计算机科学
COMPUTER SCIENCE

基于边框距离度量的增量目标检测方法

刘冬梅，徐洋，吴泽彬，刘倩，宋斌，韦志辉

引用本文

刘冬梅，徐洋，吴泽彬，刘倩，宋斌，韦志辉. 基于边框距离度量的增量目标检测方法[J]. 计算机科学, 2022, 49(8): 136-142.

LIU Dong-mei, XU Yang, WU Ze-bin, LIU Qian, SONG Bin, WEI Zhi-hui. Incremental Object Detection Method Based on Border Distance Measurement[J]. Computer Science, 2022, 49(8): 136-142.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220100132

基于边框距离度量的增量目标检测方法

刘冬梅　徐　洋　吴泽彬　刘　倩　宋　斌　韦志辉
南京理工大学计算机科学与工程学院　南京 210094
(dongmei@njust.edu.cn)

摘　要　增量学习在图像分类中已经获得了不错的效果,但是将增量学习技术直接应用于多类目标检测具有一定的挑战性。相比图像分类,目标检测是一项更复杂的任务,因为它结合了分类和边框回归的问题。目前最先进的增量目标检测器大多采用基于知识蒸馏的外部固定区域建议方法,该方法需耗费大量的时间和成本。由于单阶段检测器缺少旧类别的标注和区域建议信息,检测器通常会将旧类目标识别为背景,从而导致灾难性遗忘,因此提出了一种基于边框距离度量的标签选择算法。该算法利用旧模型检测结果和现有的数据集标签,通过度量边框重合度进行选择与合并,弥补了新数据集中旧类目标注释缺失的问题,缓解了灾难性遗忘。同时设计了一个注意力残差模块,该模块通过将注意力模块与残差模块相结合,在特征提取网络的不同深度均能提取可鉴别性特征,进一步提升了模型检测新旧类目标的精度。在单阶段检测框架中实现了该方法,同时在 PASCAL VOC 数据集上验证了该方法的有效性。与目前最好的方法相比,所提模型检测旧类别目标的平均精度值 *mAP* 高出了 2.8%,总体的平均精度值 *mAP* 高出了 2.1%。所提方法得到的伪标签有效缓解了遗忘问题,注意力残差模块的设计提升了模型的检测精度。

关键词:目标检测;标签选择;增量学习;注意力模块;灾难性遗忘;伪标签
中图法分类号　TP391

Incremental Object Detection Method Based on Border Distance Measurement

LIU Dong-mei,XU Yang,WU Ze-bin,LIU Qian,SONG Bin and WEI Zhi-hui
School of Computer Science and Engineering,Nanjing University of Science and Technology,Nanjing 210094,China

Abstract　Incremental learning has achieved good results in image classification,but it is challenging to apply incremental learning to multi-class object detection. Object detection is more complex than image classification,which combines classification and border regression. At present,the most advanced incremental object detectors adopt the external fixed region suggestion method based on knowledge distillation,which consumes a lot of time and cost. For single-stage detectors,due to the lack of annotation and region advice information for the old class,old objects are usually identified by the detector as the background,resulting in catastrophic forgetting. In this paper,a label selection algorithm based on border distance metric is proposed. It uses the detection results of the old model and the existing dataset labels to select and merge by measuring the coincidence of the bounding boxes,making up for the lack of annotations of the old objects in the new dataset and alleviating catastrophic forgetting. In addition,a module that combines the attention module with the residual module is designed to extract discriminative features at different depths in feature extraction network,to further improve the detection accuracy of model. The proposed method is implemented in the single-stage detection framework,and the effectiveness of the method is verified on PASCAL VOC dataset. Compared with the best model at present,the average accuracy value of the old object and all objects improves by 2.8% and 2.1%,respectively. The pseudo-labels obtained by the proposed method greatly alleviate the forgetting problem,and the attention residual module improves the detection accuracy of the model.

Keywords　Object detection,Label selection,Incremental learning,Attention module,Catastrophic forgetting,Pseudo label

到稿日期:2022-01-17　返修日期:2022-02-17
基金项目:国家自然科学基金(61772274,62071233,61971223,61976117);江苏省自然科学基金(BK20211570,BK20180018,BK20191409);中央高校基金项目(30917015104,30919011103,30919011402,30921011209);中国博士后基金(2017M611814,2018T110502)
This work was supported by the National Natural Science Foundation of China(61772274,62071233,61971223,61976117),Natural Science Foundation of Jiangsu Province(BK20211570,BK20180018,BK20191409),Fundamental Research Funds for the Central Universities(30917015104,30919011103,30919011402,30921011209) and China Postdoctoral Science Foundation(2017M611814,2018T110502).
通信作者:徐洋(xuyangth90@gmail.com)

1 引言

作为一种计算机视觉任务，目标检测对许多应用程序而言有着不可替代的作用，但是复杂的系统需要大量的训练时间去学习模型，最终检测器的性能在很大程度上依赖于一组有代表性的训练实例。在实际应用中，获取代表性的训练数据既昂贵又耗时，因此数据在一段时间内以小批量的形式出现并不罕见。随着时间的推移，可能会出现模型以前没有学习过的新类，如果用新类别数据直接微调基本模型，由于以前的数据不可访问而缺少旧类别的标注，这种方法将会遭受灾难性遗忘，即旧类别的检测性能严重下降，同时还会影响新类别的检测精度。

在这项工作中，我们研究增量目标检测以解决目标检测器无法适应新数据的问题。增量学习在图像分类领域的应用得到了发展，但是若直接将该方法应用在目标检测任务中具有很大的挑战性，主要原因如下：1）图像分类属于单任务，而目标检测属于多任务检测，一幅图像中可能同时含有新类和旧类多种目标；2）目标检测任务中数据集只包含新类标签，旧数据标注的缺失将导致分类时误将旧类识别为背景。

为了避免上述目标检测任务的灾难性遗忘问题，最简单的做法是使用旧类和新类全部数据重新开始训练模型，但旧数据由于隐私、存储等原因可能无法访问，且这种方法十分耗时又计算昂贵，因为需要重新标注包含到目前为止的所有类别。另一种是寻找合适的方法挑选出具有代表性的旧数据并将其存储起来，然后联合新数据一起训练模型，在这种方法下，随着需求的增加，存储的数据越来越多，需要非常大的内存开销，与上述方法存在相同的问题。

增量学习应用于目标检测任务中要求模型能在不忘记旧类别的基础上学习检测新的类别，严格来说，微调旧模型以检测新类别的目标时不能访问旧类别的数据，同时检测模型的规模应保持相对不变。结合现有的研究工作，针对图像中旧类目标和新类目标同时出现从而导致模型将旧类识别为背景的问题，本文主要做出了以下贡献。

（1）提出了一种数据标签选择算法，利用旧模型的检测结果和现有的数据标签生成用于训练的伪标签，弥补了新数据集中存在的旧类目标注释缺失问题，缓解了灾难性遗忘。

（2）提出了一种注意力残差模块，该模块引入通道注意力和空间注意力，在特征提取网络的不同深度中学习重要特征信息，进一步提升了模型检测新旧类目标的精度。

（3）在不需要访问旧数据集的前提下，在单阶段检测框架中实现了该增量检测方法，并验证了该方法的有效性。

本文第 2 节简要介绍了增量目标检测的相关工作；第 3 节详细介绍了本文提出的方法模型；第 4 节介绍了数据集和评估准则并进行了多个实验的对比，对所提方法的有效性进行了论证；最后总结全文并展望未来。

2 相关工作

本文的工作目标是在不访问旧数据集的前提下使目标检测器可以增量检测新的类别，同时不忘记如何检测旧类目标。本节首先介绍目前先进的目标检测框架，然后再讨论基于增量学习的目标检测方法。

2.1 目标检测

随着深度学习的发展，目标检测框架主要分为两类，即双阶段检测和单阶段检测。具有代表性的双阶段检测框架 RCNN（Region Convolutional Neural Network）主要由 3 个模块组成。首先通过选择性搜索或其他方法生成独立的候选区域，利用这些区域获取可用的候选集；然后通过卷积神经网络提取特征，它从每个区域提取固定长度的特征向量；最后利用特定类的线性支持向量机分类器预测每个区域内存在的目标类别并定位。由于在大量重叠区域上进行冗余的特征计算，因此检测速度非常慢。后来，He 等[1]提出了空间金字塔池化网络，该网络引入了空间金字塔池（Spatial Pyramid Pooling，SPP）层，它使 CNN 能够生成固定长度的表示，而无须对图像感兴趣的区域进行重调。更快的 RCNN[2]（Faster RCNN）引入了一个区域候选网络（Region Proposal Network，RPN），它与检测网络共享全图像卷积特征，从而实现了几乎无成本的候选区域。Lin 等[3]在此基础上提出了特征金字塔网络（Feature Pyramid Network，FPN），与之前大多数检测器只在网络的顶层进行检测不同，FPN 提出了一种具有横向连接的自顶向下体系结构，用于构建各种尺度下的高级语义特征图。

虽然双阶段检测器在检测精度上取得了不错的成果，但由于存在大量的特征冗余计算，导致检测速度非常慢。随着目标检测的广泛应用，对检测系统的速度要求越来越高，Redmon 等[4]提出了深度学习时代首个单级检测器 YOLO（You Only Look Once），它将目标检测框架视为一个回归问题，将图像划分为 $S\times S$ 个网格单元，当目标的中心落在网格中，该单元负责预测该对象，整个检测管道是单一网络，因此可以进行端到端直接优化，但是在检测速度提升的同时，检测精度有所下降。Redmon 等[5]对其不断进行优化，提出了 YOLOv3 检测模型，该模型的先验框采用聚类生成，模型设计采用特征金字塔 FPN 的思想，在多个尺度上进行预测，使速度和精度得到了较好的平衡。本文基于 YOLOv3 检测框架实现所提方法。Zhou 等[6]提出了中心网络（Centernet），它将特征图中的每个像素视为形状不可知的锚，对于特征图中的每个像素，Centernet 会预测其是否为对象中心，并回归宽度和高度，它不需要非极大值抑制（Non-Maximum Suppression，NMS）进行后处理，并且实现了速度和精度之间的良好折衷。Tian 等[7]设计的 FCOS 模型对每个对象预测对象内部的一个点，并回归从该点到边界框四边的距离，该网络利用地面真值框内的所有点来预测边界框，同时采用多级特征金字塔网络（FPN）来处理低召回率和真实边界框重叠的问题，取得了与大多数基于锚点的检测器相当的检测精度。

2.2 增量目标检测

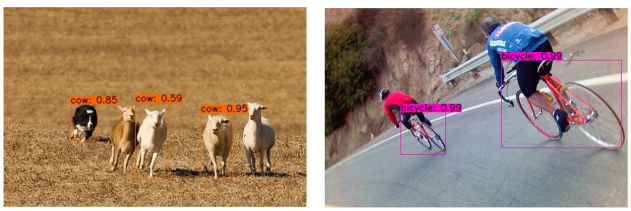
增量学习指一个学习系统能不断地从新样本中学习新的知识，并能保存大部分已经学习到的知识，在图像分类中的应用得到了很大发展，但对于检测任务来说存在极大的挑战。

目前一种常见的方法是只使用新数据对旧类模型进行正则化微调，迫使模型不忘记学到的旧类别的知识，但是这种方法容易导致新类别的学习十分有限，因为过度的正则化会导致模型难以学习新类，虽然旧类别的检测效果能基本保持，但是新类的学习却变得十分艰难，这违背了增量检测的初心。Shmelkov 等[8]将知识蒸馏思想应用于增量检测任务中，使用神经网络以端到端的方式对任务进行建模，与 Li 等[9]在图像分类中的做法相似。该方法借鉴模型压缩领域的思想，核心是使用双模型网络，最小化来自旧模型和新模型关于旧类别响应之间的差异，并且提出一个损失函数以平衡模型对新类预测之间的相互影响，即交叉熵损失和一个新的类别蒸馏损失。这种方法在基于外部建议的双阶段检测模型上得到验证，但是对于单阶段目标检测器来说仍无法解决遗忘问题，且这种方法在训练过程中存在的噪声使得模型不易收敛。Li 等[10]从旧模型中提取 3 种类型的知识，模拟旧模型在对象分类、边框回归和特征提取中的行为，同时设计了一个实时收集训练数据的算法，并使用类别和边框注释自动标记图像，最后在单阶段深度目标检测模型中进行端到端的训练。该方法的基本思想是阻止对旧类别的预测进行更改，强迫新旧模型对旧类别的预测达成共识。Zhang 等[11]认为现有的增量检测方法往往会生成一个具有偏向性的模型，因此他们提出了一种深度模型巩固(Deep Model Consolidation，DMC)的增量学习方法。该方法的核心思想是，首先使用标记数据为新类训练一个单独的检测模型，然后使用公开可用的未标记的辅助数据通过一个双蒸馏训练目标合并新旧模型，克服了原始数据不能访问导致的潜在困难。

3 本文方法

3.1 基于边框距离度量的标签选择算法

旧数据集无法访问和新数据集中缺少旧类的标注信息使增量检测变得十分艰难，通过实验，我们发现直接使用旧类检测模型检测新数据集时，对于存在的旧类别目标分类和定位比较准确，同时也存在大量的类别预测错误但边框预测相差无几的情况(如图 1(a)所示，旧模型将狗和羊预测为牛，分类错误，但位置信息基本正确，图 1(b)中预测的旧目标分类和定位都正确)。因此可以先获取旧模型生成的关于旧类别的标注，然后提取新数据的真实标签，通过计算两者的边框距离，由标签选择算法生成旧类的类别和边框标注，这样的先验信息刚好可以弥补旧类标注缺失的问题，从而使模型在学习检测新的类别的同时不忘记旧类别。



(a)Wrong label (b)Correct label

图 1 旧类检测模型生成的伪标签

Fig. 1 Pseudo-labels generated by old detection model

如图 1(a)所示，图像包含的对象和旧类别中的牛相似，导致旧类检测器分类错误，但边框检测的位置基本正确，联合数据集的真实标签可以通过标签选择将错误的标注去除。基于边框距离度量的标签选择方法受到 IoU(Intersection over Union)的启发。在一般检测任务中，IoU 是为了检测模型预测目标位置准确度的一个标准，其只关注重叠区域的不同。它的改进版本包括 GIoU(Generalized IOU)，DIoU(Distance-IoU)和 CIoU(Complete-IOU)，都是用于计算位置准确度的一种度量方式。GIoU 对尺度不敏感；DIoU 将边框的距离、重叠率和尺度都考虑进去，更加符合目标框回归的机制，但没有考虑到长宽比问题；CIoU[12]考虑了长宽比的问题，在 DIoU 的基础上加入了惩罚项，解决了前面几种度量方法存在的问题。通过实验比较，在去除错误的边框时，CIoU 可以更好地进行标签选择，对比 3 种不同的方式，CIoU 得到的结果最理想。因此本文采用 CIoU 作为衡量位置准确度的标准，其表达式为：

$$CIoU=\frac{\rho^2(b,b^{gt})}{c^2}+\alpha v \tag{1}$$

其中，$\rho(b,b^{gt})$用于度量两个边框中心点的距离，α 是权重函数，v 用来度量宽高比的一致性，定义为：

$$\alpha=\frac{v}{(1-IoU)+v} \tag{2}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{\omega^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{3}$$

式(1)中，c 表示预测框 b 和真实边框 b^{gt} 的最小包围框的对角线长度，式(3)中，w 和 h 分别表示预测框的宽和高，ω^{gt} 和 h^{gt} 分别表示真实边框的宽和高。通常旧模型分类错误但定位正确的目标是新类目标，因此该图像标签中会包含该目标的位置信息。在本文中，旧模型预测图像得到的预测框表示为 b，该图像标签中的边框表示为 b^{gt}，通过计算两个边框的 $CIoU$ 值找出结果大于设定阈值的预测框，该预测结果与真实边框高度重叠即为错误的标注，将其从预测结果中去除，直到所有的边框全部参与计算完毕，然后将剩下的正确结果与数据集的真实标签合并得到用于训练的伪标签集合。具体的算法如算法 1 所示。对于可能存在的漏检情况，在第 4 节中，通过使用全部标签训练模型检测的结果与本方法对比，表明了其对模型的检测能力不会造成直接的影响。

算法 1 基于边框距离度量的标签选择算法

输入：数据集 X，真实标签 **Y**，旧类检测模型 M，CIOU 阈值 h=0.5

输出：伪标签 **G**

1. for X_i in X：# X_i 表示数据集 X 中的训练样本(i=1,2,…,N；j=1,…)
2. $M(X_i)\rightarrow \mathbf{Y}_i'$ # 旧模型 M 从样本 X_i 中得到旧类目标的标签 $\mathbf{Y}_i'$
3. for y_j', y_j in $\mathbf{Y}_i'$, $\mathbf{Y}_i$：# 只使用 $\mathbf{Y}_i'$ 和 $\mathbf{Y}_i$ 中的边框标签
4. if CIoU ($\mathbf{y}_j'$, $\mathbf{y}_j$) > h：# $\mathbf{Y}_i\in\mathbf{Y}$ 表示样本 X_i 的真实标签
5. delete $\mathbf{y}_j$ # $\mathbf{y}_j\in\mathbf{Y}_i$；$\mathbf{y}_j'\in\mathbf{Y}_i'$；$\mathbf{y}_j$ 和 $\mathbf{y}_j'$ 表示边框坐标
6. end
7. end
8. end # 直到所有的样本全部选择处理完毕

9. $\mathbf{G}=\mathbf{Y}\ \mathbf{U}\ \mathbf{Y}'$ #将真实标签 $\mathbf{Y}$ 与经过选择后的 $\mathbf{Y}'$ 合并

其中阈值 h 的值是通过实验选取的，存在少部分相似类别的错误框不会对模型的精度造成很大影响，但阈值过低，正确的伪标签会被去除。例如，一张图片中包含了一个人骑着自行车，自行车属于旧类，人属于新类，在进行标签选择时，若阈值设置过低，两者的重叠度较大，超过阈值，自行车的标签会被去除，导致模型漏检。如植物和鸟，旧模型预测时，会将植物的部分预测为鸟，若是阈值设置过大，则无法将错误的标签去除，因此本文将阈值 h 设置为 0.5。

基于边框距离度量的标签选择算法充分利用了旧模型预测的结果，相似目标的检测结果类别差异较大，但检测模型对目标的位置并不敏感，在这种前提下，我们利用存在的新类标签找出误检的标注，尽可能地保留正确的旧目标的类别和位置信息。如图 2 所示，对于输入的图像，我们首先使用旧模型进行检测并得到预测结果，然后计算预测框与图像真实边框的 $CIoU$ 值，当结果大于预设的阈值，则将该标注从预测结果中删除，直到所有图像的预测框全部计算选择完毕，最后将经过处理的所有预测结果与数据集的真实标签合并，用于训练增量检测模型。训练新的模型时，我们通过知识迁移使用旧类检测模型的权重初始化增量检测模型 M'，最终训练的模型可以实现检测所有类别。

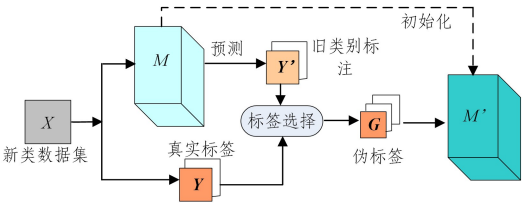


图 2 基于边框距离度量的标签选择过程

Fig. 2 Label selection process based on border distance measurement

3.2 基于残差注意力的特征提取

我们发现，对于经过标签选择算法得到的关于旧类的先验信息弥补了旧类标注缺失的问题，但在训练增量模型的过程中，随着迭代次数的增加，特征提取的学习逐渐由关注之前学习到的旧类的特征倾向于新类别的特征信息。因此，我们在特征提取阶段引入注意力模块。通过实验得出，本文设计的注意力残差模块不仅提高了模型的整体精度，同时提高了模型在旧类别上的检测精度。

注意力残差模块是受 Dhar 等[13]的启发，他们通过实验得出，在增量学习过程中，基于蒸馏思想的方法在训练检测过程中类别蒸馏损失几乎不变，但注意力关注的区域已经发生了很大变化。他们提出的加入注意力蒸馏损失的方法在图像分类任务中取得了不错的成果，验证了注意力模块关注的特征信息让整个模型的检测结果受益。因此，本文根据 Woo 等[14]提出的卷积块注意力模型(Convolutional Block Attention Module，CBAM)，将注意力模块与骨干网中的残差模块进行结合，设计了一个基于注意力残差模块的特征提取网络。注意力残差模块在网络不同层次中学习重要特征信息，具体构成如图 3 所示。

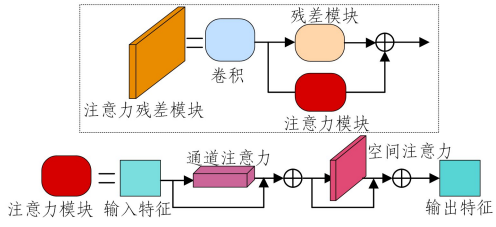


图 3 注意力残差模块

Fig. 3 Attention residual module

通道注意力模块将输入的特征分别经过基于宽度和高度的全局池化和平均池化，然后再送入一个共享网络(Multilayer Perceptron，MLP)中，最后使用逐元素求和⊗合并输出特征向量；空间注意力模块将经过通道注意力输出的特征图在空间维度上进行压缩，然后应用全局池化和平均池化进行处理，整个注意力模块的处理可表达为：

$$\boldsymbol{F}'=M_{ch}(\boldsymbol{F})\otimes\boldsymbol{F} \tag{4}$$

$$\boldsymbol{F}''=M_{sp}(\boldsymbol{F}')\otimes\boldsymbol{F}' \tag{5}$$

其中，$\boldsymbol{F}$ 是输入特征，$\boldsymbol{F}'$ 表示通道注意力的输出特征，$\boldsymbol{F}''$ 表示空间注意力的输出特征，$M_{ch}(\boldsymbol{F})$ 和 $M_{sp}(\boldsymbol{F})$ 分别表示通道注意力模块和空间注意力模块的处理过程，计算式如下：

$$M_{ch}(\boldsymbol{F})=\sigma(MLP(AvgPool(\boldsymbol{F}))+MLP(MaxPool(\boldsymbol{F}))) \tag{6}$$

$$M_{sp}(\boldsymbol{F})=\sigma(f^{7\times7}([AvgPool(\boldsymbol{F});MaxPool(\boldsymbol{F})])) \tag{7}$$

其中，σ 是 sigmoid 操作，$AvgPool(\boldsymbol{F})$ 表示平均池化操作，$MaxPool(\boldsymbol{F})$ 表示最大池化操作，7×7 表示卷积核的大小。

本文的损失函数为边框损失、类别损失、置信度损失 3 者之和，其中置信度损失和类别损失采用标准的交叉熵损失，边框损失采用 CIoU 计算损失，CIoU 将边框的距离、重叠率、尺度以及惩罚项都考虑进去，使得目标边框回归变得更加稳定，不会出现训练过程中发散等问题。损失函数如下：

$$Loss=L_{\text{local}}+L_{\text{prob}}+L_{\text{conf}} \tag{8}$$

其中，边框损失 L_{local}、类别损失 L_{prob}、置信度损失 L_{conf} 分别定义如下：

$$L_{\text{local}}=\lambda\sum\sum(2-w\times h)(1-CIoU) \tag{9}$$

$$L_{\text{prob}}=\sum I^{\text{obj}}\sum[p'\log(p)+(1-p')\log(1-p)] \tag{10}$$

$$L_{\text{conf}}=-\sum\sum I^{\text{obj}}[C'\log(C)+(1-C')\log(1-C)]-\lambda\sum\sum I^{\text{noobj}}[C'\log(C)+(1-C')\log(1-C)] \tag{11}$$

其中，I^{obj} 和 I^{noobj} 分别表示相应网格是否负责预测该目标，C 和 C' 分别表示标签置信度和预测置信度，p' 和 p 分别表示预测类别概率和标签类别概率。

3.3 两阶段训练与多尺度预测

在增量检测模型的训练过程中，通过 Nvidia GTX2080 显卡对网络进行加速训练，交并比的阈值设为 0.5，动量参数设为 0.9995，初始学习率为 0.0001，采用余弦衰减方式进行递减，批次 *batch_size* 设为 4，锚框的大小是对当前数据集进行聚类得到，模型优化采用 Adam 优化器。在实验中根据知识迁移思想用原始的旧模型权重初始化增量检测模型，将训练过程分为两个阶段，分别训练 30 轮和 40 轮。第一阶段只训练注意力模块和最后用于预测的 3 个卷积层的权重，在旧模型权重的基础上微调模型，防止模型太快遗忘旧类别的知识；

第二阶段更新所有权重参数，训练整个网络。此外，训练数据采用水平翻转和随机缩放来进行数据增强，以避免模型出现过拟合。网络模型的整体结构图如图 4 所示，最终在多个尺度上进行预测。

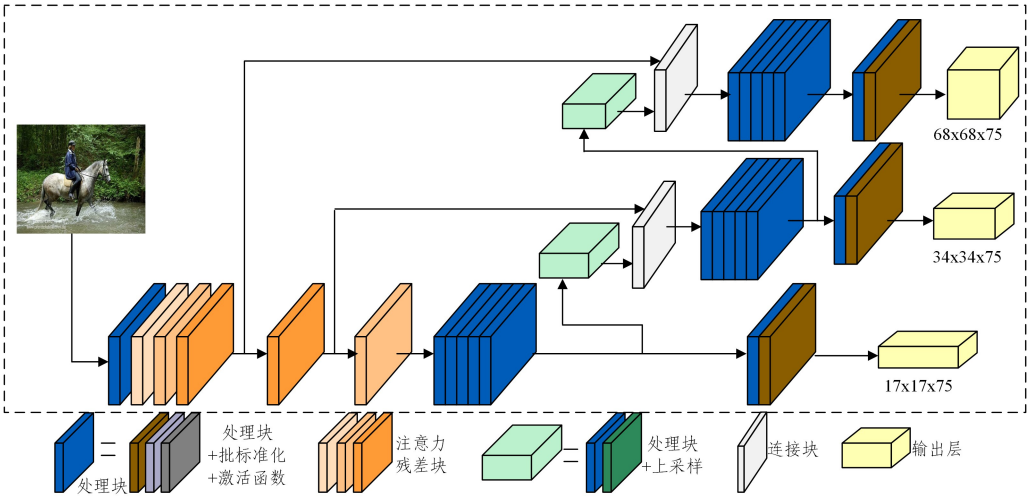


图 4 网络模型整体结构

Fig. 4 Overall structure of network model

4 实验

4.1 数据集

从 PASCAL VOC 数据集中提取包含前 10 类(1－10)的所有图片用来训练一个检测模型，作为旧模型 M，然后从 VOC2007 训练集中提取所有包含剩下的 10 类(11－20)作为增量检测的新类数据集。在训练增量检测模型时，使用的数据集中只含有新的 10 个类别的真实标签，不含有旧类别的任何标注信息。其中，训练旧模型的数据集中包含 VOC2007 和 VOC2012 前 10 类目标训练集，共计 10269 张图片；训练增量检测模型的数据集中包含 VOC2007 后 10 类目标，共计 3340 张图片。最后，所提模型的测试集使用 VOC2007 测试集，共计 4952 张图片。

4.2 检测结果

首先旧模型在包含前 10 类目标的旧数据上进行训练，得到基类检测器，从 VOC2007 测试集中提取所有包含前 10 类的图像作为旧类的测试集。图 5 给出了旧模型在该测试集上的检测结果，可以得出，旧模型在旧数据集上的平均检测精度 mAP 值为 81.6%。

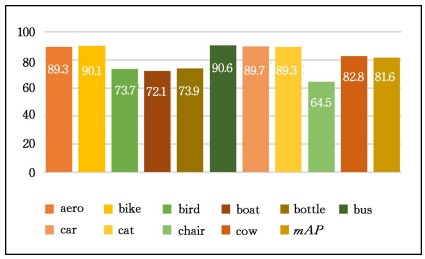


图 5 旧模型目标检测精度

Fig. 5 Accuracy of object detection based on old model

在进行增量训练时，我们首先用旧检测模型和标签选择算法得到的伪标签在 YOLOv3 检测框架上训练新类，然后在包含 20 个类别的 VOC2007 测试集上测试模型的检测精度。表 2 列出了所提模型在各个类别上的检测结果，平均精度值(mAP)达到 71.9%。其中旧类别目标精度为 71.3%，相比旧模型在旧数据集上的结果(81.6%)，旧类别的检测精度并没有大幅度遗忘，说明标签选择算法较好地缓解了遗忘问题。在该算法基础上，采用本文设计的基于注意力残差模块的特征提取网络提取目标特征，最后模型的检测结果达到72.9%，相比不使用注意力残差模块的方法，平均精度值提升了 1%，同时旧类别的精度值也提升了 1%。为了进一步说明增量方法缓解遗忘的有效性，本文在训练新数据时，使用全部类别的所有标签信息，即包括旧类目标注释的数据集训练网络，然后与本文提出的增量检测方法进行对比，结果如表 2 所列。可以看出，增量检测方法旧类别的检测精度与使用全部标签训练得到的模型检测结果差别甚微，mAP 值仅低 0.2%。

表 2 模型检测新旧类目标精度对比

Table 2 Accuracy comparison of old and new objects detected by the model

(单位：%)

方法	旧类别(1－10) mAP	新类别(11－20) mAP	全部类别 mAP
标签选择算法(不含注意力残差模块)	71.3	72.6	**71.9**
标签选择算法(含注意力残差模块)	72.3	73.5	**72.9**
使用全部标签(含注意力残差模块)	72.5	73.6	**73.1**

经过分析和实验结果的验证表明，通过增量检测方法得到的伪标签训练的模型能够较好地检测所有类别，缓解了旧类目标的遗忘问题，同时相比人工标注或从头训练模型，该方法节约了大量的时间和成本。注意力残差模块在特征提取阶段的不同深度关注目标的重要特征信息，使得整个检测模型受益，结果表明，标签选择算法和注意力残差模块的设计不仅让模型在旧类别上的检测性能得到提高，同时也提高了新类别的检测性能，进而提高了模型对新旧类别的平均检测精度，较好地缓解了增量检测任务中存在的灾难性遗忘问题。

在增量检测实验中，本文提出的基于边框距离度量的

标签选择算法中采用的是 CIoU 度量方式，为了研究不同度量方式对实验结果的影响，在增量检测数据集 VOC(11－20)上进行了增量学习的训练，如表 3 所列。

表 3 基于不同度量方式的标签选择算法对模型精度的影响

Table 3 Influence of label selection algorithms based on different metrics on model accuracy

（单位：%）

度量方式	训练数据集	测试数据集	*mAP*
GIoU	VOC2007(11－20)	VOC2007	71.5
DIoU	VOC2007(11－20)	VOC2007	71.6
CIoU	VOC2007(11－20)	VOC2007	**71.9**

在测试集 VOC2007 上进行了验证，结果表明，采用 CIoU 方式度量得到的伪标签能够更多地去除误检的标签和更多地保留正确的标签信息，原因是 CIoU 不仅将边框的距离、重叠率和尺度都考虑进去，还考虑到了长宽比问题，有效地对边框重叠度进行度量，因此模型的检测效果最好。

将模型的检测结果与最近的方法在同样的测试集(VOC2007)上进行对比，计算每个类别的平均精度(*AP*)值。其中 ILOD 方法[8]最先将压缩模型中“知识蒸馏”思想应用在目标检测任务中，该方法采用双网络模型，通过重新设计损失函数进行增量检测，其 *mAP* 值为 63.1%；RILOD 方法[10]是对 ILOD 方法的改进，其将蒸馏思想同时应用在特征层中，使检测性能得到提高，该方法的 *mAP* 值为 67.9%；DMC 方法[11]通过集成模型，然后利用额外数据来训练新模型的方法，该方法的 *mAP* 值为 68.3%。不同方法的各类别检测结果如表 4 所列。

表 4 不同方法检测的(10＋10)各类别精度

Table 4 Accuracy of each category(10＋10) detected by different methods

（单位：%）

方法	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plants	sheep	sofa	train	tv	*mAP*
ILOD	69.9	70.4	69.4	54.3	48	68.7	78.9	68.4	45.5	58.1	59.7	72.7	73.5	73.2	66.3	29.5	63.4	61.6	69.3	62.2	63.1
RILOD	71.7	81.7	66.9	49.6	58	65.9	84.7	76.8	50.1	69.4	67	72.8	77.3	73.8	74.9	39.9	68.5	61.5	75.5	72.4	67.9
DMC	73.9	81.7	72.7	54.6	59.2	73.7	85.2	83.3	52.9	68.1	62.6	75	69	63.4	80.3	42.4	60.3	61.5	72.6	74.5	68.3
Our	82.1	82.6	61.4	59.6	68.1	86.2	87.4	73.8	53.4	68.2	67.1	77.6	84.3	81.5	83.7	42.9	74.2	68.5	79.4	75.8	**72.9**

为了解决旧类别遗忘问题，上述方法在训练过程中对旧模型进行“知识蒸馏”，迫使新模型不忘记旧类别的信息，需要耗费额外的空间和大量计算时间。相比之下，本文方法不需要额外的数据和空间，通过度量边框距离的方法对生成的伪标签进行筛选，得到用于训练的伪标签集合，较好地改善了新类数据中缺失旧类别目标信息的问题，同时设计的注意力残差模块在特征提取阶段的不同深度都能很好地提取鉴别性特征，达到学而不忘的目的，模型的检测结果都得到了较大的提升。

实验结果表明，本文方法不仅在新类别的检测精度上得到了提升，对旧类别的检测精度也超过了现有方法的最高精度。目前较新的 IncDet 方法[15]基于 EWC(弹性权重巩固)思想进行了改进，使用 Faster RCNN 检测框架在 VOC2007 数据集上达到了 70.8%的 *mAP* 值。该方法的整体精度是在新类别的精度得到较大提升的基础上的结果，该方法在旧类别的检测精度上与本文方法相差较大，因为经过生成和选择后的伪标签大大弥补了旧类别缺失的标注信息，使得模型最大程度地避免了灾难性遗忘，加之注意力残差模块的设计，让网络在训练过程中学习到了不同深度层次的鉴别性特征，使得模型对旧类的检测精度得到进一步的提升。我们将现有的 4 种方法的检测结果分别对旧类、新类以及全部类别上的 *mAP* 值进行计算并与本文方法进行比较，结果如图 6 所示。

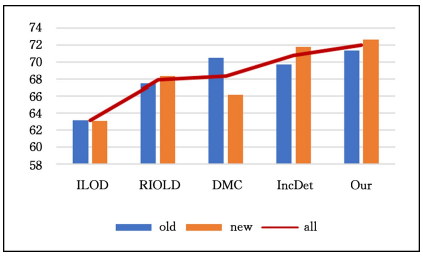


图 6 VOC2007 测试集新旧类平均检测精度

Fig. 6 Average detection accuracy of old and new categories on VOC2007 test set

图 6 表明双模型的蒸馏检测方法(DMC)能较好地缓解遗忘问题，但是会阻碍新类别的学习，导致新类别的检测精度表现较差；RILOD 方法在新旧类别上的学习有一个相对的平衡，但是整体的检测精度值较低；最新的方法(IncDet)对于新类别的学习表现较好，但是对旧类别的检测精度还有待提高；本文方法无论是新旧类别还是全部类别的平均检测精度都优于其他的增量检测方法。同时，相比原有算法，本文的增量检测方法的检测时间基本不变，保证了原有算法的检测速度，如表 5 所列。在进行增量检测时，原算法需要额外的标注和重新训练，而本文方法通过迁移学习，在旧模型权重的基础上训练网络需要的迭代次数仅为原始算法的一半，且模型能够很快收敛，有效缩短了时间并降低了成本。

表 5 算法检测时间对比

Table 5 Comparison of algorithm detection time

算法	输入尺寸	时间/ms
原始算法	416×416	32
本文算法	416×416	33
原始算法	608×608	53
本文算法	608×608	54

结束语 本文提出了一种基于边框距离度量的增量目标检测方法，通过生成一个包含新旧类目标标签信息的伪标签集合用于训练增量检测模型，解决了新数据集中存在旧类别目标时模型将其识别为背景的问题，缓解了遗忘问题。同时在特征提取网络中，本文设计了一种注意力残差模块，加强对目标特征学习和提取能力，在网络的不同深度提取可鉴别性特征信息，并在 3 个不同尺度上进行融合与传播，训练过程无须访问旧类别的标签数据，只使用新类数据集训练模型即可检测所有类别。我们在单阶段目标检测框架上实现了该算法，同时在 PASCAL VOC 数据集上验证了该方法的有效性。与其他检测方法相比，所提检测模型不仅简单且无论是旧类

别还是全部类别的目标检测精度都更高。本文提出的增量方法检测旧目标的精度与使用全部标签进行训练的模型检测结果相差甚小，能有效缓解遗忘问题，同时检测时间基本不变，保证了算法的检测速度。今后我们将探索如何更好地保留旧类目标特征信息的方法，实现即使在以序列形式逐个增量的极端情况下也能很好地检测所有目标。

参 考 文 献

[1] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[2] REN S, HE K, GIRRSHICK 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.

[3] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE Computer Society, 2017: 2117-2125.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-time Object Detection[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 779-788.

[5] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv: 1804. 02767, 2018.

[6] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as Points[J]. arXiv: 1904. 07850, 2019.

[7] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C]// IEEE International Conference on Computer Vision. 2019: 9626-9635.

[8] SHMELKOV K, SCHMID C, ALAHARI K. Incremental Learning of Object Detectors without Catastrophic Forgetting[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 3420-3429.

[9] LI Z, HOIEM D. Learning without Forgetting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(12): 2935-2947.

[10] LI D, TASCI S, GHOSH S, et al. RILOD: Near Real-time Incremental Learning for Object Detection at the Edge[C]// Proceedings of the 4th ACM. New York, 2019: 113-126.

[11] ZHANG J, ZHANG J, GHOSH S, et al. Class-incremental Learning via Deep Model Consolidation[C]// IEEE Winter Conference on Applications of Computer Vision. 2020: 1131-1140.

[12] ZHENG Z, WANG P, REN D, et al. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation[EB/OL]. https://arxiv. org/pdf/2005. 03572. pdf.

[13] DHAR P, SINGH R V, PENG K C, et al. Learning without Memorizing[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5138-5146.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// Proceedings of the European Conference on Computer Vision. 2018: 3-19.

[15] LIU L, KUANG Z, CHEN Y, et al. IncDet: In Defense of Elastic Weight Consolidation for Incremental Object Detection [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(6): 2306-2319.

LIU Dong-mei, born in 1996, postgraduate. Her main research interests include image processing and deep learning.

XU Yang, born in 1990, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include image processing and machine learning.

（责任编辑：何杨）