

基于序贯三支决策的半监督目标检测算法

宋法兴 苗夺谦 张红云

同济大学电子与信息工程学院 上海 201804

嵌入式系统与服务计算教育部重点实验室 上海 201804

(2130767@tongji.edu.cn)

摘要 深度学习对大规模数据的需求以及目标检测标注任务的复杂性促进了半监督目标检测任务的发展。近年来,半监督目标检测已经取得了很多优秀的成果。然而,伪标签中的不确定性依然是半监督目标检测研究中难以避免的问题,优越的半监督方法要求选取合适的过滤阈值来权衡伪标签的噪声信息比例和召回率,以最大程度保留准确有效的伪标签。为了解决此问题,在半监督检测的框架中引入了序贯三支决策算法,将模型输出的伪标签根据不同的筛选阈值划分为干净的前景标签、有噪声的前景标签,以及干净的背景标签,并对其采取不同的处理策略。对有噪声的前景标签采用负类学习损失来学习这些存在噪声的标签,避免学习到其中的噪声信息。实验结果表明了所提算法的性能优势,针对 COCO 数据集,在有监督数据占比只有 10% 的情况下,该方法实现了 35.2% 的检测精度,相比仅依靠有监督训练性能提升了 11.34%。

关键词: 序贯三支决策; 不确定性; 负类学习; 半监督学习; 半监督目标检测

中图法分类号 TP389.1

Semi-supervised Object Detection with Sequential Three-way Decision

SONG Faxing, MIAO Duoqian and ZHANG Hongyun

College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, China

Abstract The need for large scale data in deep learning and the complexity of object detection annotation task promote the development of semi-supervised object detection. In recent years, semi-supervised object detection has achieved many excellent results. However, the uncertainty in pseudo labels is still an unavoidable problem in semi-supervised object detection. The superior semi-supervised method requires an appropriate filtering threshold to balance the proportion of pseudo labels' noise and the recall rate, so as to retain accurate and effective labels as much as possible. To solve this problem, this paper introduces a sequential three-way decision algorithm into semi-supervised object detection, which divides the model output pseudo-labels into clean foreground labels, noisy foreground labels, and clean background labels according to different filtering thresholds, and adopts different processing strategies for them. For noisy foreground labels, we use negative class learning loss to learn these noisy labels, thereby avoiding learning noise information from them. Experimental results show the performance advantage of this algorithm. For COCO dataset, this method achieves performance of 35.2% when supervised data only accounts for 10%, which outperforms the supervised results by 11.34%.

Keywords Sequential three-way decisions, Uncertain, Negative class learning, Semi-supervised learning, Semi-supervised object detection

1 引言

在图像分类领域,半监督学习(Semi-Supervised Learning, SSL)已经取得了很多成果。然而,在目标检测领域,由于模型输出具有复杂性,直接迁移半监督分类中的方法所取得的改进效果往往有限。伪标签方法是半监督目标检测算法

已经取得了很多成果。然而,在目标检测领域,由于模型输出具有复杂性,直接迁移半监督分类中的方法所取得的改进效果往往有限。伪标签方法是半监督目标检测算法

到稿日期:2023-06-05 返修日期:2023-08-08

基金项目:国家重点研发计划(2022YFB3104700);国家自然科学基金(61976158,61976160,62076182,62163016,62006172);江西省自然科学基金重点项目(20212ACB202001);江西省“双千计划”

This work was supported by the National Key Research and Development Program of China(2022YFB3104700), National Natural Science Foundation of China (61976158, 61976160, 62076182, 62163016, 62006172), Key Project of Natural Science Foundation of Jiangxi Province, China (20212ACB202001) and Jiangxi “Double Thousand Plan”.

通信作者:苗夺谦(dqmiao@tongji.edu.cn)

中的常用方法。然而,由于模型自身准确性的限制以及有标注数据的有限性,检测模型输出的伪标签中存在很多噪声标签。常规做法是利用一个较高的阈值对模型的输出结果进行过滤,将分类置信度低于阈值的伪标签舍弃。然而,过滤阈值的选取是一个难题,需要在伪标签准确率与召回率之间进行权衡。

为了在保证伪标签准确率的情况下,尽可能多地从教师模型中进行学习,之前的工作大多利用动态阈值^[1]方法来解决上述问题。此方法根据训练过程中的一些指标(如置信度的分布等)动态地调整阈值,以达到平衡伪标签准确率与召回率的目的。

本文在前人的研究基础上,引入了序贯三支决策^[2]的思想,并将其应用到半监督目标检测中。具体来说,本文将模型输出的伪标签按照上下阈值的不同分为3类:干净的前景标签、带有噪声的前景标签以及干净的背景标签。对于带有噪声的前景标签,利用类不确定性度量和分类概率分布,选择其中无噪声的信息进行学习,即对通过第二阶段筛选的伪标签采用负类学习^[3]方法进行处理。具体来说,保留其中准确的负类标签,并舍弃不确定的前景类别标签,即不计算损失值。这种方法避免了噪声对模型性能的影响,同时最大程度地利用了伪标签所提供的信息。本文的主要贡献如下:

1) 针对半监督目标检测中的筛选阈值问题,本文提出了即插即用的序贯三支决策方法,在尽可能学习更多信息的前提下,有效地减少了噪声信息对模型性能的影响。

2) 针对噪声标签中的噪声信息,采用负类学习的方法减少对噪声信息的学习,并通过实验证明了其在半监督目标检测中的有效性。

3) 实验证明本文提出的方法有效地提高了半监督目标检测算法在 MS-COCO^[4]数据集上的半监督检测性能。

2 相关工作

2.1 目标检测

目标检测是计算机视觉领域的基础任务之一。根据对锚框的需求来分类,现有的目标检测算法可以分为 Anchor-based 和 Anchor-free 两大类。其中,Anchor-based 方法又可进一步划分为单阶段和双阶段算法。

在单阶段目标检测算法中,早期的 SSD^[5]引入了多尺寸和多长宽比的先验框以及不同尺度特征图,实现了对大小物体的检测。RetinaNet^[6]提出了 Focal Loss 损失函数,使得单阶段算法首次实现了超越双阶段检测算法的性能。

而在双阶段目标检测算法中,Girshick 等提出的 RCNN^[7]算法首次将 CNN 算法应用到了目标检测领域。随后,Fast RCNN^[8]和 Faster RCNN^[9]算法相继出现,目标检测算法的准确度得到进一步提升。

Anchor-free 检测算法^[10-11]放弃了之前方法的预设先验框的概念,改为以点为基准,直接预测长宽的方法,避免了在预设先验框中长宽比和尺寸等超参数对结果的影响。

2.2 半监督图像分类

在少量有标注数据的基础上,半监督学习方法通过对大量的无标注数据的使用来提升模型的性能。伪标签^[12]方法

利用现有模型对无标注图片进行预测,根据预测的分类概率来选择高置信度的伪标签,最终将留下的无标注数据和有标注数据联合训练,不断进行迭代优化。Noisy Student^[13]通过将上一个 Self-Training^[14]周期生成的噪声插入到数据中来改善最终的模型性能。此外,基于一致性约束的方法中,模型对同一张图片经过不同数据增强后得到的扰动图片的预测结果应该是一致的。常用的数据增强方法包括翻转、Cutout^[15]和高斯模糊。此外,Mixup^[16]和 CutMix^[17]等方法还可用于约束模型在两张图片插值上的一致性。

2.3 半监督目标检测

半监督学习方法在目标检测领域已经取得了一些显著的成果。STAC^[18]将 Faster-RCNN 中经过 NMS 筛选之后的结果使用阈值 τ 进行过滤,将最终留下的标签作为伪标签输入到模型中进行学习。RPL^[19]在 STAC 的基础上,提出了确定感知的伪标签,通过分类方法解决检测的边界框回归问题。MUM^[20]通过在不同图片之间进行块融合并在特征层面进行特征恢复,获得了更丰富的图像特征。

最近,DSL^[21]对模型输出的伪标签也根据不同的阈值划分为3个部分,但是对于中间的带噪声的伪标签,DSL 直接采取忽略操作,不计算相关的损失,这损失了噪声标签中的有价值的信息。VC^[22]对噪声标签问题采用了虚拟类别的方法,并通过收集不同训练阶段或者不同模型的伪标签来发现潜在的噪声标签集合。本文采用负类学习的方法,在一个训练阶段即可得到所有潜在类别集合。与 VC 虚拟类别的潜在标签集相比,负类学习对潜在类别的数量是自适应的,而 VC 中这个数量通常被限制为 2。

3 基于序贯三支决策的半监督目标检测算法

在半监督目标检测领域,噪声伪标签是一个难以避免的问题;当选取的阈值过高时会出现严重的漏检问题,这个噪声信息隐式地存在于图片的背景区域之中;而当阈值选取过低时,分类置信度不高的样本可能出现分类错误、前背景混淆的问题。

本章首先对半监督目标检测的结构和基本问题进行定义,然后再介绍基于序贯三支决策的半监督目标检测算法以及其中的负类损失函数的设计过程。

3.1 基本问题定义

在半监督目标检测问题中,训练数据集中包含两个子数据集:有标注数据集 D_L 和无标注数据集 D_U 。其中 $D_L = \{(x_n^l, b_n^l) \mid n=0\}^N$, $D_U = \{(x_n^u) \mid n=0\}^N$, x 和 b 分别代表数据集的图片以及图片对应的标签信息, N^l 和 N^u 分别代表有标注和无标注图片的数量。

本文参考 SoftTeacher 的基础训练范式,模型整体包含教师和学生模型两个分支,并且同时对有标注和无标注图片进行学习。如图 1 所示,对于有标注图片,将其直接输入到学生网络 S 中进行正常的前向处理,得到有监督损失 L_l 。对于无标注图片,使用强弱数据增强进行扰动;将弱增强的图片输入到教师模型 T 中,预测结果并依据阈值 τ 筛选合适的伪标签;将强增强图片输入到学生模型 S 中,使用前面得到的伪标签进行训练,基于一致性约束计算无监督损失 L_u 。学生模型的

参数同时使用有监督损失和无监督损失来更新,而教师模型的参数则由学生模型经 EMA 进行更新。

$$w^{t+1} = m \cdot w^t + (1-m) \cdot w \quad (1)$$

其中, w^{t+1} 和 w^t 为教师模型的参数, w 为学生模型的参数, m 是 EMA 模块的更新动量。

本文的模型总体结构如图 1 所示。

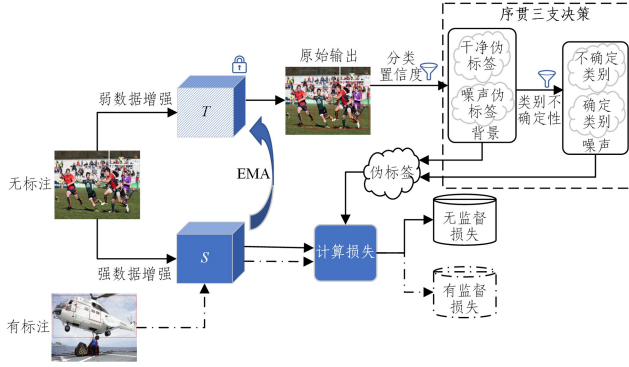


图 1 基于序贯三支决策的半监督检测架构示意图

Fig. 1 Structure of sequential three-way SSOD methods

3.2 基于序贯三支决策的半监督检测算法

噪声标签的处理十分关键。舍弃这些带有噪声的伪标签,意味着需要选取一个较高的阈值,这会导致伪标签中出现非常多的漏检问题,同时也会损失这些噪声标签中有价值的信息;而保留这些噪声伪标签,带来的问题则是噪声对模型的性能产生影响。最新的基于多个阈值的过滤方法 DSL 中,也提出了噪声伪标签的问题,但是其对这些噪声伪标签采取忽略策略,即忽略模型训练当中与这些噪声伪标签配对的候选框的损失值。这固然消除了噪声伪标签带来的漏检与错检问题,但也损失了噪声伪标签中存在的有价值信息。

本文提出了基于序贯三支决策的过滤方法,使用多个阈值将伪标签集合划分为干净前景伪标签、噪声前景伪标签和干净背景伪标签。形式化定义如下:

$$p_i^* = \begin{cases} 1, & p_i^c \geq thr_{high} \\ 0, & thr_{low} \leq p_i^c \leq thr_{high} \\ -1, & p_i^c \leq thr_{low} \end{cases} \quad (2)$$

其中, thr_{high} 和 thr_{low} 分别为过滤伪标签的上阈值和下阈值, p_i^* 取值为 1, 0, -1, 分别代表干净的前景标签、噪声前景标签和干净的背景。对于干净的前景标签和背景标签,在无监督学习的过程中正常训练即可。

对于中间域中的伪标签,考虑到噪声标签大多是正常标签集合中的离群点,输出的噪声伪标签是模型对错误标签训练过拟合的结果。因此,这些噪声标签对于输入的抗扰动性能将显著低于正常的干净标签。模型输出预测的伪标签框之后,在中间域伪标签周围采样一定数量的检测框,并将其输入教师模型以得到这些检测框的预测输出,依据分类模型得到的置信度分数,计算该噪声伪标签框的类不确定性。形式化定义如下:

$$cls_unc_i = \sqrt{\frac{\sum_{s=1}^S (p_s - \hat{p}_i)^2}{S}} \quad (3)$$

其中, S 表示采样的检测框数量, \hat{p}_i 表示伪标签框的预测类别

置信度的平均数, p_s 则是围绕该伪标签框 i 采样得到的检测框对应类别的分类置信度。

对于所有的中间域伪标签,类不确定性低于阈值 thr_{unc} 的可以被认定为具有稳定的预测类别,对此类标签采取负类学习的方法,再根据类别概率分布来学习其中较为确定的负类别的信息;而类不确定性高于阈值 thr_{unc} 的则可以被认定为是模型对错误标签的过拟合导致的,有很大概率存在噪声,在模型训练中将此类标签设置为忽略标签。序贯三支决策的整体过滤流程如图 2 所示。

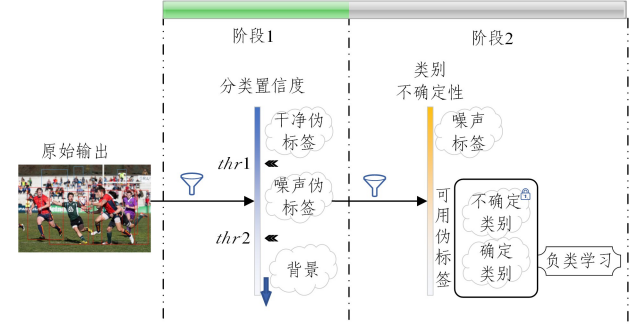


图 2 基于序贯三支决策的过滤方法

Fig. 2 Filter method based on sequential three-way decision

3.3 基于负类学习的降噪方法

所谓负类学习,是相对于正常的模型训练来定义的,是在面对噪声标签时的反向学习过程。以单个样本的多分类交叉熵损失函数为例:

$$L_{CE} = - \sum_{c=1}^{C+1} y_c \cdot \log(p_c) \quad (4)$$

其中, y_c 表示该样本是否属于类别 c ,若真实类别等于 c 则取 1,反之则取 0; p_c 表示该样本属于类别 c 的预测概率, C 表示类别的数量。

当标签是无噪声标签时,直接使用交叉熵损失作为模型的分类损失能够给模型优化提供足够的指导。对多分类任务来说,模型预测概率使用 SoftMax 函数计算得到。

$$p_c = \text{SoftMax}(z_c) = \frac{e^{z_c}}{\sum_{i=1}^C e^{z_i}} \quad (5)$$

其中, z_c 表示模型的第 c 类的输出。交叉熵损失函数中,当损失值趋近于 0 时,对应真实类别的概率趋近于 1。而同一样本不同类别之间因为概率总和为 1 的约束,当真实类别预测概率趋近于 1(100%)时,即意味着其他类别的概率将被抑制为 0。

然而,半监督方法中教师模型产生的伪标签中不可避免地存在噪声标签。对于噪声标签,因为其可能存在某几个类别的预测概率较为接近的情况,所以无法确定其真实类别,故直接使用传统交叉熵损失进行学习会面临噪声问题。但是那些预测概率极小的类别,其不属于此类是确信的,可以将这个确信的信息提供给模型进行学习。本文将这些伪标签中依据预测概率可以极大程度确定不是真实标签的类别统称为负类别,否则称为正类别。因此,对于噪声标签,损失函数设计的预期效果为将其确定的负类别概率抑制到 0,而对于不确定的正类别则忽略其损失。形式化定义如下:

$$p_c^{\text{target}} = \begin{cases} 0, & p_c^i \leq \text{thr} \\ -1 & p_c^i \geq \text{thr} \end{cases} \quad (6)$$

其中, p_c^{target} 表示预期的分类概率, p_c^i 表示教师模型预测的分类置信度, thr 表示用于选择确定负类别的阈值, c 的取值范围为 $[1, C+1]$, 共有 $C+1$ 个取值, 其中第 $C+1$ 类表示背景类别。通过这一噪声正类别挖掘策略, 本文提出的方法中的噪声正类别数量可以根据模型预测概率自适应调整, 而无须提前设置负类别数量作为超参数。此外, 背景类别也可以是噪声正类别。

关于损失函数的初步构想如下:

$$L_{\text{NCE}} = - \sum_{c=1}^{C+1} y_c \times \log(p_c^{\text{pred}}) \times \omega(p_c^{\text{target}}) \quad (7)$$

其中, $\omega(x)$ 函数表示加权函数, 即当 p_c^{target} 取值大于 0 时, 权重为 1, 否则为 0。此时, 当类别 c 为正类别时, $\omega(p_c^{\text{target}})$ 取值为 0; 当类别 c 为负类别时, y_c 取值为 0。因此, 最终损失值 $y_c \times \log(p_c^{\text{pred}}) \times \omega(p_c^{\text{target}})$ 也总保持为 0。

直接使用模权重矩阵对交叉熵损失进行加权并不能达到预期的目的。因此, 本文使用掩模矩阵对模型输出及其标量形式进行加权融合, 从而在最小化负类别概率的同时, 避免噪声信息的影响。新的损失函数形式化定义如下:

$$\begin{aligned} \tilde{p}_c^{\text{pred}} &= p_c^{\text{pred}} \cdot \omega(p_c^{\text{target}}) + \bar{p}_c^{\text{pred}} \cdot (1 - \omega(p_c^{\text{target}})) \\ L_{\text{NCE}} &= - \sum_{c=1}^{C+1} y_c \cdot \log(\tilde{p}_c^{\text{pred}}) \\ p_c^{\text{pred}} &= \text{SoftMax}(x) \end{aligned} \quad (8)$$

其中, $\tilde{p}_c^{\text{pred}}$ 为模型预测输出 p_c^{pred} 之后的标量。当类别 c 为不确定的正类别时, $\omega(p_c^{\text{target}})$ 取值为 0, $\tilde{p}_c^{\text{pred}}$ 使用的是标量, 其产生的梯度不进行回传; 当类别 c 为确定的负类别时, 梯度会正常进行回传, 交叉熵函数对原始正类别的最大化操作同时也会将 p_c^{pred} 抑制为 0。综上分析, 本文提出的 L_{NCE} 损失函数能够在避免噪声影响的情况下尽可能地学习其中有价值的信息。

当模型的分类激活函数为 Sigmoid 函数时, 上述分析也同样成立。

4 实验结果

4.1 数据集

MS-COCO: 由微软收集并标注的大规模公开目标检测数据集。本文分别采用 4 个不同的有监督数据比例 (1%, 2%, 5%, 10%) 进行了实验验证。本文使用 SoftTeacher^[23] 提供的有标注和无标注数据划分方法, 用 5 个随机数种子分别在上述不同的有监督比例下对 MS-COCO 训练集进行了划分。本文使用这些划分后的有标注和无标注子集进行半监督目标检测模型的训练, 使用 MS-COCO val2017 进行性能评估。

Pascal VOC^[24]: 对 Pascal VOC^[24] 数据集, 本文采用 VOC07-trainval 作为有标注数据, 使用 VOC12-trainval 作为无标注数据, 在 VOC07-test 集合上进行性能评估。

4.2 评价指标

本文采用目标检测数据集 MS-COCO 提供的测试工具中的全类平均正确率 (Mean Average Precision, mAP) 作为评估半监督目标检测性能的检测指标, mAP 是以 0.05 为步长在

0.5~0.95 区间上选取 IoU 阈值的各类别平均精度 AP 的均值。

4.3 实验设置

本文采用 Pytorch 和 MMDetection^[25] 框架来实现所提出的方法。采用 Faster-RCNN-FPN 作为目标检测器, 使用在 ImageNet 上预训练的 ResNet-50^[26] 作为模型的骨干网络。

为公平起见, 本文在 8 个 GPU 上进行模型训练, 每个 GPU 的批处理大小为 5, 包含 1 张有监督图片和 4 张无监督图片。迭代次数为 180000, 使用随机梯度下降作为模型的优化器, 学习率为 0.01, 且在前 1000 次迭代中从 0 开始逐步增大学习率到 0.01。无监督损失的权重设置为 2.0。序贯三支决策算法中的下阈值固定为 0.6, 上阈值为 0.7。强弱数据增强设置与 Soft Teacher 中保持一致。教师模型的参数由学生模型经过 EMA 得到, 平滑系数取值为 0.9995, 更新步长为 1。

4.4 实验结果与分析

如表 1 所列, 在 MS-COCO 数据集上, 本文的方法在不同的有监督数据比例下都产生了相当优异的结果。

表 1 MS-COCO 数据集上不同比例的性能

Table 1 Performance on MS-COCO with different label ratios

方法	1%	2%	5%	10%
Supervised	9.05	12.70	18.47	23.86
CSD ^[27]	10.51	13.93	18.63	22.46
STAC	13.97	18.25	24.38	28.64
Instant Teaching ^[28]	18.05	22.45	26.75	30.40
Interactive ^[29]	18.88	22.43	26.37	30.40
Humble Teacher ^[30]	16.96	21.72	27.70	31.61
SoftTeacher	20.46	—	30.74	34.04
Unbiased Teacher ^[31]	20.75	24.30	28.27	31.50
MUM	21.88	24.84	28.52	31.87
Ours*	20.60	25.10	31.90	35.20

首先, 通过对比有监督的 baseline 实验, 验证本文提出的半监督方法的有效性。相比有监督的 baseline 方法, 基于序贯三支决策的方法在 1%, 2%, 5%, 10% 的有标注数据比例下检测精度分别提升了 11.55%, 12.40%, 13.43% 和 11.34%。不难看出, 在与最新的其他半监督目标检测方法的比较中, 本文提出的半监督方法都取得了更加优越的性能。当有标注数据比例为 1% 时, 本文方法的检测性能略低于 MUM, 这是因为本文的阈值超参数都是依据 10% 比例下的消融实验确定的, 其在 1% 比例下可能并非最优超参组合。

为了进一步验证所提方法的泛化性, 在 Psacal VOC 数据集上也进行了半监督检测实验, 并与其他最新方法进行比较。如表 2 所列, 本文方法在 PASCAL VOC 上实现了 82.00% 的 AP50 检测精度和 54.11% 的 mAP 精度 (括号中的结果为将 PASCAL VOC 转换为 MS-COCO 格式进行测试的精度)。

表 2 Pascal VOC 数据集上的性能

Table 2 Performance on Pascal VOC dataset

方法	mAP	AP50
CSD	—	74.70
STAC	44.64	77.45
Instant Teaching	48.70	78.30
Interactive	46.23	—
Humble Teacher	53.04	—
Unbiased Teacher	48.70	77.40
MUM	50.22	—
Ours	54.11 (54.20)	82.00 (84.70)

4.5 消融实验

本文在 MS-COCO 数据集中 10% 比例有标注数据的设定下进行消融实验。此外,为了加快消融实验的进程,在不失公平的前提下,本文对部分实验选取了 FP16 精度进行加速,这将导致部分消融实验中的检测精度略低于表 1 中的数据 (MS-COCO 数据集 10% 有标注数据精度从 35.2% 下降到 34.7%)。为了方便起见,后续表格中,带 * 的结果表示 FP16 浮点精度下的实验。

1) 各模块有效性

如表 3 所列,实验 1 和实验 2 的对比结果表明,简单的一阶段三支决策方法与负类损失的结合并不能有效地实现对噪声标签的筛选操作,其精度只能达到 32.9%,远低于 baseline 的基准精度 34.2%。实验 1 和实验 4 的对比结果显示,本文通过引入类不确定性度量,进行第二阶段的三支决策,将模型精度从 34.2% 提升到了 34.7%。实验 3 和实验 4 的精度对比结果表明,引入负类损失学习可以摒弃噪声标签中无法确定的正样本类别,将模型的精度从 34.4% 提升到 34.7%。

表 3 模块消融实验

Table 3 Ablation of different modules

实验序列	Three-way	Sequential Three-way	NCE	mAP/%
实验 1				* 34.2
实验 2	✓		✓	* 32.9
实验 3		✓		* 34.4
实验 4		✓	✓	* 34.7

2) 分类置信度阈值

为了找到第一阶段三支决策中合适的上下阈值,本文进行了简单的消融实验。本文并没有通过大规模的实验对阈值进行网格搜索,而是在目前通用的过滤阈值 $thr=0.7$ 上下进行了对比实验,具体对比结果如表 4 所列(对比阈值设置时类不确定性阈值设计为 0.04)。结果表明,上下阈值设置为 0.6 和 0.7 时模型表现更佳。此外,无特殊说明,后续的消融实验中上下阈值取值均为 0.6 和 0.7。

表 4 过滤阈值的选取

Table 4 Comparison of different thresholds (单位: %)

Threshold	mAP	AP50	AP75
0.6~0.7	* 34.40	* 53.4	* 37.2
0.7~0.8	* 34.00	* 52.8	* 37.1

3) 类不确定性阈值

类不确定性是序贯三支决策第二阶段用于辅助决策的信息。考虑到噪声标签大多是正常标签集中的离群点,将类不确定性度量低于阈值的伪标签框看作有效的标签,其中前景类别的预测可能并不准确,但是其输出的分类分布中依旧存在可学习的信息。

为了找到合适的类不确定性筛选阈值,我们进行了表 5 中所展示的一系列实验,表中列举了模型在不同的类不确定性阈值下的 mAP 精度。其中,当类不确定性过滤阈值设置为 0.06 时,模型达到了最佳的精度 34.70%。

表 5 不同类不确定性阈值的精度

Table 5 mAP of different uncertainty thresholds

Uncertainty Threshold	mAP
0.02	* 34.20
0.04	* 34.40
0.06	* 34.70
0.08	* 33.9
0.10	* 33.4

通过对比不同不确定性阈值下模型训练过程中的精度变化曲线,可以观察到不同程度的噪声信息比例对模型性能的影响。从图 3 中可以看到,当阈值大于 0.06 之后,随着阈值增大,可学习的伪标签数量变多,模型的精度反而开始下降;且在阈值为 0.08 和 0.10 的实验中,模型的精度都是先上升,后剧烈波动,而其他设置下的精度曲线则较为平缓,这是噪声信息增多带来的 confirmation bias 现象。这一结果也进一步说明了噪声标签对模型性能的影响,验证了本文提出的序贯三支决策过滤方法中第二阶段过滤的必要性。

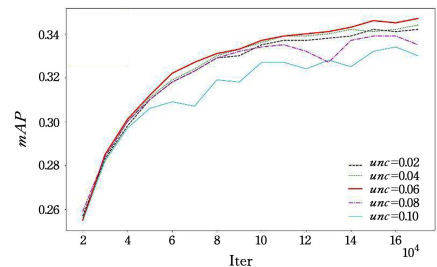


图 3 类不确定性阈值的影响

Fig. 3 Influence of uncertainty thresholds

结束语 本文提出了一种多阶段划分伪标签的半监督目标检测方法,利用序贯的三支决策算法,在第一阶段利用分类分数划分出中间域伪标签;在第二阶段联合使用伪标签的分类分布和类不确定性度量,对伪标签进行进一步的划分和利用。实验结果表明,本文提出的方法在 Pascal VOC 和 MS-COCO 数据集的半监督实验设定上有良好的性能增益。

未来的研究工作中,拟将该算法扩展到检测模型的回归分支上,进一步提升半监督目标检测的性能。此外,计划将本文提出的方法应用到 3D 目标检测任务上,以验证算法的泛化性能。

参考文献

- [1] XU Y, SHANG L, YE J, et al. Dash: Semi-supervised learning with dynamic thresholding [C] // International Conference on Machine Learning (ICML). Cambridge MA: JMLR, 2021: 11525-11536.
- [2] YUE X D, CHEN Y F, MIAO D Q, et al. Fuzzy Neighborhood Covering for Three-way Classification [J]. Information Sciences, 2020, 507: 795-808.
- [3] WEI X S, XU H Y, ZHANG F, et al. An Embarrassingly Simple Approach to Semi-Supervised Few-Shot Learning [J]. Advances in Neural Information Processing Systems, 2022, 35: 14489-14500.
- [4] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO:

- Common Objects in Context [C] // European Conference on Computer Vision (ECCV). Cham: Springer, 2014: 740-755.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] // European Conference on Computer Vision (ECCV). Cham: Springer, 2016: 21-37.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [C] // Conference on Computer Vision (ICCV). Cham: Springer, 2017: 2980-2988.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2014: 1714-1722.
- [8] GIRSHICK R. Fast R-CNN [C] // International Conference on Computer Vision (ICCV). Cham: Springer, 2015: 1440-1448.
- [9] REN S, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] DUAN K W, BAI S, XIE L X, et al. CenterNet: Keypoint Triplets for Object Detection [C] // IEEE/CVF International Conference on Computer Vision (ICCV). Cham: Springer, 2019: 1-16.
- [11] TIAN Z, SHEN C H, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection [C] // International Conference on Computer Vision (ICCV). NJ: IEEE, 2019: 9627-9636.
- [12] PHAM H, DAI Z, XIE Q, et al. Meta pseudo labels [C] // Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2021: 11557-11568.
- [13] XIE Q, LUONG M T, HOVY E, et al. Self-training with noisy student improves imagenet classification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2020: 10687-10698.
- [14] ZOPH B, GHIASI G, LIN T Y, et al. Rethinking pre-training and self-training [J]. Advances in Neural Information Processing Systems, 2020, 33: 3833-3845.
- [15] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout [J]. arXiv: 1708. 04552, 2017.
- [16] ZHANG H, CISSE M, DAUOHIN Y N, et al. mixup: Beyond empirical risk minimization [J]. arXiv: 1710. 09412, 2017.
- [17] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). NJ: IEEE, 2019: 6023-6032.
- [18] SOHN K, ZHANG Z, LI C L, et al. A simple semi-supervised learning framework for object detection [J]. arXiv: 2005. 04757, 2020.
- [19] LI H, WU Z, SHRIVASTAVA A, et al. Rethinking pseudo labels for semi-supervised object detection [C] // Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). CA: AAAI, 2022: 1314-1322.
- [20] KIM J M, JANG J Y, SEO S, et al. MUM: Mix Image Tiles and UnMix Feature Tiles for Semi-Supervised Object Detection [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2022: 14492-14501.
- [21] CHEN B, LI P, CHEN X, et al. Dense learning based semi-supervised object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2022: 4815-4824.
- [22] CHEN C, DEBATTISTA K, HAN J. Semi-supervised object detection via virtual category learning [J]. arXiv: 2207. 03433, 2022.
- [23] XU M, ZHANG Z, HU H, et al. End-to-end semi-supervised object detection with soft teacher [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). NJ: IEEE, 2021: 3060-3069.
- [24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge [J]. International Journal of Computer Vision, 2009, 88: 303-308.
- [25] CHEN K, WANG J, PANG J, et al. MMDetection: Open mmlab detection toolbox and benchmark [J]. arXiv: 1906. 07155, 2019.
- [26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [27] JEONG J, LEE S, KIM J, et al. Consistency-based semi-supervised learning for object detection [C] // ICCV 2019. 2019.
- [28] ZHOU Q, YU C, WANG Z, et al. Instant-teaching: An end-to-end semi-supervised object detection framework [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2021: 4081-4090.
- [29] YANG Q, WEI X, WANG B, et al. Interactive self-training with mean teachers for semi-supervised object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2021: 5941-5950.
- [30] TANG Y, CHEN W, LUO Y, et al. Humble teachers teach better students for semi-supervised object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). NJ: IEEE, 2021: 3132-3141.
- [31] LIU Y C, MA C Y, HE Z, et al. Unbiased teacher for semi-supervised object detection [J]. arXiv: 2102. 09480, 2021.



SONG Faxing, born in 1999, postgraduate. His main research interests include deep learning, object detection and semi-supervised object detection.



MIAO Duoqian, born in 1964, Ph. D., professor, Ph. D supervisor. His main research interests include artificial intelligence, machine learning, rough set and big data analysis.