

基于数据增强和监督均衡的鱼眼图像行人检测优化方法

司绍峰, 张赛强, 李庆, 陈本瑶

引用本文

司绍峰, 张赛强, 李庆, 陈本瑶. [基于数据增强和监督均衡的鱼眼图像行人检测优化方法](#) [J]. 计算机科学, 2022, 49(11A): 210900070-6.

SI Shao-feng, ZHANG Sai-qiang, LI Qing, CHEN Ben-yao. [Pedestrian Detection Optimization Method Based on Data Enhancement and Supervised Equalization in Fisheye Image](#) [J]. Computer Science, 2022, 49(11A): 210900070-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于点云数据的交通环境下单阶段三维目标检测方法](#)

Single-stage 3D Object Detector in Traffic Environment Based on Point Cloud Data
计算机科学, 2022, 49(11A): 210900079-6. <https://doi.org/10.11896/jsjcx.210900079>

[融合词性与声调特征的越南语语法错误检测](#)

Incorporating Part of Speech and Tonal Features for Vietnamese Grammatical Error Detection
计算机科学, 2022, 49(11): 221-227. <https://doi.org/10.11896/jsjcx.210900247>

[不同数据增强方法对模型识别精度的影响](#)

Influence of Different Data Augmentation Methods on Model Recognition Accuracy
计算机科学, 2022, 49(6A): 418-423. <https://doi.org/10.11896/jsjcx.210700210>

[基于接缝一致性准则的结构纹理图像快速合成算法](#)

Fast Structural Texture Image Synthesis Algorithm Based on Seam Consistency Criterion
计算机科学, 2022, 49(6): 262-268. <https://doi.org/10.11896/jsjcx.210400039>

[自适应权重的级联增强节点的宽度学习算法](#)

Adaptive Weight Based Broad Learning Algorithm for Cascaded Enhanced Nodes
计算机科学, 2022, 49(6): 134-141. <https://doi.org/10.11896/jsjcx.210500119>

基于数据增强和监督均衡的鱼眼图像行人检测优化方法

司绍峰^{1,2} 张赛强^{1,2} 李庆² 陈本瑶³

1 中国科学院大学 北京 100049

2 中国科学院微电子研究所 北京 100029

3 湖州市特种设备检测研究院 浙江 湖州 313000

(1029865870@qq.com)

摘要 近年来,由于鱼眼相机被广泛应用于智能监控领域,不少学者提出了针对鱼眼图像的行人检测算法。然而,鱼眼图像场景复杂且存在畸变,其导致的数据集样本分布和算法监督分配的不均衡问题会降低检测器性能。针对上述问题,首先提出了一种针对鱼眼图像行人检测任务的数据增强方法,该方法由模式采样增强和角度直方图增强两部分组成。其中模式采样增强专注于鱼眼图像难例样本挖掘,生成的新样本丰富了鱼眼图像中心附近的行人模式;角度直方图增强基于直方图均衡的思想,对鱼眼图像样本角度分布做平滑处理,缓解了检测器对单一场景的过拟合问题。此外,基于鱼眼图像 Anchor-Free 行人检测器,提出将定位质量预测与监督信息融合,将 Focal Loss 推广到连续域用以优化检测器定位分支的监督分配。实验结果表明,所提数据增强算法能够有效缓解鱼眼图像数据集的分布不均衡,在 Anchor-Based 和 Anchor-Free 检测网络上均展现了较好的效果;连续 Focal Loss 结合定位质量优化监督,在不增加 Anchor-Free 检测器推理计算复杂度的前提下,将整体性能提升了 3.8%。

关键词: 鱼眼图像;行人检测;数据增强;Anchor-Free 检测器;连续 Focal Loss

中图法分类号 TP391

Pedestrian Detection Optimization Method Based on Data Enhancement and Supervised Equalization in Fisheye Image

SI Shao-feng^{1,2}, ZHANG Sai-qiang^{1,2}, LI Qing² and CHEN Ben-yao³

1 University of Chinese Academy of Sciences, Beijing 100049, China

2 Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100029, China

3 Huzhou Special Equipment Inspection Center, Huzhou, Zhejiang 313000, China

Abstract In recent years, due to the fisheye camera is widely used in the field of intelligent monitoring, many scholars propose pedestrian detection algorithm for fisheye image. However, the fisheye scene is complex and distorted, which leads to the imbalance of data set sample distribution and algorithm supervision allocation, which will reduce the performance of the detector. To solve these problems, a data enhancement method for pedestrian detection in fisheye image is proposed, which consists of pattern sampling enhancement and angle histogram enhancement. The pattern sampling enhancement focuses on the mining of difficult samples in fisheye image, and the generated new samples enrich the pedestrian patterns near the center of fisheye image. Angle histogram enhancement is based on the idea of histogram equalization, which smooths the angle distribution of fisheye image samples to alleviate the over fitting problem of single scene. In addition, based on Anchor-free fisheye image pedestrian detector, the fusion of location quality prediction and supervision information is proposed to extend Focal Loss to continuous domain to optimize the supervision allocation of detector location branches. Experimental results show that the proposed data enhancement algorithm can effectively alleviate the uneven distribution of fisheye image data set, and show good results in both Anchor-Based and Anchor-Free detectors. The continuous Focal Loss optimization localization supervision method improves the overall performance by 3.8% without increasing the reasoning complexity of the Anchor-Free detector.

Keywords Fisheye images, Pedestrian detection, Data enhancement, Anchor-Free detector, Continuous Focal Loss

1 引言

行人检测作为图像处理的一个分支,其核心思想是将计算机视觉和数字图像处理相结合,通过对监控设备获取的

行人行为信息进行分析处理,来判断图像中是否存在行人并给出精确定位。相比普通摄像机,鱼眼摄像机因拥有超广角镜头,可以实现大范围清晰成像,而被广泛应用于各种监控场景。然而,由于鱼眼图像存在畸变失真,这对鱼眼图像中的

基金项目:基于电梯安全一体化管理与应用技术的研(NQI20200125)

This work was supported by the Research on Integrated Management and Application Technology of Elevator Safety(NQI20200125).

通信作者:李庆(liqing@ime.ac.cn)

行人检测提出了更高的要求。传统的鱼眼图像行人检测算法通常基于斑点提取和前景分割^[1-2],但该类方法的检测精度受场景变化的影响较大,应用范围有限。近年来,因深度学习在行人检测领域取得巨大成功,不少学者试图引入深度学习方法来解决鱼眼图像的行人检测问题。Demirkus 等^[3]基于鱼眼图像中行人的几何特征提出了一个俯视场景的鱼眼基准,并将该基准与最先进的行人检测器 ACF^[4]相结合用于鱼眼图像行人检测;Krams 等^[5]使用几何关系来计算人体在鱼眼图像上特定位置的大致高度,并利用这些信息来辅助行人检测。然而,以上两种算法只适用于特定的区域,而不是整个鱼眼图像,整体检测效果不佳。Qian 等^[6]利用基于 EHT 行人基准的鱼眼相机成像原理,提出了一种新的畸变行人检测基准,利用该基准,畸变的行人可以根据在鱼眼图像中的位置来进行不同的训练,但该方法提出的行人基准鲁棒性较差,极大地影响了行人检测的准确性。相比以上加入鱼眼基准等先验信息的方法,Duan 等^[7]直接将单阶段目标检测网络 YOLOv3^[8]用于鱼眼图像行人检测,使用带有角度参数的旋转锚框来锚定畸变的行人,同时使用旋转增强和颜色变换等方法来减轻鱼眼图像样本不均衡造成的影响。该方法在检测精度上优于之前的方法,但其数据增强方法针对性较弱,效果有限;且引入旋转目标框会增加计算复杂度,导致检测速度较慢,不能满足行人检测任务对高实时性的要求。

通过以上研究,本文分析了目前鱼眼图像行人检测存在的问题。

(1)鱼眼相机因其独特的成像特点,存在着行人沿半径和角度分布极不均衡的情况。训练样本不均衡会导致模型在鱼眼图像不同位置的检测效果存在较大差异,然而目前的目标检测数据增强方法并不完全适用于鱼眼图像。

(2)鱼眼图像上畸变行人的形状和朝向任意,需要使用旋转目标框来锚定。引入旋转框后,Anchor-Based 行人检测器因在 NMS 阶段需进行大量旋转框 IoU^[7]的计算,检测速度变慢,无法满足高实时性的需求。

(3)目前的 Anchor-Free 框检测器对目标框定位质量评估的监督只分配给标注框中心,而将其他位置作为负样本处理。但由于鱼眼图像存在畸变,其负样本会获得不可控的更高质量的预测,这种监督信息的不均衡分配会降低检测器检测性能。

针对目前鱼眼图像行人检测算法存在的问题,本文从鱼眼图像中的行人分布入手,提出了一种针对鱼眼图像的数据增强方法,该方法分为模式采样增强和角度直方图增强两个阶段,可有效改善鱼眼图像中行人模式和分布不均衡的问题。在检测算法上,本文首先采用 Anchor-Free(无锚框)网络结构,避免了推理阶段 NMS 操作带来的高计算复杂度;其次将定位质量的监督分配给锚框内的所有点来调节网络生成的高斯热图的正负样本比例;最后将 Focal Loss^[9]推广到连续域,得到连续 Focal Loss 用于定位分支的监督优化。实验结果表明,本文算法在满足高实时性的前提下,改善了鱼眼图像行人检测的不均衡问题,进一步提升了 Anchor-Free 网络的行人检测精度,具有较高的实际应用价值。

2 相关工作

2.1 经典目标检测网络

近些年来,目标检测算法取得了巨大的突破。具有代表

性的算法可分为两类,其中基于候选区域提案(Region Proposal)^[10]的 RCNN 系列算法,如 RCNN^[11],Fast RCNN^[12],Faster RCNN^[10],Mask RCNN^[13]等,将目标检测分为目标候选框生成和基于候选框的分类回归两个阶段。该类算法通过将一阶段生成的粗粒度候选框进行第二阶段细粒度的分类与回归,并且端到端地进行训练,使得检测精度有了较大提升。第二类检测算法如 YOLO^[14]和 SSD^[15]等,通过一阶段架构直接生成图像的细粒度回归与分类结果,其在大大提升了网络检测速度的同时,也取得了良好的检测效果。之后,由于 Focal Loss^[9]的提出可以缓解单阶段网络训练中正负样本的不均衡问题,该类算法可以达到与二阶段检测网络相当的检测精度。2018 年提出的一阶段目标检测网络 CornerNet^[16],创新性地将目标检测问题转换为关键点检测问题,并在检测精度上超越了二阶段检测网络,为目标检测领域带来了新的思路。

2.2 Anchor-Free 设计

作为一阶段目标检测算法中的一种,Anchor Free 方法采用基于关键点的无锚框设计方法,摒弃了经典算法为网络预设锚框,并在此基础上进行精细分类和回归的思路,提升了目标检测算法的鲁棒性。自 CornerNet 取得了较优的性能表现,不断有新的 Anchor-Free 算法被提出。CenterNet^[17]利用高斯核生成中心点热力图,从而保证网络在目标中心附近产生更强大的激活值。对于回归问题,CenterNet 在 CornerNet 的基础上,预测一个偏移来恢复下采样造成的离散化误差,提升了检测精度。FCOS^[18]通过将特征金字塔结构^[19](Feature Pyramid Network)引入 Anchor-Free 网络来获取图像高层语义信息,提升了网络召回能力。本文的算法框架基于 Anchor-Free 检测网络 TTFNet^[20],采用高斯核控制参与回归的样本数量,通过高斯核函数在目标中心周围构造子区域,然后从该区域提取密集的采样点作为训练样本来加速收敛,又避免了 FCOS 那样的 NMS 后处理,从而实现了训练时间、推理速度和精度之间的较好平衡。

2.3 数据增强

在目标检测领域,对数据集进行数据增强能够增加样本多样性,缓解数据分布的不均衡问题,显著提高模型的泛化能力。传统的数据增强方法主要包括图像旋转、翻转、尺度缩放、裁剪以及对图像添加噪声和颜色变换等。VGG^[21]和 AlexNet^[22]使用了随机裁剪和图像翻转的数据增强方法,有效地提升了在 ImageNet^[23]数据集上小样本类别的召回率。YOLOv4^[24]使用 mosaic 数据增强,将 4 张图片按不同比例拼接成一张用于训练,可以丰富物体检测的背景信息,让检测器检测出超出常规语境的目标。Zhang 等^[25]提出了基于邻域风险最小化的数据增强方法 Mixup,该方法使用线性插值生成新样本,并在样本及其标签的凸组合(Convex Combinations)上训练神经网络;其增强了训练样本之间的线性表达,减小了深度学习模型的泛化误差。Goodfellow 等^[26]提出的生成对抗网络 GAN 将判别网络和生成网络进行对抗训练,最终收敛到纳什平衡,其生成网络可以生成高质量的数据样本,用于扩充小样本数据集。

3 本文方法

3.1 鱼眼图像数据增强

本文对鱼眼图像行人检测常使用的两个公开数据集

(HABBOF 数据集和 CEPDOF 数据集)中的行人边界框中心点的分布进行统计,以图像平面作为坐标平面,对所有样本点和图像中心点的 L2 距离做归一化,得到分布散点图,如图 1 所示。观察发现,其分布不均衡问题主要体现在以下两个方面:

(1)沿鱼眼图像圆形区域半径方向,中心附近(L2 距离小于 0.4)的样本数量占比极低。由于鱼眼图像存在畸变,图像中心点附近的行人模式与其他区域存在显著差异,该不均衡问题会降低检测器在图像中心附近的检测性能。

(2)鱼眼图像行人边界框中心点在鱼眼图像中心位置四周的分布并不均匀,会导致模型训练特定区域重复采样,检测器对该区域过拟合,从而影响整体检测性能。因此本文提出了模式采样增强和角度直方图增强分别优化以上两个问题。

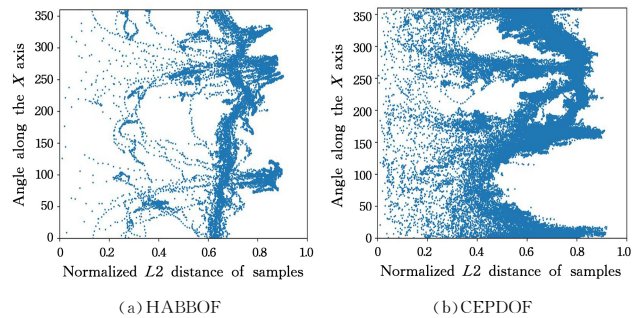


图 1 鱼眼图像数据集样本分布

Fig. 1 Sample distribution of fisheye image dataset

算法 1 模式采样增强算法

输入:同一场景下的鱼眼图像数据集(数量为 N ,图像宽为 w ,高为 h ,

图像中心位置为 $(\frac{w}{2}, \frac{h}{2})$), N 张图像中一共包含了 M 个行人

边界框)和对应的 json 标注

输出:采样增强后的输出图像(数量为 N)和对应的 json 标注

1. 分别计算每一个行人边界框中心点 (x_i, y_i) 距离鱼眼图像中心 $(\frac{w}{2}, \frac{h}{2})$ 的归一化 L2 距离 D_i ,如式(1)所示:

$$D_i = \frac{\sqrt{\left(x_i - \frac{w}{2}\right)^2 + \left(y_i - \frac{h}{2}\right)^2}}{\max\left(\frac{w}{2}, \frac{h}{2}\right)} \quad (1)$$

2. 将 D_i 在区间 $[0, 1]$ 划分成 k 等分(k 取 20),统计 k 个区间内的边界框的数量,归一化得到分布 Q ;忽略图像边缘区域畸变严重且非常小的边界框,仅仅保留 $[k * r]$ (取 r 为 0.9)组的边界框,即保留前 18 组边界框的数量,得到一个沿半径的边界框中心点的离散分布 q_i 。

3. 依据 q 分布,尝试 3 种方式得到与 q 分布 $[q_1, q_2, \dots, q_k]$ 变化趋势相反的 p 分布 $[p_1, p_2, \dots, p_k]$ 作为采样概率,结果如图 2 所示。规定 q 分布中概率为零的位置在 p 分布中概率依旧为 0,即,若 $q_i = 0$ 则 $p_i = 0$ 。

3. 1. Exchange Probability(EP 方法),对于 q 分布中的非零概率值,令在 q 分布中取极大概率值的位置,则在 p 分布对应位置取到非零概率的极小值;在 q 分布中的次大值位置,在 p 分布对应位置取到非零概率的次小值。不断重复上述操作,直到 p 分布所有位置的概率值均确定。

3. 2. Reciprocal Ratio(RR 方法),对 q 分布的非零概率位置取倒数,再归一化得到 p 分布,使得 p 分布满足 $\sum_{i=1}^k p_i = 1$ 。

3. 3. Subtraction Ratio(SR 方法),概率分布取值为 $1 - q$ 分布的非零位置,再归一化得到 p 分布,使得 p 分布满足 $\sum_{i=1}^k p_i = 1$ 。

for $i=1$ to N do

4. 利用 Q 分布和 p 分布采样边界框得到 B_{Q_i} 和 B_p 和对应的图片 I_Q 和 I_p 。这里使用的分布是关于行人边界框的,而采样是针对鱼眼图像的,这种不一致会导致有较多边界框的图像被采样的概率也较大。由于此类鱼眼图像通常为难例样本,包含更复杂的模式和更多的训练信息,故该做法会让网络聚焦于难例样本学习,从而增强网络的学习能力。

5. 将图像 I_p 对应的边界框 B_p 进行裁剪并放置到图像 I_Q 的相同位置,生成采样图像 I_s 。该操作可能会导致 B_p 和图像 I_Q 中行人边界框 B_{Q_i} 存在交叠的现象,从而产生错误的样本模式。为此使用评价指标 IoSA(Intersection over Small Area)来丢弃掉 B_p 和 B_{Q_i} 存在较大重叠的采样图像 I_s 。IoSA 含义为重叠区域面积占较小目标框面积的比例,表达式如下:

$$\text{IoSA}(B_p, B_{Q_i}) = \frac{S(B_p \cap B_{Q_i})}{\min(S(B_p), S(B_{Q_i}))} \quad (2)$$

其中, S 为计算面积,设置阈值 $T_{\text{success}} = 0.5$,对于 $\exists B_{Q_i} \in I_Q$,若 $\text{IoSA}(B_p, B_{Q_i}) > T_{\text{success}}$,则丢弃掉当前的采样结果。

6. 对仍存在较小交叠现象的采样图像 I_s ,借鉴 Mixup 的思路,如图 3 所示,对 I_s 种交叠部分 $B_p \cap B_{Q_i}$ (红框部分)采用像素级融合,使得 $I_s = \lambda I_Q + (1 - \lambda) I_p$,实验中取 $\lambda = 0.5$,表示同等看待 B_p 和 B_{Q_i} 。对于仅属于 B_p 的部分,即 $B_p \setminus (B_p \cap B_{Q_i})$,令 $I_s = I_p$,剩余部分令 $I_s = I_Q$ 。最后将边界框 B_p 的标注信息以及 I_Q 图像中所有边界框标注信息加入到采样图像 I_s 的标注信息中。

end for

7. 输出 N 张采样图像以及对应的边界框标注信息。

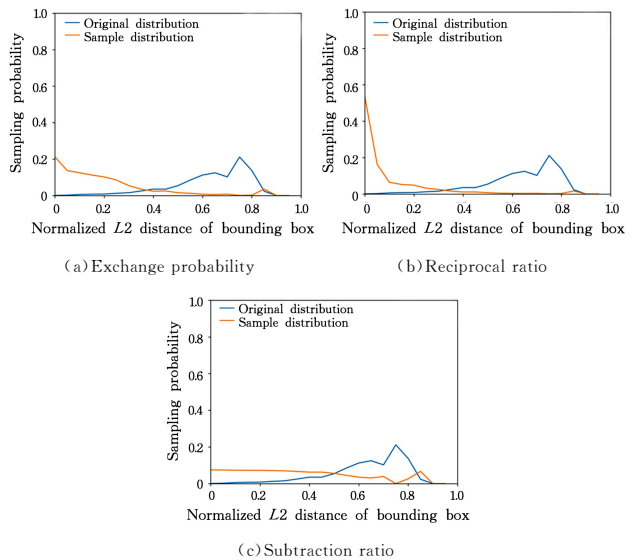
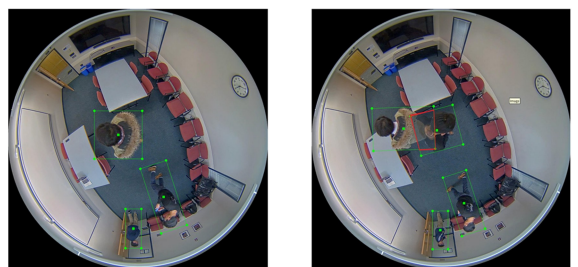


图 2 不同方法得到的采样概率

Fig. 2 Sampling probability obtained by different methods



(a) Before enhancement

(b) After enhancement

图 3 采样增强示意图

Fig. 3 Sampling enhancement diagram

算法 2 角度直方图增强算法

输入: 同一场景下的所有鱼眼图像数据集 (数量为 N , N 张图像中一共包含了 M 个行人边界框) 和对应的 json 标注

输出: M 张角度直方图数据增强后的图像和对应的 json 标注

1. 计算鱼眼图像中每一个行人边界框中心点 (x_i, y_i) 和鱼眼图像中心点 $(\frac{w}{2}, \frac{h}{2})$ 连线与图像 x 轴正方向夹角 a_i , 逆时针旋转为正, 角度范围为 $[0, 360)$ 。
for $i=1$ to M do
2. 利用直方图均衡的思想, 将每一个行人边界框的角度 a_i 依据步骤 1 中得到的边界框的角度分布计算得到 A_i , 如式(3)所示:

$$A_i = 360 \times \frac{\sum_{k=1}^M \mathbf{I}(a_k < a_i)}{M} \quad (3)$$
其中, $\mathbf{I}$ 为示性函数, A_i 为经过旋转后的边界框目标角度。
3. 将 a_i 对应的图像 I_i^{origin} 逆时针旋转 $A_i - a_i$ 角度得到图像 I_i^{angle} , 同时将 I_i^{origin} 中的标注边界框也相应地进行旋转。
end for
4. 返回 M 张增强后的图像以及对应的 json 标注文件。

角度直方图增强是对鱼眼图像旋转特定角度, 然而在进行分布统计时使用的是边界框。因此, 增强后重新统计的边界框中心点的角度分布和同一场景下的鱼眼图像中行人数量分布有关, 数量分布越集中, 增强后得到的角度分布就越均匀。

3.2 连续 Focal Loss 优化定位损失

本文的行人检测在数据增强的基础上使用无锚框检测网络 TTFNet 作为检测器, 网络头部包含目标框中心定位分支和目标框参数回归分支, 如图 4 所示。

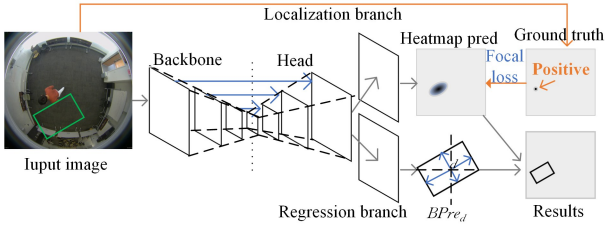


图 4 TTFNet 网络结构

Fig. 4 Architecture of TTFNet

其中, 中心定位分支输出热图特征图, 选择特征图局部最大值的位置 $(x_{\text{pred}}, y_{\text{pred}})$ 作为候选框中心点位置; 参数回归分支输出候选框中心到边界的 4 个距离以及候选框旋转角度 $B_{\text{pred}} = (d_{\text{left}}, d_{\text{right}}, d_{\text{top}}, d_{\text{bottom}}, \theta_{\text{pred}})$; 二者结合可以唯一确定目标框。

在中心定位分支的监督方面, TTFNet 采用在真实中心点附近生成对应尺度的高斯热图作为监督信号。高斯分布的峰值集为边框中心的像素, 被视为正样本, 其他像素位置均为负样本。给定目标中心点预测值 $\hat{H} \in R^{1 \times \frac{H}{r} \times \frac{W}{r}}$ 以及真实值 $H \in R^{1 \times \frac{H}{r} \times \frac{W}{r}}$ (H, W, r 分别为图片宽, 高和下采样步长), 使用 Focal Loss 作为损失函数, 如式(4)所示:

$$L_{\text{loc}} = \frac{1}{M_{\text{positive}} \cdot xy} \sum \begin{cases} (1 - \hat{H}_{ij})^{\alpha_f} \log(\hat{H}_{ij}), & \text{if } H_{ij} = 1 \\ (1 - \hat{H}_{ij})^{\beta_f} \hat{H}_{ij}^{\alpha_f} \log(1 - \hat{H}_{ij}), & \text{else} \end{cases} \quad (4)$$

其中, α^f, β^f 为 Focal Loss 损失中的超参数, 实验中设置 $\alpha^f = 2$,

$\beta^f = 4$ 。然而, 由于鱼眼图像存在畸变, 真实框中心位置与行人中心对应关系较弱, 故其高斯分布中心附近的负样本仍有可能预测质量较高的候选框, 将其视作负样本显然不合理。

为了缓解正负样本分配的不均衡问题, 本文将高斯区域内的所有点均视作正样本。且为了让特征图预测值能反映候选框的预测质量, 本文将真实值生成的高斯分布与候选框定位质量的乘积作为监督信号分配给所有正样本, 候选框的定位质量定义为候选框与真实框之间的 IoU, (x_{ij}, y_{ij}) 位置的监督信号计算式如下:

$$K_{ij}(x_{ij}^{gt}, y_{ij}^{gt}) = \exp\left(-\frac{(x_{ij}^{gt})^2}{2\sigma_x^2} - \frac{(y_{ij}^{gt})^2}{2\sigma_y^2}\right) \times \text{IoU}(B_{gt}, B_{\text{pred}}) \quad (5)$$

其中, σ_x, σ_y 用于调节高斯分布, B_{gt}, B_{pred} 分别为真实框与候选框。

改进的监督信号为 $0 \sim 1$ 上的连续值, 故式(4)中针对离散值的 Focal Loss 不再适用。因此, 本文将 Focal Loss 推广到连续区域得到广义 Focal Loss 作为定位分支的损失函数, 如式(6)所示:

$$L_{\text{loc}}^{\text{new}} = \frac{1}{M_{\text{positive}} \cdot xy} \sum -|\hat{H}_{ij} - H_{ij}|^{\beta} ((1 - H_{ij}) \log(1 - \hat{H}_{ij}) + \hat{H}_{ij} \log(\hat{H}_{ij})) \quad (6)$$

其中, 真实值 $H_{ij} \in (0, 1)$, 并且令 $\hat{H}_{ij} = H_{ij}$ 可得到损失函数的全局最小值, 保证了回归方向的正确性, β 为正负样本权重因子, 用于平衡正负样本的比例 (实验中设 $\beta = 2$)。该损失函数不仅保留了 Focal Loss 可以优化正负样本不平衡问题的能力, 同时使得定位分支的预测值既能够定位候选框中心, 又能反映候选框预测质量, 使得预测结果可以获得更高的定位质量。

4 实验结果与分析

4.1 实验数据集

本文在 3 个公开的鱼眼图像行人数据集 (HABBOF, CEPDOF 以及 MW-R) 上进行实验。为了验证上述算法的有效性, 实验采用交叉验证的方式, 选择任意两个数据集作为训练集, 剩下的数据集作为测试集。

4.2 数据增强实验

由于 3 个数据集之间的场景差异较大, 对混合场景使用模式采样增强会出现错误数据模式, 故数据增强算法均在单一数据集上进行。对 MW-R 数据集进行模式采样增强前后, 其数据分布对比如图 5 所示, 可见 EP, RR 和 SR 3 种采样方法都在不同程度上增加了图像中心附近区域的难例样本数量, 缓解了样本图像半径方向分布的不均衡问题, 如图 5 所示; 角度直方图增强算法则改善了样本在图像上的角度分布, 如图 6 所示。

另外本文通过计算数据增强前后的数据信息熵值来定量衡量数据增强对数据集的信息增益, 如表 1 所列, 对计算结果进行分析, 模式采样增强在保证数据集图片总数不变的前提下, 间接地增加了中心附近的行人样本, 因此在半径和角度方向都提高了数据分布信息熵值; 而角度直方图由于是对图像做特定角度旋转, 未增加样本数量, 因此在半径方向无信息熵增益, 而在角度方向有较大增益。综合来看, 二者均增加了数据分布的不确定度, 丰富了数据集的信息含量。

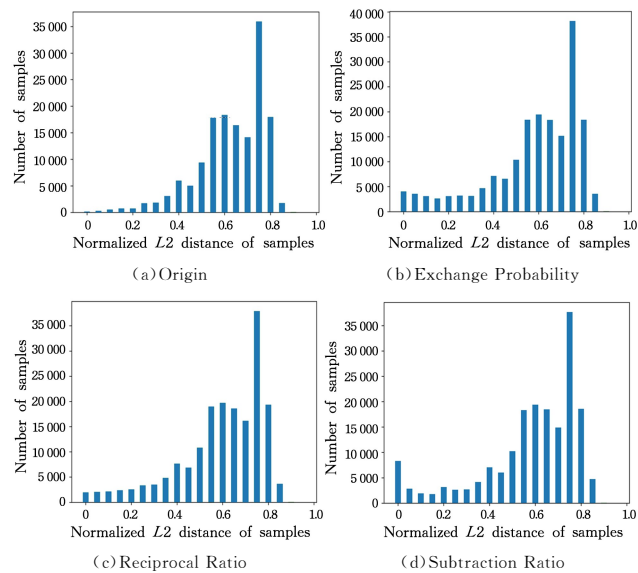


图5 模式采样增强前后样本沿半径分布

Fig. 5 Sample distribution along radius before and after pattern sampling enhancement

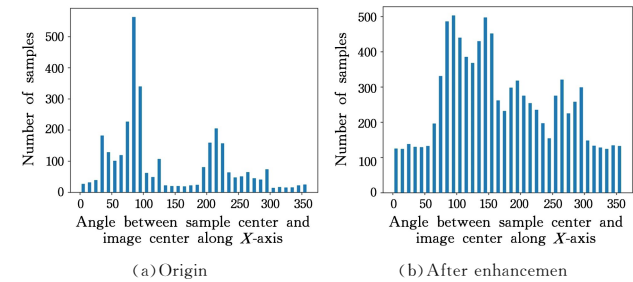


图6 角度直方图增强前后样本角度分布

Fig. 6 Angle distribution of samples before and after angle histogram enhancement

表1 数据增强前后的熵值

Table 1 Entropy before and after data enhancement

	模式采样增强 分布熵值				角度直方图增强 分布熵值	
	Origin	EP	RR	SR	Origin	角度直方图
半径	2.30	2.55	2.54	2.51	2.30	2.30
角度	3.08	3.23	3.19	3.22	3.08	3.48

为了验证增强算法对模型检测能力的提升,对 HABBOF 和 CEPDOF 数据集进行数据增强,使用增强后的数据集分别训练基于锚框(Anchor-Based)的鱼眼图像行人检测网络 RAPID 和本文所使用的无锚框(Anchor-Free)网络 TTFNet,并在 MW-R 数据集上进行测试,使用平均准确率(Average Precision,AP)作为评价指标。

模式采样增强实验中,为了体现该方法对鱼眼图像中心附近行人检测能力的效果,对 MW-R 图像进行划分,当图像中存在行人边界框,且中心点到鱼眼图像中心的归一化距离 $l \leq 0.3$ 时,将该图划分到 MW-R 0-0.3 中,同理划分 MW-R 0-0.5 和 MW-R 0-0.7。TTFNet 与 RAPID 分别采用 ResNet18 和 Darknet53 作为骨干网络,训练图像和测试图像分辨率均为 512×512 。训练策略上,使用随机梯度下降法更新网络参数。其中 Darknet53 基础学习率为 0.001875,动量为 0.9;ResNet19 基础学习率为 0.002,动量为 0.9。预训练阶段采取骨干网络学习率和检测网络学习率 10:1 迭代 82 000 次。

微调阶段先冻结骨干网络迭代 8 000 次,再采取骨干网络学习与检测网络学习 10:1 迭代 3 000 次,得到最佳表现结果。模式采样增强的实验结果表 2 所列。

表2 模式采样增强结果

Table 2 Results of patten sampling enhancement

模型框架	方法	MW-R <i>mAP</i> /%	MW-R <i>AP</i> ₅₀ /%	<i>AP</i> ₇₅ /%		
				MW-R 0-0.3	MW-R 0-0.5	MW-R 0-0.7
无锚框	Origin	49.91	94.26	24.05	43.31	47.88
	EP	49.47	94.0	27.06	43.05	48.24
	RR	50.27	94.13	26.44	43.83	49.92
	SR	50.04	93.57	24.66	41.71	47.49
带锚框	Origin	51.43	95.45	25.41	43.89	49.03
	EP	53.22	95.84	30.81	48.66	54.14
	RR	51.73	95.22	27.36	45.31	53.44
	SR	52.1	95.66	30.53	46.54	52.12

从表 2 无锚框 MW-R 0-0.3,MW-R 0-0.5 以及 MW-R 0-0.7 的实验结果可以看到,EP,RR,SR 3 种方法均能够提升距离鱼眼图像中心较近目标的检测能力;从带锚框实验结果可知,EP 方法能够在提升距离鱼眼图像中心较近目标的检测能力的同时提升在 MW-R 数据集上的整体检测性能。

在角度增强中,测试集为原始的 MW-R 数据,输入图像分辨率为 512×512 ,训练策略与模式采样实验中相同。实验结果如表 3 所列。在无锚框框架中使用角度直方图增强后的数据损害了模型性能,分析原因为:无锚框网络边界框回归是基于真实边界框中的点。对原始图像进行旋转后,其依然是基于真实边界框中的点进行回归,因此边界框回归具有旋转不变性。此外,角度直方图增强对原始图像的旋转角度变动较小,训练数据的大量增加使得网络在角度直方图增强后的数据上产生了局部过拟合,导致性能降低。而基于锚框检测网络框架中是对预设锚框进行变换,此时边界框回归并非是非旋转不变的,因此能够取得较优的检测结果。

表3 角度直方图增强结果

Table 3 Results of angle histogram enhancement

模型框架	方法	MW-R <i>mAP</i> /%	MW-R <i>AP</i> ₅₀ /%	MW-R <i>AP</i> ₇₅ /%
无锚框	Origin	49.91	94.26	46.55
	角度直方图	48.89	92.30	46.28
带锚框	Origin	51.43	95.45	48.62
	角度直方图	53.07	95.54	54.0

4.3 连续 Focal Loss 消融实验

为了验证连续 Focal Loss 在优化无锚框网络监督信息分配中的作用,本文使用与数据增强实验相同配置的 TTFNet 进行消融实验。由表 4 可知,在监督信息中加入预测框质量度量(IoU)并使用离散 Focal Loss 作为损失对其检测性能并无明显提升,因为离散 Focal Loss 将高斯分布峰值之外所有的点都视为负样本,预测框质量度量信息被离散 Focal Loss 权重抑制,故对网络优化无明显作用。单独使用推广的连续 Focal Loss 因为将高斯核内部所有点均视作正样本计算损失,进一步缓解了正负样本不平衡问题;将连续 Focal Loss 与预测框 IoU 监督信息结合使用可获得最佳性能表现,其综合性能(*mAP*)提升约 3.8%,检测准确率(*AP*₅₀)可达 97.2%,优于 Anchor-Based 结构网络 RAPID。且由于连续 Focal Loss 结合 IoU 优化监督只作用在网络训练阶段,对网络推理阶段无影响,故单张图片检测速度(*fps*)远优于 RAPID。由此可见,本文所提方法在保证 Anchor-Free 检测器高实时性

的前提下,可以有效缓解检测器训练监督信息分配不均衡的问题,进一步提升了其检测精度。

表 4 连续 Focal Loss 消融实验结果

方法	MW-R mAP/%	MW-R AP ₅₀ /%	MW-R AP ₇₅ /%	<i>fps</i>
Anchor-Based(RAPiD)	51.43	95.45	48.62	17.1
Origin(TTFNet)				
Focal Loss	49.91	94.26	46.55	
TTFNet				
Focal Loss+IoU 监督	49.80	93.96	47.21	
TTFNet				66.3
连续 Focal Loss	51.53	95.47	48.33	
TTFNet				
连续 Focal Loss+IoU 监督	53.71	97.22	51.62	

结束语 针对鱼眼图像行人检测存在的数据分布不均衡问题,本文提出了一种针对鱼眼图像的两阶段的数据增强方法,其中模式采样增强丰富了样本模式,生成的新样本让网络专注于难例学习;角度直方图增强基于平滑样本角度分布,以提升检测模型对复杂多变场景的鲁棒性。此外本文将 Focal Loss 推广到连续域,结合定位质量监督信息用于网络优化。实验结果表明,本文算法可有效缓解鱼眼图像行人检测中的数据和训练中的不均衡问题,进一步提升了无锚框行人检测器的检测性能。当然,本文方法在模式采样增强和角度直方图增强上的反馈支撑还存在不足,今后将依据鱼眼图像成像机理对其进行进一步研究并做出改进。

参 考 文 献

[1] KUBO Y, KITAGUCHI T, YAMAGUCHI J. Human tracking using fisheye images[C]//SICE Annual Conference. 2007:2013-2017.

[2] MEINEL L, FINDEISEN M, HES M, et al. Automated real-time surveillance for ambient assisted living using an omnidirectional camera[C]//IEEE International Conference on Consumer Electronics. IEEE, 2014.

[3] DEMIRKUS M, WANG L, ESCHEY M, et al. People detection in fish-eye top-views[C]//VISIGRAPP. 2017:141-148.

[4] SOLICHIN A, HARJOKO A. Deteksi Pejalan Kaki pada Video dengan Metode Fastest Pedestrian Detector in the West(FPDW) [J]Journal of TICOM, 2013, 2(1):21-37.

[5] KRAMS O, KIRYATI N. People detection in top-view fisheye imaging[C]//2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS). IEEE, 2017.

[6] QIAN Y, MING Y, WANG C, et al. Self-adapting part-based pedestrian detection using a fish-eye camera[C]//2017 IEEE Intelligent Vehicles Symposium(IV). IEEE, 2017:779-788.

[7] DUAN Z, TEZCAN M O, NAKAMURA H, et al. RAPiD: Rotation-Aware People Detection in Overhead Fisheye Images[J]. arXiv:2005.11623, 2020.

[8] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv:1804.02767, 2018.

[9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[C]//IEEE Transactions on Pattern Analysis & Machine Intelligence. IEEE, 2017:2999-3007.

[10] REN S, HE K, 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.

[11] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for Fine-grained Category Detection[C]//European Conference on Computer Vision. 2014:2700-2709.

[12] GIRSHICK R. Fast r-cnn[J]. arXiv:1504.08083, 2015.

[13] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//IEEE Transactions on Pattern Analysis&Machine Intelligence. IEEE, 2017.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition. IEEE, 2016.

[15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision. Cham: Springer, 2016.

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

[17] ZHOU X, WANG D, KRHENBÜHL P. Objects as Points[J]. arXiv:1904.97850, 2019.

[18] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020:516-520.

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

[20] LIU Z, ZHENG T, XU G, et al. Training-Time-Friendly Network for Real-Time Object Detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020:11685-11692.

[21] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv:1409.1556, 2014.

[22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.

[23] 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.

[24] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv:2004.10934, 2020.

[25] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond Empirical Risk Minimization[J]. arXiv:1710.09412, 2017.

[26] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]//Advances in Neural Information Processing Systems. 2014:2672-2680.



SI Shao-feng, born in 1996, postgraduate. His main research interests include computer vision and object tracking.



LI Qing, born in 1972, Ph.D, professor, Ph.D supervisor. His main research interests include artificial intelligence and multi-source information fusion.