文章编号　1004-924X(2023)13-2000-08

采用SVD协同训练的半监督实例级目标检测

王　睿[1*]，樊思杨[1]，许婧文[1]，温志庆[2]

(1. 北京航空航天大学 仪器科学与光电工程学院 精密光机电一体化教育部重点实验室，北京 100191；
2. 季华实验室 智能机器人工程研究中心，广东 佛山 528200)

摘要：在室内实例物体目标检测中，传统深度学习需要大量人工标注的训练样本进行网络训练，费时费力，为此提出并实现了一种采用奇异值分解(Singular Value Decomposition，SVD)和协同训练的半监督实例级目标检测网络SVD-RCNN。挑选关键样本进行人工标注并预训练SVD-RCNN，以确保其获取更多先验知识，采用基于SVD的收敛-分解-微调策略，在SVD-RCNN中得到两个较强独立性的检测器以满足协同训练的要求，最后提出一种自适应的自标注策略，获得高质量的自标注及检测结果。在多个室内实例数据集上对该方法进行测试，在GMU数据集上只需人工标注199个样本，均值平均精度(mean Average Precision，mAP)达到了79.3%，相较于需标注3 851个样本的全监督Faster RCNN的81.3% mAP仅下降了2%。消融实验及系列实验证明了本文方法的有效性和普适性，本文提出的方法仅需人工标注5%的训练数据，即可达到与全监督学习相当的实例级目标检测精度，有利于智能机器人高效识别不同实例物体的实际应用。

关　键　词：计算机视觉；目标检测；半监督学习；自标注

中图分类号：TP394.1；TH691.9　　**文献标识码**：A　　**doi**：10.37188/OPE.20233113.2000

Semi-supervised instance object detection method based on SVD co-training

WANG Rui[1*]，FAN Siyang[1]，XU Jingwen[1]，WEN Zhiqing[2]

(1. *Laboratory of Precision Opto-mechatronics Technology*，*Ministry of Education*，*Institute of Instrumentation Science and Opto-electronics Engineering*，*Beihang University*，*Beijing* 100191，*China*；
2. *Engineering Research Center for Intelligent Robotics*，*Ji Hua Laboratory*，*Foshan* 528200，*China*)
* *Corresponding author*，*E-mail*：*wangr@buaa.edu.cn*

Abstract：Detecting indoor instance objects is useful for various applications. Traditional deep-learning methods require a large number of labeled samples for network training，making them time-consuming and labor-intensive. To address this problem，SVD-RCNN—a semi-supervised instance object detection network based on singular value decomposition (SVD) and co-training—is proposed. First，key samples are selected for manual labeling to pre-train SVD-RCNN，to ensure that it acquires more prior knowledge. Second，a convergence，decomposition，and finetuning strategy based on SVD is used to obtain two detectors with strong independence in SVD-RCNN to satisfy the requirements of co-training. Finally，an adap-

收稿日期：2022-07-27；**修订日期**：2022-09-11.
基金项目：国家自然科学基金资助项目(No. 61673039)；佛山市产业领域科技攻关专项(No. 2020001006807)

tive self-labeling strategy is used to obtain high-quality self-labeling and detection results. The method was tested on multiple indoor instance datasets. On the GMU dataset, it achieved a mean average precision of 79.3% with 199 manually labeled samples. This was only 2% lower than that (81.3%) of Faster RCNN with fully supervised learning, which required labeling 3 851 samples. Ablation studies and a series of experiments confirmed the effectiveness and universality of the method. The results indicated that the method only needs to manually label 5% of the training data to achieve instance-level detection accuracy comparable to that of fully supervised learning; thus, it is suitable for applications in which intelligent robots must efficiently identify different instance objects.

Key words: computer vision; object detection; semi-supervised learning; self-labeling

1　引　言

智能机器人技术已逐渐应用于室内场景以完成智能家居、助老助残等任务，这些任务中智能机器人要能够在结构复杂的不同室内场景中识别检测到特定的实例物体，有别于常见的类别级目标检测技术，实例级目标检测[1]通过提取和分析目标的特征，对特定的实例目标进行分类和定位。

近年来，学者们将卷积神经网络(Convolutional Neural Network, CNN)应用于类别级目标检测[2]，通常能够获得较为稳定、准确的检测效果。由于任务的相似性，通过迁移学习将相应的CNN改进后可以应用于实例级目标检测，但CNN通常需要10^3及10^4以上数量级的带标签训练数据进行有监督训练，而实例级目标检测任务中有时每个实例只有少量带标签的训练样本可用，相较于类别级样本，人工标注实例样本更为繁琐耗时，从而限制了卷积神经网络在实例级物体检测领域的发展与应用。因此，利用易获取的未标注样本训练模型的想法应运而生，其中，通过学习少量有标注样本和大量未标注样本而获得高精度模型的半监督学习方法[3-4]倍受关注。

当前主流的半监督目标检测方法可以分为两类：基于一致性的方法[5]和基于伪标签的方法[6-8]。基于一致性的方法[5]采用对未标注样本进行数据增强和一致性约束提高模型的性能，但是相对于全监督来说，检测精度下降较大。为了能够提升半监督目标检测算法的检测精度，Sohn等[6]提出了基于伪标签的STAC算法，利用教师模型输出的较高精度的伪标签训练学生模型。在此基础上，Liu等[7]提出了采用Focal loss的Unbiased Teacher算法，Guo等[8]提出了解决物体尺度差异大的问题SED框架，Li等[9]则是实现了能够进行带噪伪标签学习的PseCo。但是上述半监督目标检测方法主要是面向类别级目标检测，面向实例级目标检测的半监督方法目前几乎没有。

借鉴面向类别级目标检测的半监督方法，本文提出了一种基于伪标签的半监督实例级目标检测方法。该方法使用协同训练[10]的方式训练目标检测网络，包括关键样本选择策略、基于奇异值分解(Singular Value Decomposition, SVD[11])的收敛-分解-微调训练策略和自适应的自标注策略，确保协同训练的稳健进行，仅需人工标注5%的训练样本即可达到与全监督学习相当的实例级检测精度。

2　原　理

协同训练[10]在半监督学习中的应用范围较为广泛，它通过两个学习器的独立性来互相标记样本扩充训练集，从而利用未标注样本提升学习器的性能。

SVD[11]提出对于任意一个$m\times n$维的矩阵$\boldsymbol{W}$可以分解为：

$$\boldsymbol{W}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}=\boldsymbol{U}\begin{bmatrix}\sigma_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma_m\end{bmatrix}\boldsymbol{V}^{\mathrm{T}},\tag{1}$$

其中：$S\in\mathbf{R}^{m\times n}$是$m\times n$维的对角矩阵，其对角线元素为$\boldsymbol{W}$从大到小排序的非负奇异值$\sigma_1,\cdots,\sigma_m$；$\boldsymbol{U}\in\mathbf{R}^{m\times m}$，$\boldsymbol{V}\in\mathbf{R}^{n\times n}$均为正交矩阵，矩阵中的向量两两正交，相互独立。因此，本文通过SVD来满足协同训练中对学习器独立性的要求。

本文以经典的 Faster RCNN[12]作为基础网络，提出了 SVD-RCNN 检测网络，如图 1 所示。与原网络不同，SVD-RCNN 将最后的全连接层替换为参数矩阵为正交矩阵的 SVD 层，并将其输出的特征向量分为两组，作为特征向量 1 和特征向量 2，分别输入两个检测器中完成协同训练。

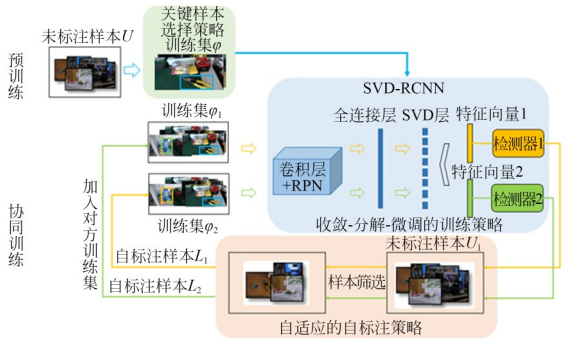


图 1　SVD-RCNN 训练的流程

Fig. 1　Flow chart of training SVD-RCNN

SVD 协同训练分为预训练阶段和协同训练阶段。首先，通过预训练阶段获得较为可靠的 SVD-RCNN，再将它迁移至协同训练阶段继续训练，使其最终获得较高的检测精度。

算法 1：SVD 协同训练流程如下：

输入：未标注样本 U

预训练：

(1)从未标注样本 U 中挑选关键样本进行人工标注作为训练集 ϕ，U 中剩余样本作为未标注样本 U_1；

(2)使用 ϕ 对 SVD-RCNN 采用基于 SVD 的收敛－分解－微调策略训练至收敛；

协同训练：

(3)将预训练后的 SVD-RCNN 迁移至协同训练阶段，将 ϕ 中的样本信息赋值给训练集 ϕ_1 和 ϕ_2；

(4)repeat；

(5)使用 SVD-RCNN 对 U_1 进行自标注，得到自标注样本 L_1 和 L_2 并加入对方训练集，$\phi_1=\phi_1\cup L_2$，$\phi_2=\phi_2\cup L_1$；

(6)使用 ϕ_1 和 ϕ_2 对 SVD-RCNN 采用收敛-分解-微调的策略训练至收敛，更新 SVD-RCNN；

(7)until 无法从 U_1 中得到 L_1 和 L_2；

输出：高检测精度的 SVD－RCNN 和数据集的标注结果。

根据以上流程，本文提出了关键样本选择策略、收敛-分解-微调的训练策略和自适应的自标注策略以确保协同训练稳健进行。

2.1　关键样本选择

为有效实现协同训练，本文首先对 SVD-RCNN 使用关键样本进行预训练，确保 SVD-RCNN 能够获取更多先验知识且更具泛化能力，从而为 SVD-RCNN 的性能提升打下基础，为此提出了关键样本选择策略，该策略能挑选出信息量丰富的代表性样本，其框架如图 2 所示。

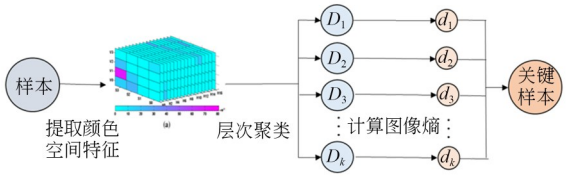


图 2　关键样本选择的框架

Fig. 2　Framework of key sample selection

本文采用差异性准则和不确定性准则选择对 SVD-RCNN 最具信息量的关键样本，实现关键样本信息量和信息冗余之间的平衡。如图 2 所示，根据差异性准则，首先使用聚类[13]将未标注样本分类。对于一张 RGB 图像，联合使用颜色特征和空间特征：将图像分为 4 个子块，提取每个子块 R，G，B 三通道的颜色直方图作为颜色特征，然后拼接四个子块的颜色特征，得到颜色空间特征。以轮廓系数为指标，采用层次聚类将未标注样本分为 K 类，每一类中分别包含 D_1，…，D_K 张图像。根据不确定性准则，本文选用图像的信息熵[14]来衡量样本的不确定度，其计算公式为：

$$E=-\sum_{i=0}^{255}p_i\log p_i, \tag{2}$$

其中：p_i 表示图像中灰度值为 i 的像素所占的比例，熵越大说明样本的不确定度越大，所含信息量越高。选择每一类中熵最大的前几张样本图片进行人工标注，如按熵值大小在第 1 类中选择较大的 d_1 个样本进行人工标注，以此类推，K 类

样本共可产生出 $d_1+\cdots+d_k$ 张样本图片（训练集 ϕ），称为关键样本。

2.2　基于SVD的收敛-分解-微调的训练策略

对于图1的SVD-RCNN来说，协同训练成功的关键在于检测器1和检测器2之间的独立性，因此本文提出一种收敛-分解-微调的策略。该策略训练得到的SVD-RCNN中包含上千个互不相关的特征向量，分成两组分别输入两个检测器可以获取独立的检测器，协同工作实现半监督学习。

如图1所示，SVD-RCNN中最后全连接层的参数矩阵可以看作由其中列向量组成的线性空间，其输出特征是输入特征经过该线性空间的投影，因此只需将最后全连接层的投影方向变成正交投影方向，即可使其输出的特征向量1和2相互正交从而去相关。单轮训练SVD-RCNN的具体操作为：

（1）收敛。在预训练阶段使用训练集 ϕ 或者在协同训练阶段使用训练集 ϕ_1 和 ϕ_2 训练SVD-RCNN至收敛，使其提取的特征具有一定的表达能力。

（2）分解。将SVD-RCNN中最后全连接层的参数矩阵 $\boldsymbol{W}$ 进行SVD分解（$\boldsymbol{W}=\boldsymbol{USV}^{\mathrm{T}}$[11]），使用 $\boldsymbol{US}$ 替换 $\boldsymbol{W}$ 得到SVD层。

（3）微调。将SVD层的参数矩阵固定，继续训练SVD-RCNN至收敛。

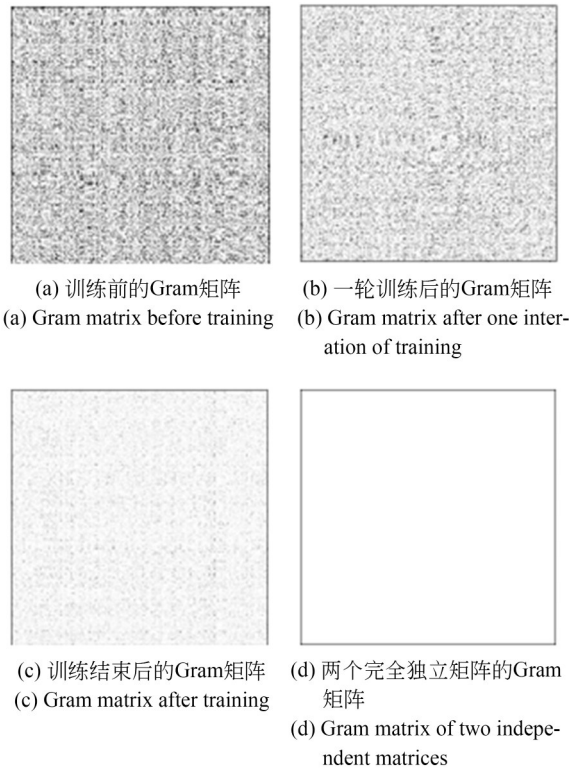
如算法1所示，上述单轮操作方法是整个半监督学习的一个环节，需要重复多次，且要不断更新扩充标记样本，直至所有的无标记样本被充分利用时，才可以获取一个高检测精度的SVD-RCNN。

为证明收敛-分解-微调策略能够增强检测器1和2之间的独立性，构建检测器的参数矩阵 $\boldsymbol{W}_k\in\mathbf{R}^{m\times n}(k\in\{1,2\})$ 的格莱姆Gram矩阵[15]，如下：

$$\boldsymbol{G}=\boldsymbol{W}_1{}^{\mathrm{T}}\boldsymbol{W}_2=\begin{bmatrix}\vec{w}_{11}^{\mathrm{T}}\vec{w}_{21} & \cdots & \vec{w}_{11}^{\mathrm{T}}\vec{w}_{2n}\\ \vdots & \ddots & \vdots\\ \vec{w}_{1n}^{\mathrm{T}}\vec{w}_{21} & \cdots & \vec{w}_{1n}^{\mathrm{T}}\vec{w}_{2n}\end{bmatrix}=\begin{bmatrix}g_{11} & \cdots & g_{1n}\\ \vdots & \ddots & \vdots\\ g_{n1} & \cdots & g_{nn}\end{bmatrix},\tag{3}$$

其中：$\vec{w}_{ki}\in\mathbf{R}^{m\times 1}$ 为 $\boldsymbol{W}_k$ 的第 i 列向量，g_{ij} 表示 $\boldsymbol{W}_1$ 中第 i 列向量和 $\boldsymbol{W}_2$ 中第 j 列向量的内积。此时Gram矩阵中的每个元素 g_{ij} 实际上就代表着 $\boldsymbol{W}_1$ 的 i 通道与 $\boldsymbol{W}_2$ 的 j 通道的特征向量 $\vec{w}_{1i}$，$\vec{w}_{2j}$ 之间的相关程度，g_{ij} 越小，表示相关性越低，g_{ij} 值为0则表示参与内积的两个向量正交。利用Gram矩阵对训练前后的特征向量（$\vec{w}_{1i}$，$\vec{w}_{2j}$）进行相关性评估，以验证SVD-RCNN的有效性，如果 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 之间任意两向量正交，则说明 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 完全相互独立。

图3将Gram矩阵中 g_{ij} 用灰度值进行了可视化表示，g_{ij} 越小，其灰度值越大。图3(a)～3(c)反映了使用收敛-分解-微调策略进行不同迭代次数时Gram矩阵中各 g_{ij} 值的分布情况，图3(d)为仿真生成的两个完全独立矩阵的Gram矩阵，即当 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 之间任意两向量正交时，Gram矩阵表示为全白。图3表明：随着协同训练迭代的递增，两个检测器之间的独立性逐渐增强，且能很好地逼近完全独立的状态。



(a) 训练前的Gram矩阵
(a) Gram matrix before training

(b) 一轮训练后的Gram矩阵
(b) Gram matrix after one interation of training

(c) 训练结束后的Gram矩阵
(c) Gram matrix after training

(d) 两个完全独立矩阵的Gram矩阵
(d) Gram matrix of two independent matrices

图3　收敛-分解-微调前后Gram矩阵的对比

Fig. 3　Comparison of Gram matrix before and after convergence, decomposition and finetuning

2.3　自适应的自标注策略

样本在协同更新时，如何判断检测器的自主标注是否可信非常重要，对于目标检测任务来说，高置信度的样本应该包含高准确率和高召回率两个指标。因此，本文提出了一种自适应的自标注策略，综合准确率与召回率对自标注结果进行打分，依据分数进行样本的筛选与更新，以确保将高置信度的自标注样本添加到训练集 ϕ_1 和 ϕ_2 中，提升后续训练中 SVD-RCNN 的精度。对于一张待检测的图像 x，其得分规则如下：

$$score(x)=\sum_{i=1}^{n} I(accuracy_1(w_i)>\delta_{acc})\\ [I(accuracy_2(w_i)>\delta_{acc})\cdot I(IoU(wi)>\delta_{IoU})], \tag{4}$$

其中：$I(\cdot)$为指示函数，在·为真和假时分别取值为 1，0，w_i 是检测器 1 和检测器 2 都检测出的目标，$IoU(w_i)$是目标 w_i 在两个检测器中所属预测框的交并比，根据文献[16]，δ_{IoU} 设置为 0.5。$accuracy_1$ 和 $accuracy_2$ 分别是两个检测器对目标 w_i 所属预测框的置信度，本文在 GMU 数据库上研究了 δ_{acc} 对训练 SVD-RCNN 的影响。实验结果表明，将 δ_{acc} 设置为 0.8，能够较好地平衡自标注结果中的准确率和召回率，使得 SVD-RCNN 的性能最好。

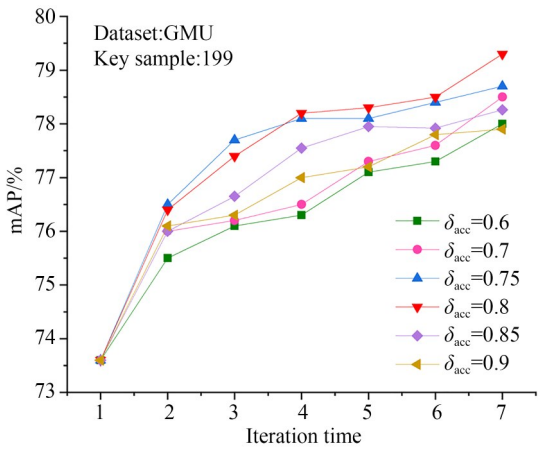


图 4　不同 δ_{acc} 下 SVD-RCNN 的迭代精度

Fig. 4　Iteration accuracy of SVD-RCNN with different δ_{acc}

式(4)中，accuracy 和 IoU 考虑了预测框的准确率，计算满足要求的预测框的个数则是衡量预测框的召回率，因此在该规则下得到的高得分自标注结果可以被认为是既精确又全面的。

其次，设置阈值 τ_k 以筛选样本并更新训练集，为平衡自标注样本的质量与数量，使用一种自适应的方式设置阈值 τ_k，即：

$$\tau_k=\frac{\beta}{N_k}\sum_{i=1}^{N_k} score(x_i^k), \tag{5}$$

其中：N_k 是第 K 类中包含的图像数量，β 是调节每轮自标注样本数量的参数，在协同训练的第一轮设置为 β_0，随着迭代次数的增加，β 值每轮递减。在实验中，设置每轮递减 6.5%β_0 能够取得较好的精度。这是由于初始迭代时 SVD-RCNN 的精度较低，需设置较高的阈值避免将置信度低的自标注样本添加到训练集中造成噪声累积；随着迭代的进行，SVD-RCNN 能够进行较高质量的自标注，因此需降低阈值以选出更多数量的自标注样本，防止 SVD-RCNN 过拟合。

3　实验结果与分析

3.1　实验数据集

为验证 SVD-RCNN 的性能，分别在 GMU[17]，AVD[18] 公共数据集和 BHID 自制数据集上进行了系列实验。GMU 数据集为厨房场景下的实例级目标检测数据集，包含 11 个实例对象，总共有 6 728 张图像，其中 3 851 张图像用于训练，2 877 张图像用于测试；AVD 数据集为家庭场景下的实例级目标检测数据集，包含 30 个实例对象，本文从中挑选出厨房场景下的 6 个实例对象，共计 7 262 张图像，其中 5 083 张图片用于训练，2 179 张用于测试；BHID 数据集为在办公场景下的自制实例级数据集，包含 7 个实例对象，共计 1 721 张图像，其中 1 110 张图片用于训练，611 张用于测试。

3.2　评价指标

实验采用均值平均精度（mean Average Precision，mAP）作为 SVD-RCNN 的目标检测性能的评价指标。mAP 是各实例目标类的平均精度（Average Precision，AP）的均值，AP 是准确率-召回率（Precision-Recall）曲线下的面积，因此 mAP 是从准确率和召回率两个角度对 SVD-RCNN 检测不同实例物体的性能进行综合评价。

3.3　实　验

实验分为两个部分，第一部分在 GMU 数据

集上探究了关键样本选择策略、收敛-分解-微调策略和自适应的自标注策略对SVD-RCNN性能的影响，以证明各个策略的有效性；第二部分在GMU，AVD和BHID数据集上将本文方法与全监督学习和其他半监督学习进行比较，以证明本文方法的有效性与普适性。

第一部分实验首先探讨了不同数量的关键样本对SVD-RCNN性能的影响。按照2.1所述的策略，从GMU数据集的训练集中分别选择2%（82张），5%（199张）和10%（396张）的图像作为关键样本对SVD-RCNN进行训练，检测精度分别达到了73.8% mAP，79.3% mAP和79.3% mAP，说明使用5%的训练样本作为关键样本足以表示训练集的整体分布，且可以保证SVD-RCNN的性能可持续提升。其次，探究了关键样本选择策略对预训练SVD-RCNN泛化能力的影响，在GMU的训练集中分别使用随机选择的方法和关键样本选择的方法获取5%的图像作为初始样本对SVD-RCNN进行训练，检测精度分别达到了66.5% mAP和73.6% mAP。图5列举了预训练SVD-RCNN对6类实例物体的检测精度。

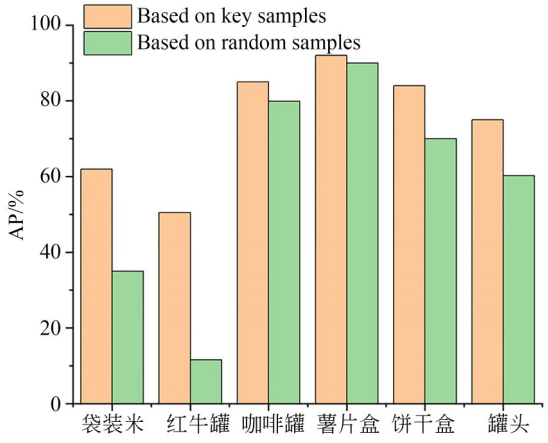


图5　预训练SVD-RCNN在GMU数据集上的检测精度

Fig. 5　Detection accuracy of pre-training SVD-RCNN on GMU dataset

在预训练SVD-RCNN时，人工标注关键样本比随机标注样本在单个实例物体的检测精度上有一定程度的提高，说明关键样本选择策略可以保证预训练时人工标注的样本是最具信息量的关键样本，从而令SVD-RCNN预训练后获取更多的先验知识并具有较好的泛化能力。

为探究收敛-分解-微调策略和自适应的自标注策略的有效性，在GMU的训练集中挑选5%（199张）的图像进行消融实验，结果如表1所示。

表1　消融实验结果

Tab. 1　Results of ablation experiment

数据集	关键样本数量	收敛-分解-微调	自适应的自标注	mAP/%
GMU	199			72.1
	199	√		77.6
	199		√	75.8
	199	√	√	79.3

从表1可以看出，单独采用收敛-分解-微调的训练策略提升了5.5%的检测精度，说明相比于两个相关性较强的检测器，独立性强的检测器能增强互相检查与纠错的能力。单独实施自适应的自标注策略提升了3.7%的检测精度，说明在协同训练时，如果将置信度低的自标注结果添加到训练集中会造成噪声累积，从而对最终迭代的SVD-RCNN产生负面影响。将两者结合使用，检测精度提升了7.2%，由此证明了收敛-分解-微调策略和自适应的自标注策略的有效性。

为验证本文方法的有效性和普适性，在GMU，AVD和BHID数据集上与全监督学习的Faster RCNN进行了对比实验。分别从3个数据集的训练集中选择5%的图像作为关键样本进行人工标注，并使用本文方法训练SVD-RCNN，在GMU数据集上的检测精度达到了79.3% mAP，

表2　训练方法在不同数据集上检测精度的对比

Tab. 2　Comparison of detection accuracy of training methods on different datasets

数据集	关键样本数量	方　法	mAP/%
GMU	199	半监督SVD-RCNN	79.3
	3 851	全监督Faster RCNN	81.3
AVD	255	半监督SVD-RCNN	75.6
	5 083	全监督Faster RCNN	78.5
BHID	55	半监督SVD-RCNN	76.1
	1 110	全监督Faster RCNN	78.4

相比于全监督学习的 Faster RCNN 的 81.3%，mAP 仅下降 2%，说明本文方法在不同数据集上仅需人工标注 5% 的训练样本，就可达到与全监督相当的目标检测精度。

SVD-RCNN 在不同数据集上对部分实例物体的检测结果如图 6 所示。可以看出，SVD-RCNN 在结构复杂的环境中对每个数据集均可达到很好的检测效果，甚至在遮挡，物体不完整的情况下仍能找到物体，如图 6(b)中的第一幅图所示。这说明 SVD-RCNN 可以在结构复杂的不同室内场景中胜任特定实例物体的检测工作。

在 GMU 数据集上，将本文算法与采用学生教师模型的两个性能较好的半监督方法进行了对比，实验结果如表 3 所示。在使用相同标注样本数量的情况下，采用协同训练的本文算法的检测精度比 Unbiased Teacher[7] 提升了 2.8%，比 Soft Teacher[19] 提升了 1.0%。

图 6　在 GMU，AVD 和 BHID 数据集上的检测结果

Fig. 6　Detection result on GMU, AVD and BHID datasets

表 3　在 GMU 数据集上半监督算法的对比

Tab. 3　Comparison of semi-supervised methods on GMU

数据集（标注样本数量）	基础网络	方　法	mAP/%
GMU (199)	Faster RCNN	Unbiased Teacher[7]	76.5
		Soft Teacher[19]	78.3
		本文算法	79.3

4　结　论

本文提出了一种采用 SVD 和协同训练的半监督实例级目标检测网络 SVD-RCNN。为保证 SVD-RCNN 具有较好的实例目标检测先验知识，只需人工标注少量关键样本对其进行预训练，并将预训练好的 SVD-RCNN 迁移到协同训练阶段；利用基于 SVD 的收敛-分解-微调策略，在保证网络收敛的前提下，训练出两个独立不相关的检测器，以满足协同训练对于相互独立视图的要求，最后采用一种自适应的自标注策略获得高质量的实例级目标的自标注和检测结果。实验结果表明，在人工标注 5% 的训练样本的条件下，本文方法可以达到与全监督学习相当的实例级目标检测精度，大大节约了人工标注的开销，有利于智能机器人高效识别不同实例物体。

参考文献：

[1] GRAUMAN K, LEIBE B. Visual object recognition[J]. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2011, 5(2): 1-181.

[2] 范丽丽，赵宏伟，赵浩宇，等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学 精密工程，2020，28(5)：1152-1164.
FAN L L, ZHAO H W, ZHAO H Y, *et al*. Survey of target detection based on deep convolutional neural networks[J]. *Opt. Precision Eng.*, 2020, 28(5): 1152-1164. (in Chinese)

[3] 徐哲，耿杰，蒋雯，等. 联合训练生成对抗网络的半监督分类方法[J]. 光学 精密工程，2021，29(5)：1127-1135.
XU ZH, GENG J, JIANG W, *et al*. Co-training generative adversarial networks for semi-supervised classification method [J]. *Opt. Precision Eng.*, 2021, 29(5): 1127-1135. (in Chinese)

[4] 黄鸿，唐玉枭，段宇乐. 半监督多图嵌入的高光谱影像特征提取[J]. 光学 精密工程，2020，28(2)：443-456.
HUANG H, TANG Y X, DUAN Y L. Feature extraction of hyperspectral image with semi-supervised multi-graph embedding [J]. *Opt. Precision Eng.*, 2020, 28(2): 443-456. (in Chinese)

[5] JEONG J, VERMA V, HYUN M, *et al*. Interpo-

lation-based semi-supervised learning for object detection[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 20-25, 2021, *Nashville*, *TN*, *USA*. IEEE, 2021: 11597-11606.

[6] SOHN K, ZHANG Z, LI C, *et al*. *A Simple Semi-supervised Learning Framework for Object Detection*[EB/OL]. 2020: *arXiv*: 2005.04757. https://arxiv.org/abs/2005.04757.

[7] LIU Y C, MA C Y, HE Z, *et al*. *Unbiased Teacher for Semi-supervised Object Detection*[EB/OL]. 2021: *arXiv*: 2102.09480. https://arxiv.org/abs/2102.09480.

[8] GUO Q S, MU Y, CHEN J Y, *et al*. Scale-equivalent distillation for semi-supervised object detection [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 18-24, 2022, *New Orleans*, *LA*, *USA*. IEEE, 2022: 14502-14511.

[9] LI G, LI X, WANG Y J, *et al*. *PseCo: Pseudo Labeling and Consistency Training for Semi-supervised Object Detection*[M]. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 457-472.

[10] NING X, WANG X R, XU S H, *et al*. A review of research on co-training[J]. *Concurrency and Computation: Practice and Experience*, 2021: doi: 10.1002/cpe.6276.

[11] DAN K. A singularly valuable decomposition: the SVD of a matrix[J]. *The College Mathematics Journal*, 1996, 27(1): 2-23.

[12] REN S Q, 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 and Machine Intelligence*, 2017, 39(6): 1137-1149.

[13] RUI W, YING L. Real-time 3D object detection in unstructured environments[C]. 2017 *First International Conference on Electronics Instrumentation & Information Systems* (*EIIS*). *June* 3-5, 2017, *Harbin*, *China*. IEEE, 2018: 1-6.

[14] VERDU S. Fifty years of Shannon theory[J]. *IEEE Transactions on Information Theory*, 1998, 44(6): 2057-2078.

[15] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 2414-2423.

[16] EVERINGHAM M, GOOL L, WILLIAMS C K, *et al*. The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.

[17] GEORGAKIS G, ALIMOOR REZA M, MOUSAVIAN A, *et al*. Multiview RGB-D dataset for object instance detection[C]. 2016 *Fourth International Conference on 3D Vision* (*3DV*). *October* 25-28, 2016, *Stanford*, *CA*, *USA*. IEEE, 2016: 426-434.

[18] AMMIRATO P, POIRSON P, PARK E, *et al*. A dataset for developing and benchmarking active vision[C]. 2017 *IEEE International Conference on Robotics and Automation* (*ICRA*). *May* 29 - *June* 3, 2017, *Singapore*. IEEE, 2017: 1378-1385.

[19] XU M D, ZHANG Z, HU H, *et al*. End-to-end semi-supervised object detection with soft teacher [C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 10-17, 2021, *Montreal*, *QC*, *Canada*. IEEE, 2022: 3040-3049.

作者简介：

王　睿(1965－)，女，北京人，博士，副教授，1987年、1990年于清华大学分别获得学士和硕士学位，2006年于北京航空航天大学获得博士学位，研究方向为信号测试与处理、图像处理、机器学习等。E-mail: wangr@buaa.edu.cn

温志庆(1964－)，男，博士，正高级工程师，1994年于清华大学获得博士学位，研究方向为智能机器视觉、模式识别、3D成像、光学工程、人工智能等。E-mail: wenzq@jihualab.com