计算机科学
COMPUTER SCIENCE

面向智能视频监控的人体小目标检测

杨溢，申昇，窦知阳，李元，韩振军

引用本文

杨溢，申昇，窦知阳，李元，韩振军. 面向智能视频监控的人体小目标检测[J]. 计算机科学，2023，50(9): 75-81.

YANG Yi, SHEN Sheng, DOU Zhiyang, LI Yuan, HAN Zhenjun. Tiny Person Detection for Intelligent Video Surveillance [J]. Computer Science, 2023, 50(9): 75-81.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230400204

面向智能视频监控的人体小目标检测

杨 溢[1] 申 昇[2] 窦知阳[3] 李 元[1] 韩振军[1]
1 中国科学院大学电子电气与通信工程学院 北京 101408
2 北京控制与电子技术研究所 北京 100045
3 吉林大学通信工程学院 长春 130012
(yysolon@163.com)

摘 要 人体目标检测对社会治理和城市安全具有很重要的现实意义,监控数据是数据安全的重要来源。小目标检测是目前受到广泛关注的安全检测问题中一项具有挑战性的任务,其检测对象为大型图像中少于20个像素的目标。小目标的特征难以表征,其中一个主要挑战是,用于预训练/共同训练检测器的数据集(如COCO)与用于微调检测器的数据集(如TinyPerson)之间存在尺度不匹配的情况,这给小目标检测器的性能带来了负面影响。为了解决这个问题,文中提出了一种优化策略,用于匹配不同数据集的尺度,称其为尺度分布搜索(Scale Distribution Search,SDS),同时平衡图片的信息收益(数据集之间的尺度相近)和信息损失(信噪比(SNR)的降低)。该策略使用高斯模型对数据集中目标的尺度分布进行建模,通过迭代的方式寻找最优分布参数;并对比数据集中目标的特征分布和检测器的性能,以找到最佳的尺度分布。通过SDS策略,主流目标检测方法在TinyPerson上实现了更好的性能,证明了SDS策略在提升预训练/共同训练效率上的有效性。

关键词:智能视频监控;小目标检测;尺度搜索;预训练
中图法分类号 TP391.41

Tiny Person Detection for Intelligent Video Surveillance

YANG Yi[1],SHEN Sheng[2],DOU Zhiyang[3],LI Yuan[1] and HAN Zhenjun[1]
1 School of Electronic,Electrical and Communication,University of Chinese Academy of Sciences,Beijing 101408,China
2 Beijing Institute of Control and Electronics Technology,Beijing 100045,China
3 School of Communication Engineering,Jilin University,Changchun 130012,China

Abstract Person detection has significant practical implications for social governance and urban security. Monitoring data is an important source of data security. Tiny object detection,which focuses on less than 20 pixels objects in large-scale images,is a challenging task. One of the main challenges is the scale mismatch between the dataset used for pre-training/co-training the detectors,such as COCO,and the dataset used for fine-tuning the detectors,such as TinyPerson,which negatively affects the performance of detectors on tiny object detection. To address this challenge,this paper proposes an optimization strategy called scale distribution searching(SDS) to match the scale of different datasets for tiny object detection,which also balance the information gain and loss. The Gauss model is used to model the scale distribution of targets in the dataset,and the optimal distribution parameters are found through iteration. The feature distribution and the performance of the detector is comparedto find the best scale distribution. Through the SDS strategy,mainstream object detection methods have achieved better performance on TinyPerson,demonstrating the effectiveness of the SDS strategy in improving pre-training/co-training efficiency.

Keywords Intelligent video surveillance,Tiny object detection,Scale distribution search,Pre-train

1 引言

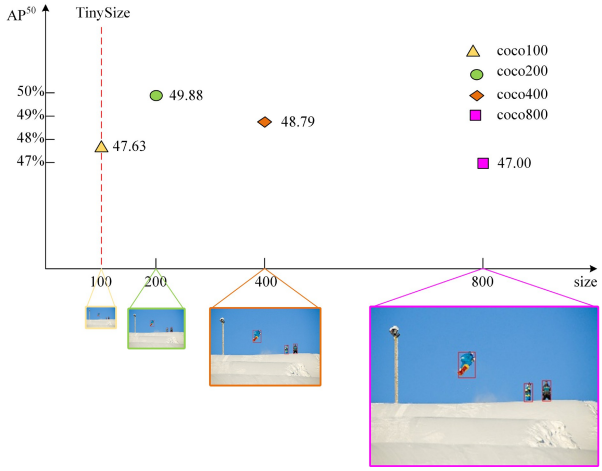
人体目标检测作为计算机视觉研究的核心课题之一,具有广泛的实践应用,这些应用跨越了智能监控、驾驶辅助系统、移动机器人以及应急救援等多个领域。对于社会治理和城市安全来说,人体目标检测技术的重要性不言而喻。

随着深度卷积神经网络的发展和进步,人体目标检测技术,特别是行人检测技术取得了显著的突破。然而,对于小尺度的人体目标检测,例如目标尺寸小于20个像素的情况,其检测问题尚未得到有效的解决。这仍然是我们面临的重要挑战,需要对其进行进一步的研究。

相比合适尺度的物体,小物体由于其尺寸较小以及信噪比较低,其检测任务具有更大挑战性。这两个因素导致小物体特征表征十分困难,进一步影响了先进物体检测器的性能。针对人体小目标检测,本文提出了一种称为尺度分布搜索(SDS)的简洁且有效的方法。如图1所示,使用原始MS CO-

到稿日期:2023-04-27 返修日期:2023-07-11
通信作者:李元(liyuan23@ucas.ac.cn)

CO[1]数据集预训练的检测器[2]性能不佳，但改变 MS COCO 图像尺寸后检测器[2]的性能出现波动，在 COCO400 设定下，性能有所提升。基于此，本文提出了 SDS 方法，其核心思想是寻找最佳尺度分布的预训练/联合训练数据集，以便与微调数据集的大小匹配。具体而言，我们考虑使用高斯模型对数据集中目标的尺度分布进行建模，通过迭代的方式寻找最优分布参数。



注：COCO400 为原尺寸，COCO800 是对其 2 倍上采样版本，COCO200 和 COCO100 则分别是其 0.5 倍和 0.25 倍下采样版本。其中，TinyPerson 的尺寸与 COCO100 类似。

图 1 不同 COCO 预训练尺寸的 Faster R-CNN[3]在 TinyPerson 数据集上的性能

Fig. 1 Performance of Faster R-CNN[3] with different COCO pre-training sizes on TinyPerson

匹配不同数据集的大小可以缩小数据集中图片之间的差距，从而使卷积神经网络(CNN)更好地提取不同大小的目标的特征。因为在下采样的过程中，特征图的分辨率降低，导致图像的信噪比变差，从而削弱了卷积神经网络的能力；而 SDS 的本质在于它可以更好地探索和利用尺度信息来平衡信息的增益和损失。本文的主要贡献包括：

1)提出了一种简单但有效的方法，即尺度分布搜索(SDS)，通过研究和利用比例信息，来实现更好的预训练。SDS 给出了一种分析卷积神经网络预训练策略的新视角。

2)将本文提出的 SDS 方法与主流目标检测方法(如 Faster-RCNN)结合之后，在 TinyPerson 上实现了更好的性能。与众多目标检测方法相比，该性能具有较强的竞争力，证明了 SDS 策略在提升预训练/共同训练效率上的有效性。

2 相关工作

1)基于 CNN 的检测。近年来，随着卷积神经网络(CNN)的发展，一些经典数据集(如 ImageNet [4]，Pascal [5]，MS COCO [1])上的分类、检测和分割性能已远远超过传统机器学习算法。CNN 起源于 LeNet [6]，并随着 AlexNet [7]的流行而变得流行。之后，NIN [8]，VGGNet [9]，GoogLeNet [10]和 ResNet [11]等模型被提出，推动了图像识别的发展。OverFeat [12]和区域卷积神经网络(RCNN)[13]成为了流行的检测架构。OverFeat 采用 Conv-Net 作为滑动窗口检测器在图像金字塔上进行检测。R-CNN 采用基于选择性搜索的区域提议方法，使用 Conv-Net 对尺度归一化的提议进行分类。空间金字塔池化(SPP)[14]在单个图像尺度上提取的特征图上采用了 R-CNN，证明了这种基于区域的检测器可以更有效地被应用。Fast R-CNN[15]和 Faster R-CNN [3]以多任务的方式创建了一个统一的物体检测器。EfficientDet-D7[16]采用更深的网络结构、更大的输入分辨率和更多的特征金字塔级别，实现了在目标检测任务中更准确和更细粒度的目标定位和分类能力。

基于区域的方法复杂且耗时，因此单阶段检测器(如 YOLO[17]和 SSD[18])被提出，用于加快处理速度，但这会导致检测性能下降，尤其是在小目标任务上。此后，性能更好的单阶段检测器被提出，即 Retinanet [19]和 FreeAnchor [20]。Retinanet 将类别不平衡和 Focal loss 用于检测，这对于小目标检测也非常关键，因为小目标的尺度较小。FreeAnchor 通过将检测器训练制定为最大似然估计(MLE)过程，将手工锚定分配改进为“自由”锚定匹配。FreeAnchor 的目标是让网络学习最能解释对象类别的特征，包括分类和定位。

为了突破预分配锚定所带来的限制，最近的无锚点方法采用像素级监督和中心性边界框回归。CornerNet[21]和 CenterNet [22]用关键点监督代替边界框监督。MetaAnchor[23]方法使用子网络从任意自定义的先前框中学习产生锚点。GuidedAnchoring [24]利用语义特征来指导锚点的预测，同时用预测的锚点替换密集锚点。FCOS[25]在训练过程中完全避免了与锚定框相关的复杂计算，如计算重叠度。更重要的是，FCOS 还避免了与锚定框相关的所有超参数，最终检测性能通常对这些超参数非常敏感。在文献[26]中，FoveaBox 通过预测针对类别敏感的语义地图来直接学习对象存在的可能性和边界框坐标，而不需要锚参考，并为可能包含对象的每个位置生成类别不可知的边界框。在文献[27]中，RepPoints 是一种用于目标检测的新的表示形式，它包括一组指示对象空间范围和语义显著局部区域的点，从而实现更细粒度的定位以便分类。然而，由于缺乏针对小目标检测的特定设计，这个问题并没有得到很好的解决。

2)基于 Transformer 的检测。Carion 等提出了 DETR[28]，一种使用 Transformer 架构的端到端目标检测模型。DETR 通过序列到序列的方式进行目标定位和分类，建立了目标检测的新范式。文献[29]在 DETR 的基础上引入了可变形卷积，大大加快了模型在训练阶段的收敛速度，同时提升了模型对目标形变的感知能力和准确性。PnP-DETR[30]通过引入逐点学习机制，实现了对局部信息的更好建模，提高了目标检测的性能。Conditional-DETR[31]利用条件编码器，将语义分割掩码作为附加信息，增强了目标检测模型对物体边界的精细感知能力。Anchor-DETR[32]通过引入锚点机制，以及在 Transformer 编码器中使用注意力融合策略，提升了目标检测模型的性能和鲁棒性。

3)小目标检测。随着卷积神经网络(CNN)的快速发展，研究人员开始专门探索小目标检测的框架。Lin 等[2]提出了特征金字塔网络(FPN)，使用自上而下的架构和侧向连接作为优雅的多尺度特征扭曲方法。基于 FPN，许多研究人员专注于人脸检测和遥感目标检测，其中物体大小通常较小。

Zhang 等[33]提出了一个基于 Anchor 的尺度不变的人脸检测器。针对小目标人脸检测性能低的问题，该检测器的优化设计包含 3 个方面：提高 Anchor 设置的密度，增加目标被匹配的 Anchor 的数量，降低真实标签为背景的 Anchor 被误认为目标的概率。最近，Li 等[34]提出了 DSFD 用于人脸检测，除了设计新的特征增强模块和优化 Anchor 匹配策略外，还针对小目标设计了一种渐进式损失函数。首先分阶段设置了不同大小的 Anchor，其中第一阶段 Anchor 大小只有第二阶段的一半，用于提取高分辨的特征信息来提高小目标的检测性能，然后据此设计了两个分支对应的不同的损失函数。所提方式在 WIDER FACE 和 FDDB 数据集上都取得了最好的表现。Pang 等[35]提出了用于大规模遥感图像中快速小目标检测的遥感区域卷积神经网络(R2-CNN)。R2-CNN 由 TinyNet 骨干网络、中间的全局注意块以及最终的分类器和检测器组成，整个网络在计算和内存消耗方面都非常高效，可以有效地检测小目标。Yang 等[36]采用特征融合的方式使用残差网络不同层次的特征，来解决小目标信息不充分的问题。同时通过改变特征图的大小来灵活地选择合适的 Anchor 的步长，避免了传统网络中过大的 Anchor 步长设计会忽略过小的物体的缺点。该方法还考虑了物体的旋转信息，因此在遥感检测领域具有广泛的应用价值。

在多类别和多实例标注的小目标检测场景中，Lee 等[37]提出了一种新的结构，可以基于给定的用户输入，利用局部和整体信息提高小目标的检测性能。Kim 等[38]发现，在行人检测中，小尺度的行人相比大尺度特征信息更模糊，更难以和背景分离。Kim 等从人的记忆力机制出发，认为如果人充分了解大尺度目标的相关信息，那么当观测那些特征不充分的小尺度目标时，可以通过“线索召回”的思维方式来识别物体。据此，他们设计了一个记忆网络来学习大尺度目标的先验信息，然后和小尺度目标的检测网络共享权重来模拟“线索召回”的过程。Xu 等[39]回顾了在航天图像小目标检测中使用 IoU 作为匹配预测框和真实框的标准的缺点，并提出了一个新的计算两者相似度的系数，即归一化的高斯瓦瑟斯坦距离。为了缓解极小物体在训练过程中匹配 Anchor 数量过少的问题，Xu 等还优化了 Anchor 的分配策略，使用基于排名的相对分配策略来替换使用阈值的绝对分配策略，并在消融实验中展示了这两个具体改进的有效性。

3　基于尺度分布搜索的人体小目标检测方法

3.1　尺度分布建模

1)三阶段训练。深度学习视觉模型训练通常分两阶段：首先在大型通用数据集(如 ImageNet[6])上进行监督或自监督训练，接着在特定子任务数据集上对预训练模型进行微调。但是，在某些领域，大规模数据集难以获取。此时，将第二阶段进一步细分为两个子阶段，可提升训练效果，即先在大规模数据集上进行预训练，然后在子任务数据集上进行微调。例如，在目标检测中，可以先在 COCO 数据集上进行预训练，然后在特定子任务数据集如人脸检测或交通标志数据集上进行微调。这种三阶段训练策略在自然语言领域也被广泛应用。我们用以下符号来表示三阶段的训练过程：

$$train(M,UD)\rightarrow train(M,RD)\rightarrow train(M,D) \tag{1}$$

其中，M 表示神经网络模型，UD，RD，D 分别表示上述 3 种类型的数据集。第三个 $train$ 特指在子任务数据集 D 上进行微调。

2)数据集尺度的变换。在深度学习训练中，如果第二阶段预训练使用的数据集尺度与特定子任务数据集尺度不匹配，则可能会影响预训练的效果。这一问题在子任务为小目标检测、预训练数据集为通用数据集时尤为显著，原因是忽略了“尺度”这一关键特征。为解决此问题，Yu 等[40]提出了一种“尺度匹配”的方法，调整 COCO 图片的尺寸，使其与 TinyPerson 目标的尺度分布匹配。我们使用 $T(RD)$ 表示数据集的尺度变换：

$$T(RD)=\{Resize(I;s),s\sim Scale(D)\mid I\in RD\} \tag{2}$$

3)使用高斯分布建模数据集的尺度。因为 COCO 和 TinyPerson 中的目标尺寸存在显著差异，因此大部分对 COCO 的尺度变换都是缩小的。虽然这种直接缩放 COCO 数据集的预训练方法提高了 TinyPerson 上的任务性能，但相比使用原始 COCO，尺度变换可能导致图像信息损失。我们期望找到一种最佳的尺度变换方法，以平衡这种性能提升与信息损失。尺度分布搜索(SDS)的框架如图 2 所示，该框架用于小物体检测。

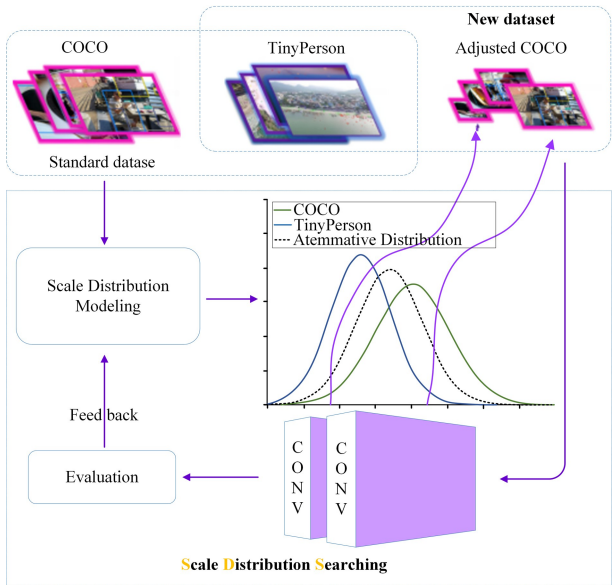


图 2　尺度分布搜索(SDS)的框架

Fig. 2　Framework of scale distribution searching(SDS)

我们在大量数据集上的研究发现，RD 类型的数据集与 D 类型的数据集的尺度分布的对数基本符合高斯分布。因此，我们考虑使用高斯模型对数据集中目标的尺度分布进行建模，从而将寻找 RD 和 D 之间最优尺度变换的问题转化为寻找最优的高斯分布的问题。因此，在三阶段训练过程中，第二步将按照式(3)和式(4) 进行修改：

$$\mu^*,\sigma^*=\underset{\mu,\sigma}{\operatorname{argmax}} \tag{3}$$

$$train(M,T(RD;\mu^*,\sigma^*)) \tag{4}$$

式(1)被转换成：

$$train(M,UD)\otimes train(M,T(RD;\mu^*,\sigma^*))\otimes$$
$$train(M,D)\otimes evaluation(M,D_{eva}) \tag{5}$$

其中，$a\otimes b$ 表示操作 a 的输出作为操作 b 的输入，D_{eva} 指 D 类型数据集的验证集。

3.2 尺度分布搜索算法

为寻找最优的高斯分布，本文提出了一种搜索算法——尺度分布搜索(SDS)。SDS 的输入包括随机初始化的均值和标准差、算法迭代次数 M_e、近邻数 nei、衰减系数 $step\_decay\_exp$，输出则是最优参数(μ^*，σ^*)，其具体步骤如下。

1)初始化。本文使用随机生成的均值和标准差(μ_0，σ_0)作为初始值，根据式(5)完成整个实验。模型的性能e_x以及对应的参数(μ_0，σ_0)被保存在列表 E 中。我们将最佳性能e_{best}设置为e_x，并将(μ_0，σ_0)分配给 x。

2)参数更新。我们通过生成 x 的 K 个近邻数来搜索最佳参数，表示为$\{nei\}=\{nei=(\mu_i,\sigma_i),1\leqslant i\leqslant K\}$。通过式(5)得到每个近邻条件下模型的性能，并将结果添加到列表 E 中。如果使用(μ_i，σ_i)模型的性能(表示为e_{nei})高于当前e_{best}的值，则更新e_{best}和x_{best}。

3)迭代策略。本文经过多次迭代来更新参数。经过参数更新后，我们得到了 K 个参考点(μ_i，σ_i)，$1\leqslant i\leqslant K$ 和它们对应的模型性能。将 K 个性能归一化得到 K 个参考点对应的转移概率，下一次迭代的均值和方差在 K 个参考点中选择。通过生成(0，1)之间的随机数来确定(μ_i，σ_i)。

算法 1 尺度分布搜索

输入：初始尺度分布的均值和方差(μ_0，σ_0)，实验最大迭代次数 max_iter，近邻数 nei，步长 step，衰减系数 step_decay_exp

输出：最优尺度分布的均值和方差($\mu*$，$\sigma*$)

1. x←(μ_0，σ_0)
2. e_x←experiment(x)
3. E←{(x，e_x)}
4. e_{best}←e_x
5. t_e←0
6. While t_e< max_iter do
7. t_e←t_e+1
8. neis ← choose_neighbour(x，step，K)
9. For nei∈neis do
10. If not∃(nei，e_{nei})∈E then
11. e_{nei}←experiment(nei)
12. E←E∪{(nei，e_{nei})}
13. End if
14. If e_{nei}>e_{best} then
15. e_{best}←f_{nei}
16. x_{best}←x
17. End if
18. End for
19. $P_{x'}$←$\exp^{\gamma e_{x'}}/\sum_{x'\in neis\cup\{x,x_{best}-x\}}\exp^{\gamma e_{x'}}$
20. x←random_choice(x′)~$P_{x'}$
21. If e_{best} have not update for t experiments then
22. step←step * decay
23. End if
24. End while
25. Return x_{best}

4)衰减系数。在实验中，我们观察到参数e_{best}的值较早地停止了更新。为了不让e_{best}陷入局部最优，我们在 SDS 中添加了一个衰减系数来解决这个问题。当e_{best}在 t 次实验后仍然没有更新，则使用衰减系数去降低生成 K 个近邻数时的步幅。

4 实验结果及分析

首先，介绍用于训练和评估模型的数据集。其次，详细阐述用于验证 SDS 算法正确性和有效性的实验设置。然后，将提出的方法与最先进的小目标检测方法以及常见的目标检测方法进行比较。最后，强调本文算法的特性，即无需特定的检测器和数据集。

4.1 数据集

TinyPerson[40]数据集是专门用于小目标检测的集合，包括 1610 张图像，涵盖了 72651 个低分辨率的人类对象注释。为确保在对象数量合理的图像上进行实验，我们选取了含有少于 200 个对象的图像进行训练和测试。所有图像被划分为 640×512 的子图像，训练集包含 8256 个子图像，测试集则有 17693 个子图像，我们并未使用验证集。

4.2 实验设置

1)区域忽略。在 TinyPerson 数据集中，存在一些难以被归类为前景(正样本)或背景(负样本)的区域，通常这些区域包含多个个体或物体，需要在训练过程中予以忽视。解决此问题有两种策略：(1)用训练集图像的平均值替换这些忽略区域；(2)忽略来自这些区域的梯度，即在训练过程中不进行反向传播。本文采取了第一种策略。

2)图像切割。TinyPerson 数据集中大部分图像具有较大尺寸，这可能导致 GPU 显存在训练和评估过程中不足。为解决此问题，我们将原始图像切分为带有重叠区域的较小子图，然后利用非极大值抑制(NMS)策略将子图结果融合为单一输出以进行评估。虽然这种策略有助于更有效地使用 GPU 资源，但存在两个潜在缺点：(1)对于特征金字塔网络(FPN)，不含任何目标的纯背景图像并不会用于训练，但图像切割可能导致许多子图成为纯背景图像；(2)在特定情况下，NMS 可能无法有效融合重叠区域的结果。

3)实施细节。本文的实验基于 Facebook 的 maskrcnn-benchmark 并选用 ResNet50 作为网络主干，而默认检测器设为 Faster RCNN-FPN。网络训练周期为 12 个 epochs，初始学习率设为 0.01，并在第 6 和第 10 个 epoch 时均将其乘以 0.1。实验是在两个 2080Ti GPU 上进行的。为了确定锚点大小，我们将其聚类得出(8.31，12.5，18.55，30.23，60.41)并设定纵横比为(0.5，1.3，2)。鉴于 TinyPerson 数据集中部分图像含有密集目标，我们将每张图像的最大检测器输出框数(DETECTIONS PER IMG)设定为 200。

在 3.2 节所示的尺度搜索算法中，我们通过多次实验来确定最大迭代次数 $max\_iter=10000$，近邻数的数量 $K=8$，步长 $step=0.01$，衰减系数 $step\_decay\_exp=0.2$。

4)数据增强。我们在训练数据增强中仅使用了水平翻转。区别于其他基于 FPN 的检测器，我们并未调整所有图像至相同尺寸，而是保持了原始图像/子图像的尺寸。在模型性能评估方面，我们采用了 Tinybenchmark 推荐的平均精度(AP)和漏报率(MR)作为度量指标。AP 指标旨在反映检测

结果的精准度和召回率，而由于 TinyPerson 主要为行人数据集，因此额外采用了 MR 作为评估标准。本文设定的 IoU 阈值为 0.25，0.5 和 0.75。为了更全面地评估小目标检测性能，Tinybenchmark 将尺度范围[2，20]进一步细分为 tiny1[2，8]，tiny2[8，12]和 tiny3[12，20]这 3 个子区间并将其进行比较。

4.3　与 SOTA 检测器的对比

为了比较 TinyPerson 上的 SDS 性能与最先进的检测器，我们选择了 RetinaNet[19]，FCOS[22]和 Faster RCNN-FPN[2]。其中，RetinaNet，FCOS 和 Faster RCNN-FPN 分别代表了单阶段锚点检测器、无锚点检测器和双阶段锚点检测器。为了保证收敛，我们使用了训练 Faster RCNN-FPN 网络时学习率的一半作为 RetinaNet 的学习率，并将 FCOS 的学习率设置为 1/4。对于自适应 FreeAnchor[20]，我们使用了与自适应 RetinaNet 相同的学习率和骨干网络设置，同时保持其他设置与 FreeAnchor 的默认设置相同。表 1 和表 2 中还列出了一些其他方法，如 SCRDet[36]和 DSFD[34]，它们分别用于弱小人脸检测以及小型、杂乱和旋转对象检测等具有类似绝对尺度分布的对象检测。

1)位置精度较差。当 IoU 阈值从 0.25 增加到 0.75 时，性能显著下降，表明位置精度较差。由于小物体有绝对和相对尺寸，因此很难在 TinyPerson 中具有高的位置精度。

2)更好的检测器。对于 COCO 这样的数据集，其中包含许多大尺寸的目标，我们发现 RetinaNet 和 FreeAnchor 的性能优于 Faster RCNN-FPN。这说明只要解决了样本不平衡问题，单阶段检测器也能超过双阶段检测器。与 RetinaNet 和 Faster RCNN-FPN 相比，无锚点检测器 FCOS 的表现更出色。然而，当目标的尺寸变得非常小，比如在 TinyPerson 数据集中。在这种情况下，RetinaNet 和 FCOS 的性能下降得更为严重，这表明双阶段检测器相对于单阶段检测器在这方面具有优势，如表 1 和表 2 所列。

表 1　基于 TinyPerson 数据集的 MR 性能对比

Table 1　MR performance comparison on TinyPerson

Detector	MR_{50}^{tiny1}	MR_{50}^{tiny2}	MR_{50}^{tiny3}	MR_{50}^{tiny}	MR_{50}^{small}	MR_{25}^{tiny}	MR_{75}^{tiny}
FCOS[22]	99.23	96.56	91.67	96.28	84.16	90.34	99.56
Adaptive Reppoints	95.89	91.20	85.64	93.08	79.48	85.73	98.88
RetinaNet[19]	95.45	89.13	86.98	92.86	82.02	82.56	99.11
GCNet[41]	90.57	85.57	82.56	89.67	74.38	84.16	98.50
Adaptive Free Anchor	90.79	83.39	82.34	89.66	73.88	79.61	98.78
Adaptive Retina Net	90.46	83.99	82.96	90.08	75.59	80.18	98.67
Libra RCNN	90.93	84.64	81.62	89.22	74.86	82.44	98.39
Doube Head	88.00	83.35	79.45	88.26	72.32	77.76	98.37
Cascade RCNN[42]	88.70	82.87	79.11	88.26	72.84	79.62	98.40
Grid RCNN[43]	88.31	82.79	79.55	87.96	73.16	78.27	98.21
Faster RCNN-FPN[2]	89.02	81.88	79.52	87.85	71.72	78.76	98.46
SCRDet[36]	98.23	94.62	89.65	95.31	82.20	88.23	99.63
DSFD[34]	96.41	88.02	86.84	93.47	78.72	78.02	99.48
SDS-FPN	**88.82**	**80.49**	**79.23**	**87.45**	**71.41**	**78.33**	**98.06**

表 2　基于 TinyPerson 数据集的 AP 性能对

Table 2　AP performance comparison on TinyPerson

Detector	AP_{50}^{tiny1}	AP_{50}^{tiny3}	AP_{50}^{tiny3}	AP_{50}^{tiny4}	AP_{50}^{small}	AP_{25}^{tiny}	AP_{25}^{tiny}
FCOS[22]	2.88	12.95	31.15	17.90	40.54	41.95	1.50
Adaptive Reppoints	15.00	30.28	44.33	30.54	51.00	50.79	3.84
Retina Net[19]	11.37	38.64	46.52	33.37	48.43	60.95	2.40
GCNet[41]	28.68	45.76	53.05	43.09	62.21	61.33	5.32
Adaptive Free Anchor	25.99	49.37	55.34	44.26	60.28	67.06	4.35
Adaptive Retina Net	26.14	49.19	55.58	44.27	58.42	67.18	4.62
Libra RCNN	27.08	49.27	55.21	44.68	62.65	64.77	6.26
Doube Head	30.33	50.08	58.15	46.88	64.45	67.52	6.17
Cascade RCNN[42]	30.89	50.75	57.83	46.97	62.19	67.01	6.00
GridRCNN[43]	30.65	52.21	57.21	47.14	62.48	68.89	6.38
FasterRCNN-FPN[2]	30.15	51.77	58.54	47.56	62.68	67.71	5.88
SCRDet[36]	4.19	18.54	36.51	21.95	48.14	51.15	1.46
DSFD[34]	13.85	37.24	49.31	33.65	56.64	63.18	1.94
SDS-FPN	30.45	51.95	58.70	47.82	62.91	68.06	6.24

除此以外，我们发现通过研究和利用尺度信息可以缓解性能的下降。通过应用本文的 SDS，Faster RCNN-FPN 在 TinyPerson 数据集上实现了更好的性能，MR_{50} 达到了 87.45，AP_{50} 达到了 47.82，证实了本文提出的预训练策略的有效性。

结束语　随着深度卷积神经网络的兴起，视觉目标检测取得了前所未有的进步。然而，在大规模图像中检测小物体（例如小于 20 像素的小目标）仍然没有得到很好的研究。庞大且复杂的背景则增加了误检测的风险，小物体的表征仍然是一个艰难的挑战。基于此，本文提出了一种简单而有效的方法——尺度分布搜索（Scale Distribution Search，SDS）。SDS 利用高斯模型建模小目标数据集和预训练数据集的尺度特征，搜索两个数据集尺度特征之间一种最优的变换方式。我们测试了不同变换方式下小目标数据集在相同检测模型下的性能，并筛选出了一种最优的变换方式作为最终的实验

参数。SDS可以更高效地利用预训练/协同训练数据集，特别是那些尺度分布差异很大的数据集，例如TinyPerson和MS COCO。实验结果表明，本文方法在小目标检测方面相比现有最先进的检测器有显著的性能提升，但是SDS应用的搜索策略在每次迭代搜索参数时都要重新训练模型，效率较低。如何在保证现有搜索精度的条件下，利用之前训练的模型的信息来提升搜索效率是下一步值得努力的方向。

参 考 文 献

[1] LINT Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// 13th European Conference (ECCV 2014). 2014: 740-755.

[2] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.

[3] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]// NIPS. 2016.

[4] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248-255.

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

[6] ISLAM M R, MATIN A. Detection of COVID 19 from CT image by the novel LeNet-5 CNN architecture[C]// 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE, 2020: 1-5.

[7] ALEX K, ILYA S, GEOFFREY E H. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.

[8] LIN M, CHEN Q, YAN S C. Network in network [J]. arXiv: 1312.4400, 2013.

[9] KAREN S, ANDREW Z. Very deep convolutional networks for large-scale image recognition[C]// 3rd International Conference on Learning Representations (ICLR 2015). Conference Track Proceedings, 2015.

[10] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[11] ZHANG C, BENZ P, ARGAW D M, et al. Resnet or densenet: introducing dense shortcuts to resnet[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3550-3559.

[12] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[C]// 3rd International Conference on Learning Representations (ICLR 2014). 2014.

[13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[14] 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.

[15] GIRSHICK R. Fast r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.

[16] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.

[17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.

[18] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// Computer Vision-ECCV 2016: 14th European Conference, Amsterdam. Springer International Publishing, 2016: 21-37.

[19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.

[20] ZHANG X, WAN F, LIU C, et al. FreeAnchor: Learning to Match Anchors for Visual Object Detection[J]. arXiv: 1909.02466, 2019.

[21] LAW H, DENG J. CornerNet: Detecting Objects as Paired Keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.

[22] DUAN K, BAI S, XIE L, et al. CenterNet: Keypoint Triplets for Object Detection[J]. arXiv: 1904.08189, 2019.

[23] YANG T, ZHANG X, LI Z, et al. MetaAnchor: Learning to Detect Objects with Customized Anchors[J]. arXiv: 1807.00980, 2018.

[24] WANG J, CHEN K, YANG S, et al. Region proposal by guided anchoring[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2965-2974.

[25] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9627-9636.

[26] KONG T, SUN F, LIU H, et al. Foveabox: Beyound anchor-based object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 7389-7398.

[27] YANG Z, LIU S, HU H, et al. Reppoints: Point set representation for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9657-9666.

[28] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// Computer Vision-ECCV 2020: 16th European Conference. Springer International Publishing, 2020: 213-229.

[29] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection[J]. arXiv: 2010.04159, 2020.

[30] WANG T，YUAN L，CHEN Y，et al. Pnp-detr：Towards efficient visual analysis with transformers[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：4661-4670.

[31] MENG D，CHEN X，FAN Z，et al. Conditional detr for fast training convergence[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：3651-3660.

[32] WANG Y，ZHANG X，YANG T，et al. Anchor detr：Query design for transformer-based detector[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022：2567-2575.

[33] ZHANG S，ZHU X，LEI Z，et al. S3fd：Single shot scale-invariant face detector[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017：192-201.

[34] LI J，WANG Y，WANG C，et al. DSFD：dual shot face detector[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019：5060-5069.

[35] PANG J，LI C，SHI J，et al. R2-CNN：Fast tiny object detection inlarge-scale remote sensing images[J]. arXiv：1902. 06042，2019.

[36] YANG X，YANG J，YAN J，et al. Scrdet：Towards more robust detection for small，cluttered and rotated objects[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019：8232-8241.

[37] LEE C，PARK S，SONG H，et al. Interactive Multi-Class Tiny-Object Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022：14136-14145.

[38] KIM J U，PARK S，RO Y M. Robust small-scalepedestrian detection with cued recall via memory learning[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021：3050-3059.

[39] XU C，WANG J，YANG W，et al. Detecting tiny objects in aerial images：A normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2022，190：79-93.

[40] YU X，GONG Y，JIANG N，et al. Scale match for tiny person detection[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020：1257-1265.

[41] CAO Y，XU J，LIN S，et al. Gcnet：Non-local networks meet squeeze-excitation networks and beyond[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[42] CAI Z，VASCONCELOS N. Cascade r-cnn：Delving into high quality object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：6154-6162.

[43] LU X，LI B，YUE Y，et al. Grid r-cnn[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019：7363-7372.

YANG Yi，born in 1981，Ph. D candidate. His main research interests include machine learning and computer vision.

LI Yuan，，born in 1987，Ph.D，assistant lecturer，Her main research interests include computer vision and intelligent emergency.

（责任编辑：喻藜）