, 因而这两部分与具体检测类别相关性较小; 候选框

分类模块 θ_{cls} 、候选框回归模块 θ_{Reg} 则需识别候选框包含何种目标并对目标进行精确定位, 与待检测类别密切相关^[21]。因此, 本文提出了一种多阶段微调训练策略, 以使得将学习到的域知识更好地迁移到目标任务中。

3.1 第一阶段基类模型训练

第一阶段基类网络学习过程中, 利用源域数据

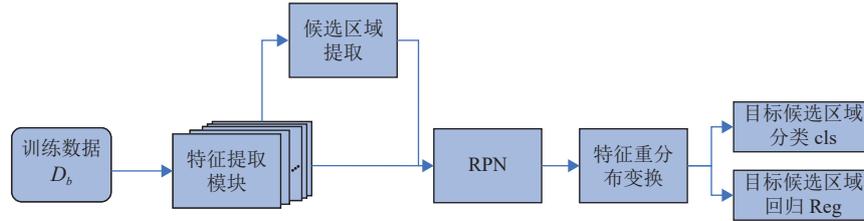


图 2 第一阶段基类模型训练

该阶段网络的损失函数 L 与文献 [25] 保持一致, 为:

$$L = L_{RPN} + L_{cls} + L_{Reg} \quad (5)$$

式中, L_{RPN} 为候选区域提取模块的损失函数; L_{cls} 为候选区域分类模块的交叉熵损失函数; L_{Reg} 为 Smooth L1 损失函数。

3.2 第二阶段微调训练

在基类模型训练完成后, 特征提取模块 θ_{bc} 、候选区域提取模块 θ_{RPN} 具有非常好的性能, 且具有较好的泛化能力, 因此在第二阶段微调训练时将这几个模块的参数冻结(即参数在训练过程中不进行更新), 只对候选区域分类 θ_{cls} 及回归模块 θ_{Reg} 进行

训练, 其流程如图 3 所示。图中, 灰色部分表示在训练中需冻结的模块, 即冻结 $\theta_{bc} + \theta_{RPN}$, 只对 θ_{cls} 和 θ_{Reg} 进行微调训练。

第二阶段微调过程中使用目标域数据 D_s 训练, 对于 N -way K -shot 任务, D_s 数据集包含 N 个类别, 每个类别均只有 K 张图像。

该阶段微调训练的损失函数与基类训练的损失函数保持一致。但由于微调训练数据集 D_s 数量较少、微调过程只对模型的少数参数进行更新等原因, 若采用与第一阶段训练过程中相同的学习率, 微调模型难以跳出鞍点, 使得模型难以收敛。因此, 在该阶段训练中, 学习率设置为第一阶段基类训练学习率的 0.1 倍。

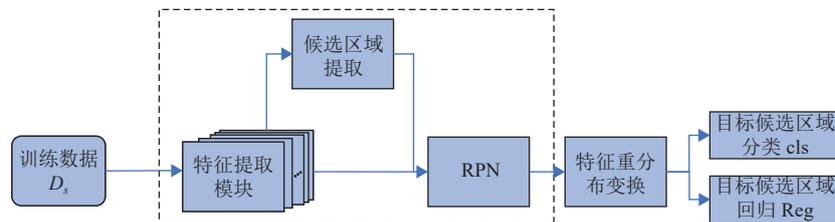


图 3 第二阶段微调训练

3.3 第三阶段微调训练

在第二阶段训练结束后, 网络对目标类别 C_n 有了一定的检测能力, 但却舍弃了对源域类别 C_b 的检测性能; 此外, 虽然 Faster RCNN 网络的最后一层(分类、回归模块)对小样本任务有非常重要的作用, 但是对于新类别, 候选区域提取模块也起着至关重要的作用。因此本文提出在第二阶段微调模

型的基础上进行进一步地学习, 将候选区域提取模块学习到的知识更好地迁移到小样本任务中。该阶段的网络流程如图 4 所示。

第三阶段微调使用的数据 D_f 由两部分构成, D_s 和 D_{b_f} , 其中, D_s 为第二阶段训练数据, D_{b_f} 为对源域数据集 D_b 进行重采样后组成的基类小样本数据集。 D_{b_f} 的采样过程如下: 针对 N -way K -shot

任务, 对源域数据集 D_b 中的每个类别 $l_i \in C_b$ 随机选取 K 个样本, 若源域类别 C_b 有 N_b 个类别, 则 D_{b_f} 共有 $N_b \times K$ 张图像。

同样地, 图 4 中灰色部分表示在训练中冻结 θ_{bc} 的参数。该阶段微调训练的学习率设置为第一阶段基类训练学习率的 0.05 倍。

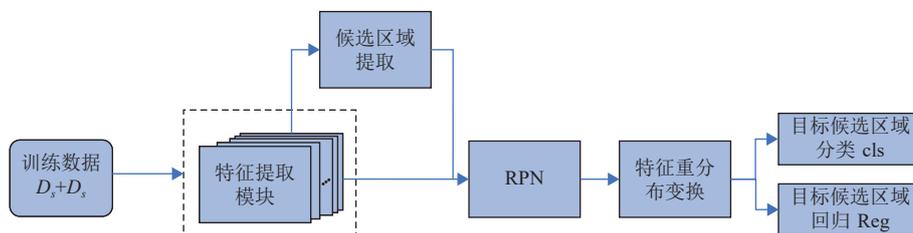


图 4 第三阶段微调训练

本文的算法流程实现如下:

1) 数据集准备。源域训练集 D_b , 对 D_b 进行重采样的少量数据集 D_{b_f} , 目标域数据集 D_s 。

2) 第一阶段训练。训练过程如 3.1 节所述, 所得模型记为 $\theta_1 = \theta_{bc_1} + \theta_{RPN_1} + \theta_{cls_1} + \theta_{Reg_1}$ 。

3) 第二阶段训练。将 θ_1 的最后一层删除并根据 D_s 的类别数对最后一层进行随机赋值, 记为 θ_1' ; 冻结 θ_1' 中 θ_{bc} 和 θ_{RPN} 的参数, 使用数据 D_s 对 θ_1' 进行训练。训练过程如 3.2 节所述, 所得模型记为 $\theta_2 = \theta_{bc_1} + \theta_{RPN_1} + \theta_{cls_2} + \theta_{Reg_2}$ 。

4) 第三阶段训练。将 θ_1 和 θ_2 进行融合做为第三阶段训练的初始模型 θ_2' , 即 $\theta_2' = \theta_{bc_1} + \theta_{RPN_1} + (\theta_{cls_1} + \theta_{cls_2}) + (\theta_{Reg_1} + \theta_{Reg_2})$, 可以看出, θ_2' 对源域类别和目标域类别已具有一定的检测能力; 为了进一步提高模型的检测性能, 进行进一步的微调训练: 冻结 θ_2' 中的 θ_{bc_1} , 对其余模块参数进行训练。训练过程如 3.3 节所示。所得模型记为 $\theta_3 = \theta_{bc_1} + \theta_{RPN_3} + \theta_{cls_3} + \theta_{Reg_3}$ 。

4 实验结果及分析

为了评估本文算法的有效性, 在 VOC 数据集上与目前多个小样本目标检测方法进行对比。

4.1 数据集及实验设置

本节使用的 VOC 数据集包括 2007 和 2012 两个版本。VOC2007 包含 20 类图像共 9 963 张, 其中训练集 5 011 张, 测试集 4 952 张。VOC2012 是 VOC2007 数据集的升级版, 共 20 类 11 530 张。本文主要研究 N -way K -shot 任务下小样本目标检测的精度, 其中, $N = 5$, $K = \{1, 2, 3, 5, 10\}$ 。为了验证算法的鲁棒性, 本文对 VOC 数据集进行划分, 以构造小样本任务, 划分方式与文献 [14-20] 保持一致。数据集的划分共有 3 种分组方式, 每次划分

从 VOC 数据集的 20 个类别中选取 15 类作为源域数据集 D_b 进行第一阶段训练, 其余 5 类作为目标域数据 D_n , 小样本待检测目标类别记为 C_n 。每个类从 2007 和 2012 版本的集合中随机采样 K 张图像作为目标域训练数据集 D_s , 该类别的其余图像作为目标域测试集 D_q 。精度衡量指标使用 AP50, 置信度为 0.5。本文实验环境为 Ubuntu18.04, 编程语言为 Python1.6、Pytorch1.6 深度学习框架, 显卡为 Nvidia v100s, 显卡支持为 CUDA10.0。

特征提取网络模块使用在 ImageNet 数据集上预训练的 ResNet101 模型。多阶段训练时 batchsize 设置为 16, 优化算法使用 SGD。第一阶段基类模型训练时, 初始学习率设置为 0.02, 训练次数为 20 000 次; 第二阶段微调训练时, 初始学习率设置为 0.002; 第三阶段微调训练时, 初始学习率设置为 0.000 1。

文献 [26] 表明, 当 $\lambda = 0.7$ 时, 该变换在小样本图像分类任务中取得了最好的效果。因此, 本文在后续实验中, 均选取上述超参。

4.2 精度分析

本节实验将对 VOC 数据集 3 种不同划分方式分别记为 Novel1、Novel2、Novel3。为了更好地衡量算法的整体检测性能, 本文采用文献 [18] 中的评价指标: 其中, 对源域类别的平均检测精度记为 bAp, 目标类别的平均检测精度记为 nAp, 对 VOC 数据集全部 20 个类别的检测精度记为 Ap。

在不同划分下, 首先利用源域数据集 D_b 进行第一阶段训练, 算法对源域类别 C_b 的检测精度如表 1 所示。由表 1 可知, 特征重分布算法在大量训练数据下对目标检测精度略有影响, 较之原 Faster RCNN 网络对源域类别的检测精度提升约 0.15%。

在完成第一阶段训练后, 分别对不同划分进行第二、三阶段的微调训练, 对比结果如表 2 所示, 其中加黑部分为各任务下的最优精度。

表 1 第一阶段源域类别检测精度 %

划分方式	本文算法精度	原Faster RCNN
Novel1	81.07	80.79
Novel2	81.22	81.17
Novel3	81.67	81.54

从表 2 可以看出, 本文算法在 3 个划分方式上对多个任务都达到了最好精度, 较其余算法均有明显提升, 尤其是样本少的时候提升更为明显 (如对 1-3 shot 任务)。在多个任务上, 较其余最好精度提升了 0.42%~9.06%, 验证了本文算法的有效性; 此外, 在不同划分下, 随着样本数的增加, 目标检测

精度也随之不断提升。表 3 为不同算法在 VOC 数据集不同划分不同任务下对新类别的检测精度, 其中加黑部分为各任务下对新类别的最优精度。由表 3 可以看出, 各任务下本文算法对多数类别的检测精度都有明显提升。Meta-RCNN 对某些类别的检测精度可以达到最优, 然而会损耗对其余类别的检测精度。如对于 Novel2 划分下 5-shot 任务, Meta-RCNN 对 cow、horse 两个类别达到了最优精度, 然而对 bottle 的检测精度最低, 只有 0.3%。此外, 对于 Novel2 划分, 本文算法和对比算法的检测精度较之 Novel1 和 Novel3 均有较大差距。这是由于 Novel2 中, 新类别的形状、颜色、尺寸等变化较大, 目标尺寸较小, 仅使用 5 个样本难以学习到表征能力足够的类别相关特征。而对于类间变化较小的类别, 仅使用少量样本如 1-shot、2-shot, 即可达到 50% 左右的精度。

表 2 本文算法与现有算法在 VOC 数据集不同任务上的精度对比 %

方法/数据划分方式	FSRW ^[22]	Meta R-CNN ^[18]	LSTD ^[24]	Meta-Det ^[20]	RepMet ^[19]	TFA ^[21]	MS	MSFR	
Novel1	1	14.7	17.9	8.22	26.1	26.1	39.8	33.72	41.97
	2	15.5	25.5	11.02	32.9	32.9	36.1	38.93	45.16
	3	26.7	35.0	12.40	34.4	34.4	44.7	35.36	46.03
	5	33.9	45.7	28.78	38.6	38.6	55.7	49.18	56.66
	10	47.4	51.5	38.50	41.3	41.3	56.0	52.63	57.32
Novel2	1	15.7	10.4	11.38	17.2	17.2	23.5	20.89	27.25
	2	15.3	21.4	3.80	22.1	22.1	26.9	23.81	30.48
	3	22.7	29.6	5.00	23.4	23.4	34.1	33.58	34.52
	5	30.1	34.8	15.76	28.3	28.3	35.1	35.16	36.11
Novel3	10	40.5	45.4	31.00	35.8	35.8	39.1	40.13	40.59
	1	21.3	14.3	12.54	27.5	27.5	30.8	30.51	32.04
	2	25.6	18.2	8.56	31.1	31.1	34.8	33.71	40.71
	3	28.4	27.5	14.98	31.5	31.5	42.8	36.90	42.28
	5	42.8	41.2	27.30	34.4	34.4	49.5	48.78	50.56
	10	45.9	48.1	36.28	37.2	37.2	49.8	50.02	51.12

表 3 本文算法与现有算法在 VOC 数据集不同划分不同任务下对新类别的检测精度 %

样本数	方法	Novel1					Novel2					Novel3				
		bird	bus	cow	motorbike	sofa	aeroplane	bottle	cow	horse	sofa	boat	cat	motorbike	sheep	sofa
1	Meta-RCNN	6.1	32.8	15.0	35.4	0.2	23.9	0.8	23.6	3.1	0.7	0.6	31.1	28.9	11.0	0.1
	FSRW	13.5	10.6	31.5	13.8	4.3	11.8	9.1	15.6	23.7	18.2	10.3	41.4	29.1	16.2	9.4
	LSTD	13.4	21.4	6.3	0	0	0.1	1.5	10.4	44.9	0	0	27.8	25.0	9.7	0.2
	MSFR	20.63	59.85	56.64	66.65	6.08	28.58	9.81	38.46	33.85	25.80	12.20	49.96	57.86	36.30	3.88
2	Meta-RCNN	17.2	34.4	43.8	31.8	0.4	12.4	0.1	44.4	50.1	0.1	10.6	24.0	36.2	19.2	0.8
	FSRW	21.2	12.0	16.8	17.9	9.6	28.6	0.9	27.6	0	19.5	6.3	47.1	28.4	28.1	18.2
	LSTD	17.3	12.5	8.6	0.2	16.5	3.0	1.5	13.9	0.6	0	0.2	27.3	0.1	15.0	0.2
	MSFR	28.50	59.25	54.82	64.83	18.42	48.84	4.93	31.32	27.18	40.13	18.7	54.25	47.50	52.80	30.31

% 续表

样本数	方法	Novel1					Novel2					Novel3				
		bird	bus	cow	motorbike	sofa	aeroplane	bottle	cow	horse	sofa	boat	cat	motorbike	sheep	sofa
3	Meta-RCNN	30.1	44.6	50.8	38.8	10.7	25.2	0.1	50.7	53.2	18.8	16.3	39.7	32.6	38.8	10.3
	FSRW	26.1	19.1	40.7	20.4	27.1	29.4	4.6	34.9	6.8	37.9	11.7	48.2	17.4	34.7	30.1
	LSTD	23.1	22.6	15.9	0.4	0	12.6	0.7	11.3	0.4	0	0	36.6	21.4	16.9	0
	MSFR	28.11	51.16	54.55	57.36	38.98	60.10	4.70	31.60	42.57	33.61	18.96	57.78	47.50	53.17	34.01
5	Meta-RCNN	35.8	47.9	54.9	55.8	34.0	28.5	0.3	50.4	56.7	38.0	16.6	45.8	53.9	41.5	48.1
	FSRW	31.5	21.1	39.8	40.0	37.0	33.1	9.4	38.4	25.4	44.0	14.8	59.1	49.6	45.0	45.6
	LSTD	24.1	30.2	24.0	40.0	25.6	11.6	9.1	15.2	42.9	0	0.2	51.5	37.2	26.9	20.7
	MSFR	34.24	70.05	63.37	70.89	44.77	56.12	9.40	42.54	28.03	44.47	20.04	61.35	65.97	55.58	49.86
10	Meta-RCNN	52.5	55.9	52.7	54.6	41.6	52.8	3.0	52.1	70.0	49.2	13.9	72.6	58.3	47.8	47.6
	FSRW	30.0	62.7	43.2	60.6	40.6	41.8	14.0	42.7	63.4	40.7	22.4	59.9	59.8	46.1	41.4
	LSTD	22.8	52.5	31.3	45.6	40.3	41.5	9.3	29.2	38.9	36.1	10.5	53.3	41.9	36.2	39.5
	MSFR	27.17	75.54	70.14	69.81	43.95	51.58	11.23	43.91	46.57	49.65	19.13	69.36	65.60	59.74	41.77

由于本文算法在微调训练的数据集 D_f 不仅包含目标域数据集 D_s , 也包含源域数据 $D_{b,f}$, 因此与基于元学习的算法在训练完成后只保留对新类别数据的检测能力相比, 本文算法在大幅提升对新类别检测精度的同时, 仍保留了对源域数据类别的检测能力。3种不同划分5-shot任务下对源域类别、目标类别、全部类别的检测精度如表4所示。

表4 同划分方式下在5-shot任务下的检测性能

数据集	Ap	bAp	nAp
Novel1	73.33	78.89	56.66
Novel2	67.50	77.97	36.11
Novel3	72.12	79.30	50.56
Mean	70.98	78.72	47.78

通过表3和表4可知, 在不同划分下本文算法对源域类别的检测精度基本一致, 平均可达到78.72%, 较第一阶段精度仅下降约3%左右, 有效证明了本文算法在提升小样本目标检测精度的同时, 仍保持了对源域类别的检测性能。

4.3 计算复杂度分析

与Faster RCNN相比, 本文提出的MSFR网络结构有两点改进: 特征重分布模块和余弦相似度分类模块。因此, 本文针对这两个模块分别进行复杂度分析。

设候选区域特征维度为 N , 本实验中 $N = 512$ 。

特征重分布模块: 由式(4)可知, 特征重分布模块中包含加法、幂运算、减法、除法4种运算,

因此该模块的计算量为 $FLOPs = 4N = 2048$

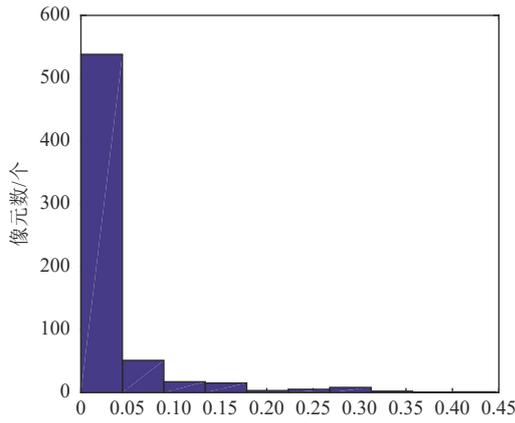
余弦相似度分类模块: 为了简化余弦相似度计算过程, 本实验在进行余弦相似度计算过程时, 首先对 w_j 和 $f_{ROI}(x_i)$ 进行标准化处理, 则式(3)简化为 $d_{i,j} = w_j f_{ROI}(x_i)$ 。可以看出, 简化后的余弦相似度计算与全连接层的计算一致。因此, 该模块中增加的计算量在于特征向量标准化计算过程。标准化计算包含减法、除法两种运算, 则该模块较之全连接分类器增加的计算量为 $FLOPs = 2 \times 2N = 2048$ 。

MSFR增加的计算量为 $FLOPs = 4096$ 。与Faster RCNN相比, MSFR额外增加的计算量可以忽略不计。

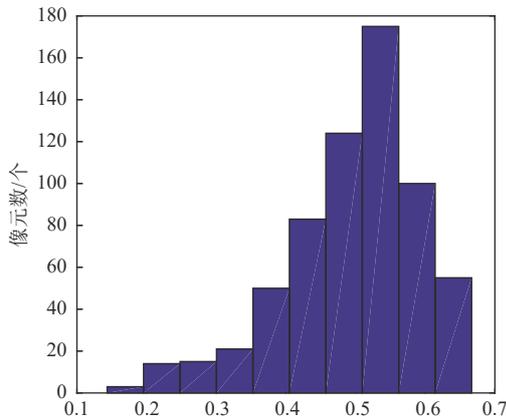
4.4 消融实验

为了验证特征向量重分布算法的有效性, 本文对该模块进行了消融实验, 实验结果如表2所示。其中, MS表示无特征重分布模块的多阶段学习算法, 可以看出, 特征向量重分布算法对小样本目标类别的检测精度有着直接影响。较之无特征向量重分布模块, 本文算法在不同任务上对目标类别的平均检测精度提高了4.62%, 样本数量越少, 提升越显著, 如对于 $K \leq 3$ 的任务, 精度提升最高达10.67%, 平均提升5.89%。

为了验证特征重分布变换对特征分布的影响, 本文在测试过程中提取重分布变换层输入和输出的特征向量, 并对其特征分布进行可视化分析, 如图5所示。其中, 图5a为特征变换前分布, 图5b为变换后的特征分布。可以看出, 特征向量在变换前是一种长尾分布, 而经重分布变换后的特征分布为一个类高斯分布。



a. 变换灰度级前分布



b. 变换灰度级后分布

图 5 VOC 数据集上 5-shot 检测结果示意图

图 6 展示了本文算法在 VOC 数据集上 5-shot 任务下对不同划分中新类别的检测结果示意图。可以看出, 本文算法在不同场景下对目标类别都具有较好的检测能力。

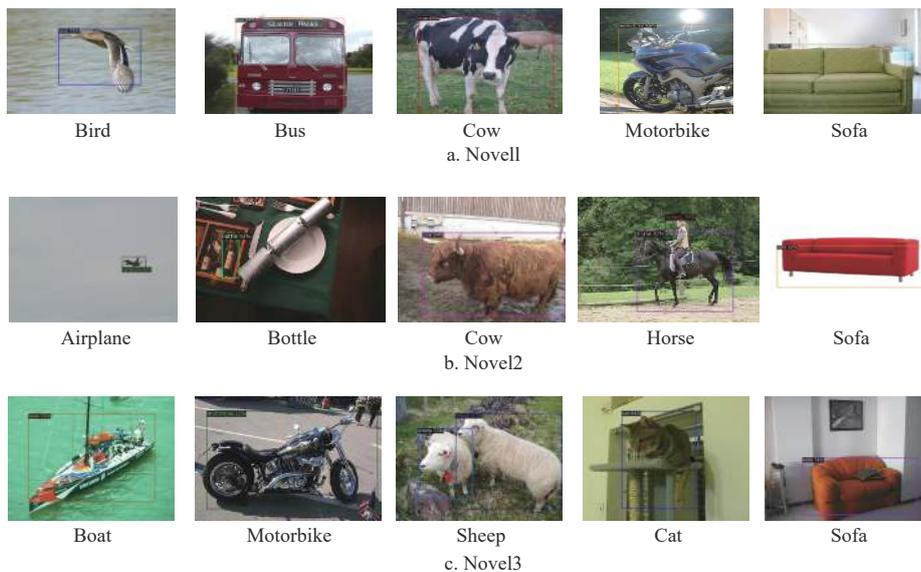


图 6 特征向量重分布变换

5 结束语

针对现有小样本目标检测的研究未考虑数据分布知识迁移效率的影响及检测精度较低的问题, 本文提出了一种多阶段特征重分布的小样本目标检测算法, 在 VOC 数据集的大量实验都取得了较高的检测精度。本文结论如下: 1) 提出了基于特征重分布算法的小样本目标检测算法。该方法将特征向量分布变换为高斯分布或类高斯分布, 以保证小样本目标域数据和源域数据有相同或相似的数据分布。2) 提出了一种多阶段微调训练策略, 提高了知识迁移效率。该方法通过分阶段冻结不同模块参数, 将不同模块知识逐步迁移到小样本任务中。本文算法在不同任务下对小样本目标域类别的检测精度提高了 0.42%~9.06%。尤其对典型的 5-shot 任务, 本文算法对源域类别的平均检测精度达 78.72%, 较第一阶段检测精度仅下降 2.60%; 对小样本目标域类别的最高检测精度可达 56.66%。大量实验表明, 本文算法在提升小样本目标检测精度的同时, 对源域类别的检测性能未有显著降低。

实验结果表明, 本文算法对于尺寸较小的目标检测精度较差。因此在将来的研究工作中, 考虑使用注意力机制来抑制背景区域特征对目标的干扰, 以提高模型对小目标的检测性能, 同时将本文算法推广到红外、遥感等图像的小样本目标检测任务中。

参 考 文 献

- [1] 宋闯, 赵佳佳, 王康, 等. 面向智能感知的小样本学习研究综述[J]. 航空学报, 2020, 41(S1): 15-28.
SONG C, ZHAO J J, WANG K, et al. A survey of few shot learning based on intelligent perception[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(S1): 15-28.
- [2] CHEN Z, FU Y, CHEN K, et al. Image block augmentation for one-shot learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2019: 3379-3386.
- [3] ANTONIOU A, STORKEY A, EDWARDS H. Data augmentation generative adversarial networks[EB/OL]. [2021-11-12]. <https://arxiv.org/pdf/1711.04340.pdf>.
- [4] XIAN Y, SHARMA S, SCHIELE B, et al. f-VAEGAN-D2: A feature generating framework for any-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2019: 10275-10284.
- [5] QI H, BROWN M, LOWE D G. Low-Shot learning with imprinted weights[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2018: 5822-5830.
- [6] YU Z, CHEN L, CHENG Z, et al. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2020: 12856-12864.
- [7] CHOBOLA T, VAŠATA D, KORDÍK P. Transfer learning based few-shot classification using optimal transport mapping from preprocessed latent space of backbone neural network[C]//AAAI Workshop on Meta-Learning and MetaDL Challenge. Menlo Park: PMLR, 2021: 29-37.
- [8] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[EB/OL]. [2021-11-18]. <https://arxiv.org/pdf/1703.03400.pdf>.
- [9] 刘颖, 雷研博, 范九伦, 等. 基于小样本学习的图像分类技术综述[J]. 自动化学报, 2021, 47(2): 297-315.
LIU Y, LEI Y B, FAN J L, et al. Survey on image classification technology based on small sample learning[J]. Acta Automatica Sinica, 2021, 47(2): 297-315.
- [10] GRAVES A, WAYNE G, DANIHELKA I. Neural Turing machines[EB/OL]. [2021-11-25]. <http://de.arxiv.org/pdf/1410.5401>.
- [11] SANTORO A, BARTUNOV S, BOTVINICK M, et al. Meta-learning with memory-augmented neural networks[C]//International Conference on Machine Learning. New York: [s.n.], 2016, 4: 2740-2751.
- [12] RIEMER M, CASES I, AJEMIAN R, et al. Learning to learn without forgetting by maximizing transfer and minimizing interference[C]//International Conference on Learning Representations. New Orleans: Elsevier, 2018: 1-31.
- [13] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition[C]//ICML Deep Learning Workshop. New York: PMLR, 2015: 1325-1355.
- [14] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems. Barcelona: [s.n.], 2016: 3637-3645.
- [15] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//Advances in Neural Information Processing Systems. Long Beach: [s.n.], 2017: 4077-4087.
- [16] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2018: 1199-1208.
- [17] FU K, ZHANG T, ZHANG Y, et al. Meta-SSD: Towards fast adaptation for few-shot object detection with meta-learning[J]. IEEE Access, 2019, 7: 77597-77606.
- [18] RAJESWARAN A, FINN C, KAKADE S M, et al. Meta-learning with implicit gradients[C]//Advances in Neural Information Processing Systems. Vancouver: [s.n.], 2019: 113-124.
- [19] KARLINSKY L, SHTOK J, HARARY S, et al. Repmet: Representative-based metric learning for classification and few-shot object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2019: 5197-5206.
- [20] WANG Y X, RAMANAN D, HEBERT M. Meta-learning to detect rare objects[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Computer Society, 2019: 9925-9934.
- [21] WANG X, HUANG T, GONZALEZ J, et al. Frustratingly simple few-shot object detection[C]//International Conference on Machine Learning. New York: PMLR, 2020: 9919-9928.
- [22] KANG B, LIU Z, WANG X, et al. Few-Shot object detection via feature reweighting[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Computer Society, 2019: 8420-8429.
- [23] WU J, LIU S, HUANG D, et al. Multi-Scale positive sample refinement for few-shot object detection[C]//European Conference on Computer Vision. Glasgow: Springer, 2020: 456-472.
- [24] CHEN H, WANG Y, WANG G, et al. LSTD: A low-shot transfer detector for object detection[EB/OL]. [2021-11-10]. <https://arxiv.org/pdf/1803.01529.pdf>.
- [25] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015: 91-99.
- [26] LIU L, WANG L, MA Z, et al. Few-Shot classification via feature redistribution[C]//2020 IEEE 6th International Conference on Computer and Communications (ICCC). Chengdu: IEEE, 2020: 1822-1826.