

远红外车载图像实时行人检测与自适应实例分割

于博, 马书浩, 李红艳, 李春庚*, 安居白

大连海事大学信息科学技术学院, 辽宁 大连 116026

摘要 针对红外图像检测与分割任务中颜色信息缺失, 特征细节模糊并带有噪声, 当目标数量较多时传统方法提取过程速度较慢等问题, 提出一种用于远红外图像的优化 YOLO 检测与分割网络模型。提出的两个优化点分别为: 综合分析实验使用的两种远红外数据集后使用 K-means++ 聚类算法寻找多尺度预测标记锚点框尺寸; 使用局部检测位置自适应阈值分割方法对检测目标进行像素级实例分割。本文优化算法在 FLIR 公开数据集与本文数据集中的检测速度分别为 29 frame/s 与 28 frame/s, 保证了实时输出的要求; 行人检测准确率分别达到 75.3% 与 77.6%, 分割结果平均交并比达到 70%~90%。实验结果表明, 本文算法具有良好的稳健性和普适性, 在远红外图像中可快速有效地检测行人并生成实例掩模。

关键词 机器视觉; 远红外图像; 行人检测; 实例分割; 聚类算法; 实时输出

中图分类号 TP183

文献标志码 A

doi: 10.3788/LOP57.021507

Real-Time Pedestrian Detection for Far-Infrared Vehicle Images and Adaptive Instance Segmentation

Yu Bo, Ma Shuhao, Li Hongyan, Li Chungeng*, An Jubai

School of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China

Abstract In infrared image detection and segmentation tasks, the color information is lost, the features are fuzzy with noise, the target number is large, and the traditional extraction method is slow. Therefore, we propose an optimized YOLO detection and segmentation network model for far-infrared images. The two proposed optimization points are as follows. We use the K-means++ clustering algorithm to determine the multi-scale prediction anchor size after the analysis of two far-infrared databases. We also perform pixel-level instance segmentation of detection targets using localized adaptive threshold segmentation. The experimental results show that the proposed algorithm performs pedestrian detection at detection speeds of 29 frame/s and 28 frame/s on the FLIR dataset and the dataset used in this paper, respectively, ensuring the requirement of real-time output. The pedestrian detection accuracies in these datasets reach 75.3% and 77.6%. Moreover, the average intersection over the union of the segmentation results is 70%–90%. In summary, the algorithm performs well with respect to robustness and universality. The algorithm provides a valuable reference method for pedestrian detection and segmentation in far-infrared fields.

Key words machine vision; far infrared image; pedestrian detection; instance segmentation; clustering algorithm; real-time output

OCIS codes 150.0155; 150.1135; 100.4996; 100.4999

1 引言

多目标行人检测与分割是计算机视觉领域中一项非常重要的研究工作, 目的是在真实场景图像中快速有效地识别出多个行人目标与其所在位置, 在

自动与辅助驾驶^[1]、智能安防^[2]、灾难救援等^[3]领域都具有重要的应用价值。当前大部分工作集于分析判断可见光图像, 可见光成像受复杂的背景、高要求的光照条件、较短的夜间可视距离等影响, 其检测结果稳定性欠佳、精准度不足。而远红外成像受上述

收稿日期: 2019-04-17; 修回日期: 2019-05-28; 录用日期: 2019-07-09

基金项目: 国家自然科学基金(61471079)

* E-mail: li_chungeng@dlnu.edu.cn

条件的影响较小,是一种被动式非接触感光技术^[4-5],主要通过物体本身的红外辐射强度呈现图像,温度越高的物体辐射强度越强,所在像素点的亮度就越高。该技术不受光照条件等自然状况的影响,稳定性较高,探测能力较强,作用距离较远。同时其成像原理依据目标的主动红外辐射,与外界光照无关,因此可在完全无光或恶劣天气情况下实行全天候行人检测。由于远红外呈现的图像对比度较低,且实际应用中受硬件价格及成本影响,远红外成像采用廉价镜头呈现的图像分辨率不高,细节特征不强,因此传统方法较难在该类图像中检测出目标^[6-9]。如何在远红外图像上快速有效地检测与分割出多个行人目标成为了一个极具挑战的研究任务。

当前国内外学者也对使用远红外图像检测行人进行了广泛研究。Bertozzi等^[10]提出使用特定尺寸和不同纵横比的多分辨率图像对目标进行定位,然后使用特定滤波器过滤掉道路设施等与行人无关的假正例,最后使用人类形态模板与热成像特征进行对应匹配。此方法需要行人完整出现在图像中且形状特征需完整,当许多行人同时出现时要求相互间不能遮挡,这在实际中难以达到理想要求。Zhang等^[11]在分析可见光图像的检测识别任务后,发现两类图像共享部分特征,因此提出将两个特征类[边缘与方向梯度直方图(HOG)特征]和两个分类模型[AdaBoost和级联支持向量机(SVM)]迁移到远红外图像中,获得了与可见光图像大致相当的准确率,但算法在运行效率上比较低。Suard等^[12]在文献^[11]的基础上使用HOG提取图像的特征描述子,并将其和SVM联合起来共同评价检测效率,在调整完特征描述子和分类器后,集成到一个完整的系统中来处理立体红外图像,算法需要输入较小的图像来获得良好的检测结果。Ge等^[13]提出将三个模块(感兴趣区域生成、目标分类和跟踪)放入级联器中,每个模块利用互补的视觉特征来判别杂乱的背景,使用双阈值分割算法生成感兴趣区域(ROI)并根据角点及梯度方向特征在不同尺寸的不相交子集图像中训练分类器,最终得到一种树状的两级检测器用于解决行人检测问题,但此方法效率较低。马也等^[14]首先采用基于改进的混合高斯模型背景减除法对人体目标进行分割,通过多个带有权值的高斯函数来描述复杂变化的背景,对模型个数、权值、学习率进行更新,然后对分割得到的ROI采用融合边缘方向累加和特性的HOG进行特征描述,利用

SVM实现对人体目标分类检测,该算法在行人重叠和遮挡方面表现不足。刘峰等^[15]提出一种基于多特征级联的红外行人检测方法,利用ROI长宽比特征和头部Haar特征组成初级分类器;改进了HOG特征提取算法和自适应缩放因子获取算法,并基于改进算法完成最终行人识别,有效减少了帧间处理时间;针对目标被遮挡情况,提出遮挡情况判断和局部特征识别功能,该功能在多人遮挡的情况下只会检测出单一人,会造成部分漏检。苏晓倩等^[16]采用基于区域估计的帧差法实现车载红外图像中的行人检测,再通过行人的几何特征和运动速率特征进行行人识别,最后运用轮廓矩匹配算法对行人进行跟踪,但文中未提及方法的实时性情况。

近几年来,深度学习在远红外图像目标检测上的表现逐渐大放光彩。车凯等^[17]提出一种改进的快速区域卷积神经网络(Fast R-CNN)红外图像行人检测方法,并结合红外图像的特点提出一种自适应ROI提取算法。该算法在不影响检测准确率的前提下,降低了ROI数量,减小了网络的计算量;使用一种加权锚点框的定位机制,对基于3种不同宽高比锚点框的检测置信度进行坐标加权,获得了更准确的定位框。许茗等^[18]对热红外图像进行基于频域的显著性检测,生成对行人目标全覆盖的显著图;然后结合热红外原图像生成ROI图作为输入,以行人目标概率图为输出,搭建全卷积网络;最后对热红外行人检测系统进行端对端训练,获取网络输出的行人目标概率图,进而实现行人目标检测。

本文使用优化的YOLO v3深度网络对远红外图像行人目标进行检测,使用改进的聚类方法得到远红外图像行人的基础预测框比例,在检测的结果中使用自适应阈值分割方法对目标进行实例分割,得到像素级的分割掩模。实验结果表明,本算法能快速有效地检测到行人目标,根据动态调整的阈值滑窗大小可获得良好的像素分割掩模,在精确度和效率上较先前算法均有不同程度的提升。

2 基本原理

2.1 YOLO v3 网络简介

目前,深度学习检测识别网络大致可分为两类:以Fast R-CNN^[19]和Faster R-CNN^[20]为代表的“建议框生成”与“目标检测识别”结合的两步处理方法;以SSD^[21]和YOLO^[22-24]为代表的“建议框与检测识别”一步处理方法。两步处理方法精度较高,但检测

效率差,一步处理方法检测效率高但精度较低。属于一步处理方法的YOLO v3经过多个版本迭代后可保证在较高精度的情况下提升检测速度,其性能获得大幅提升。

YOLO v3的核心思想为:将输入图片固定在相同尺度上,经过特征提取网络后生成特征图;之后对

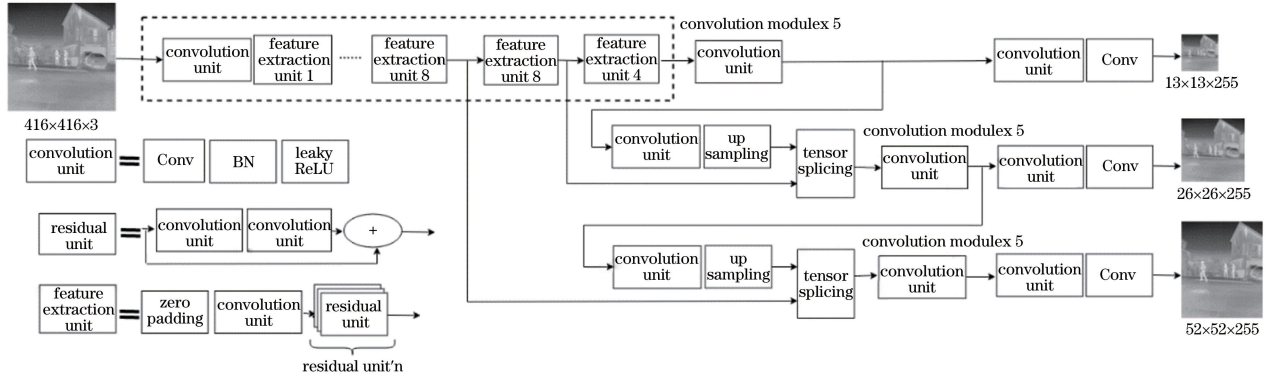


图1 YOLO v3结构示意图

Fig. 1 Structural diagram of YOLO v3

该版本算法的一大亮点是使用了经上采样融合拼接的多尺度特征图,在三种不同大小的特征图上共同预测目标的位置和类别信息,提高了对小目标的检测性能。在图1中,使用 13×13 大小的特征图预测,之后将该层特征图上采样至 26×26 ,与前端相同尺度大小的特征图融合拼接后进行第二次预测,最后将 26×26 大小的特征图上采样至 52×52 ,同前端相同大小特征图整理拼接后进行最后的预测。得益于这种多尺度检测的设计思想,算法对小目标的检测精度有了大幅度提升。算法采用回归偏移值的方法获得预测框坐标及大小,即

$$b_x = \sigma(t_x) + c_x, \quad (1)$$

$$b_y = \sigma(t_y) + c_y, \quad (2)$$

$$b_w = p_w \exp(t_w), \quad (3)$$

$$b_h = p_h \exp(t_h), \quad (4)$$

$$\sigma(t_0) = P_{\text{object}} \times I_{\text{IOU}}, \quad (5)$$

式中: t_x, t_y, t_w, t_h 为网络需要学习的目标偏移值; p_w, p_h 为预设锚点框的宽和高; c_x, c_y 为对应特征图左上角特征点的坐标值; $\sigma(t_0)$ 为目标预测概率值和当前预测框与真实框交并比(IOU)的乘积; b_x, b_y, b_w, b_h 分别为预测框的坐标及宽高值。该步骤得到等同于特征图大小的目标框的预测位置,经过缩放处理最终得到等同于输入图像大小的目标框的最终位置。

2.2 改进聚类的基础预测框

在预测框输出阶段,为更加精准地生成与真实框大小大致相同的形状,算法使用锚点机制的固定

固定的三个特征层进行上采样融合操作,并基于这三种特征层上的特征像素生成建议框位置、置信度和每个类别的概率值;经过计算代价函数和反向传播后来训练网络,最终经过宽高压缩比例转换得到与输入图像同等大小的检测框与多标签分类类别。整体网络结构如图1所示。

框作为预测框的基本比较对象,继而求得偏移量,使用K-means聚类算法^[25]来寻找符合要求的锚点框。K-means算法初始随机选取 N 个聚类中心点,部分情况下得到的结果并不完全是整个数据中最具代表性的聚类中心,因此本文提出使用改进的K-means++聚类算法^[26]来生成初始的9种锚点框。算法的大致过程如下。

1) 从当前的 M 个聚类数据中随机选取一个初始聚类中心 k_1 。

2) 计算其余样本点 x_m 到当前所有聚类中心 k_n 的最短距离,用 $D_{\text{dist}}(x_m)$ 表示。

3) 计算其余样本点被选为下一个聚类中心点的概率值,即

$$p(x_m) = \frac{D_{\text{dist}}(x_m)^2}{\sum_{m \in M} D_{\text{dist}}(x_m)^2}. \quad (6)$$

4) 将得到的概率值映射到 $(0, 1)$ 轴上的 $(M-1)$ 个左闭右开的间断段,每个段长对应第 m 个值的概率值,然后随机生成一个 $0 \sim 1$ 间的随机数,该点所在的间断段即为被选择的下一个聚类中心 k_2 。

5) 重复2)~4),直至选出 N 个聚类中心 $k_1 \sim k_n$ 。

6) 针对数据集中剩余的 $(M-N)$ 个样本,计算其到 N 个聚类中心的距离并将它们归至距离最短的聚类中心所属类中。

7) 求得所有类别的最小平方误差 E ,针对每个类别 k_n ,计算当前属于该类所有点的聚类中心,其

中 ξ_i 为当前类别的质心。

$$E = \sum_{i=1}^n \sum_{x \in k_i} \|x - \xi_i\|_2^2, \quad (7)$$

$$\xi_i = \frac{1}{|k_i|} \sum_{x \in k_i} x. \quad (8)$$

8) 重复 6)~7), 直至所有得到的聚类中心不再发生变化。

表 1 为算法默认锚点框大小和改进算法在

FLIR 数据集与自采集数据集上分别得到的 9 种有效初始锚点形状, 其中 s 为宽高长, r 为宽高比。经过数据分析比对后发现对于行人的检测, 宽高比为 0.30 的锚点框具有极强的代表性, 同时也符合图像中的人体比例。最终结合两种数据集的目标大小, 选择 9 种尺度比例在 0.30 左右的预测框作为网络的基本锚点框, 从而提升了检测精度, 结果如表 1 中粗体数字所示。

表 1 九种锚点框大小比例
Table 1 Nine anchor size ratios

No.		①	②	③	④	⑤	⑥	⑦	⑧	⑨
Default	s	(10, 13)	(16, 30)	(33, 23)	(30, 61)	(62, 45)	(59, 119)	(116, 90)	(156, 198)	(373, 326)
	r	0.77	0.53	1.43	0.49	1.38	0.50	1.29	0.79	1.14
FLIR dataset	s	(3, 11)	(5, 20)	(7, 14)	(8, 23)	(10, 34)	(11, 49)	(16, 52)	(23, 78)	(41, 140)
	r	0.27	0.26	0.53	0.36	0.30	0.22	0.32	0.30	0.30
Our dataset	s	(9, 33)	(13, 45)	(15, 61)	(19, 86)	(21, 69)	(26, 101)	(32, 133)	(44, 178)	(70, 271)
	r	0.28	0.29	0.26	0.22	0.30	0.26	0.24	0.25	0.26
Final choose	s	(5, 20)	(10, 34)	(13, 45)	(16, 52)	(21, 69)	(23, 78)	(32, 133)	(41, 140)	(70, 271)
	r	0.26	0.30	0.29	0.32	0.30	0.30	0.24	0.30	0.26

图 2 为默认锚点框在三种尺度上的可视化表达, 图 3 为本文锚点框的可视化表达。可以看出本

文选取的锚点框符合行人的外观轮廓特征, 使得网络的检测框回归定位有了更加精准的参照标准。

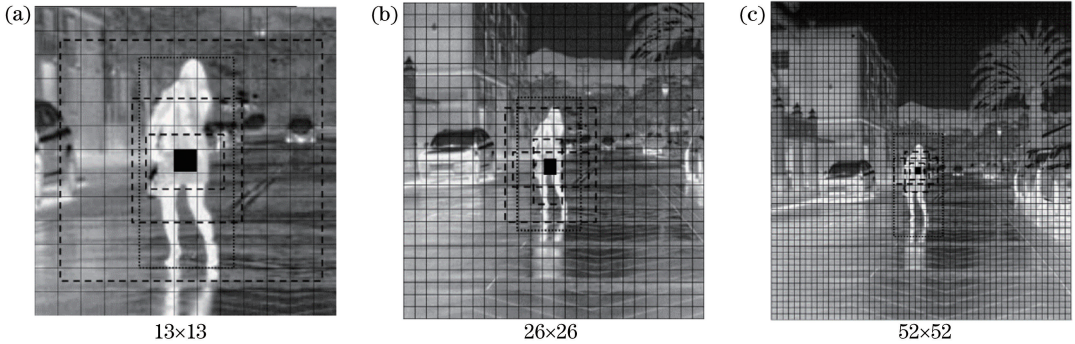


图 2 默认锚点框的可视化表达

Fig. 2 Visual representation of default anchor box

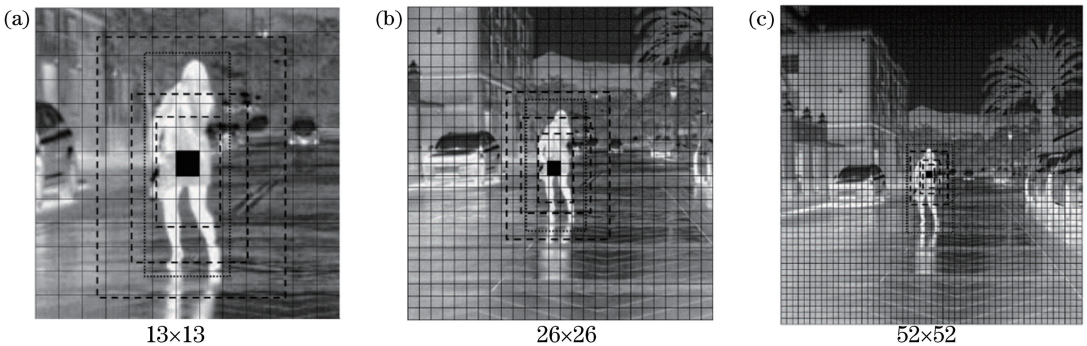


图 3 本文锚点框的可视化表达

Fig. 3 Visual representation of our anchor box

2.3 目标区域的自适应实例分割

实例分割是语义分割与目标检测结合而生的技

术, 其目的是针对每一个实体目标, 找到目标具体对应的像素点并与其他同类目标相互区别, 是计算机

视觉中具有挑战性的任务之一。其生成的掩模像素在自动驾驶、辅助驾驶中具有重要的判别定位作用。目前使用全卷积网络对目标图像生成实例分割掩模是较常用的手段,此方法的应用前提是需要人工大量标注训练图像中的目标掩模,使用监督训练的方式得到具有分辨能力的深度网络模型,此方法的前期工作量较大,且标记图像的自动化水平较低。本文通过分析远红外图像独有的温度成像特征性质,提出自适应的目标实例分割方法,该方法可有效地在最终目标上生成实例掩模,同时代替了大部分的人工实例标注工作,提高了后续使用监督方式训练的工作效率。

算法首先读入当前检测图像,生成单通道灰度图像,利用检测器得到的坐标位置及宽高长度在灰度图中对处理目标进行定位,如图4所示。远红外图像的一个显著特点是温度较高的地方像素值越高,即需要检测的人体目标较局部背景区域像素值



图4 目标区域定位图

Fig. 4 Target area location map

偏高,因此选择一个合适的分割阈值可将目标与局部背景分离。将截取的灰度图像转换为图像灰度直方图,如图5所示。将目标提取为独立图像并使用高斯滤波对局部区域进行去噪,结果如图6所示。其方差是灰度分布均匀性的一种衡量方式,目标与背景类间方差越大,这两部分在图像中呈现方式便越明显,任何背景被分为目标或目标被分为背景的操作都会使得这两类的方差增大,最终可找到局部区域分割的最佳阈值。

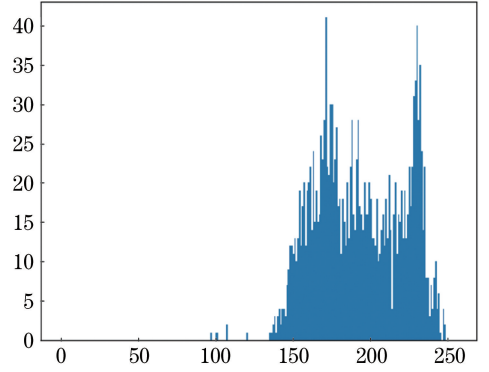


图5 行人目标图像灰度直方图

Fig. 5 Gray histogram of pedestrian object image



图6 去噪图像

Fig. 6 Denoising image

$$\omega_1 = \frac{m'n'_0 + m'n'_1 + \dots + m'n'_t}{M' \times N'}, \quad (9)$$

$$\omega_2 = \frac{m'n'_{t+1} + m'n'_{t+2} + \dots + m'n'_{255}}{M' \times N'}, \quad (10)$$

$$\mu_1 = \frac{0 \times m'n'_0 + 1 \times m'n'_1 + \dots + t \times m'n'_t}{m'n'_0 + m'n'_1 + \dots + m'n'_t}, \quad (11)$$

$$\mu_2 = \frac{(t+1) \times m'n'_{t+1} + (t+2) \times m'n'_{t+2} + \dots + 255 \times m'n'_{255}}{m'n'_{t+1} + m'n'_{t+2} + \dots + m'n'_{255}}, \quad (12)$$

$$m'n'_0 + m'n'_1 + \dots + m'n'_{255} = M' \times N', \quad (13)$$

$$\omega_1 + \omega_2 = 1, \quad (14)$$

$$\mu_1 \omega_1 + \mu_2 \omega_2 = \mu, \quad (15)$$

$$(\mu - \mu_1)^2 \omega_1 + (\mu - \mu_2)^2 \omega_2 = g, \quad (16)$$

式中: $0 \sim 255$ 为像素大小; t 为当前分割阈值大小, $0 \sim t$ 为背景像素值, $(t+1) \sim 255$ 为前景像素值;

$M' \times N'$ 为图像大小; $m'n'_0, m'n'_1, \dots, m'n'_{255}$ 为像素值对应的像素点个数; ω_1 为背景像素所占比例,

ω_2 为前景像素所占比例; μ_1 为背景平均灰度, μ_2 为前景平均灰度, μ 为图像总平均灰度; g 为方差。在遍历所有 t 值后选择最大的 g 值所对应的 t 值作为最佳分割阈值。本例中, 最终分割阈值为 195, 处在直方图波谷区域, 分割图像如图 7 所示。



图 7 阈值分割图像

Fig. 7 Threshold segmentation image

上述为大津法 (OTSU) 的阈值选取过程^[27], 由于数据集的图像分辨率及清晰度存在差异, 仅使用大津法分割会使得人物掩模存在较多空洞及边界模糊, 因此在其基础上使用自适应动态滤波器对分割图像进行闭运算操作, 来填充目标前景中的细小空洞, 平滑边界。在 FLIR 公共数据集中, 行人细节较明显, 图像分辨率宽为 640 pixel, 使用 3×3 大小的闭运算滤波器进行处理效果较好; 本文自行采集的数据集行人细节较为模糊, 摄像头分辨率宽为 160 pixel, 使用 8×8 大小的闭运算滤波器进行处理效果较为理想。根据上述信息建立自适应线性关系, 即

$$y = -\frac{1}{96}x + \frac{29}{3} (y \geq 1), \quad (17)$$

式中: x 为输入图像宽边分辨率; y 为自适应的闭运算滤波器大小 (临近取整)。处理结果如图 8 所示。



图 8 闭运算图像

Fig. 8 Closed operation image

最后在闭运算图像中添加像素点访问属性, 遍历二值图像素, 将找到的第一个非 0 像素点访问标记设置为真; 并遍历该像素点的 8 邻域像素点, 将非 0 像素点压入连通栈; 继续遍历栈顶元素, 被遍历的像素点访问标记均设为真; 遍历结束后将栈顶元素压出, 当连通栈为空时, 一个连通域遍历完成, 并记录该栈最大元素个数。重复上述步骤, 直至所有非

0 像素点遍历完成, 像素点数量最多的栈即为找到的前景目标像素, 最终完成对远红外图像的实例分割, 结果如图 9 所示。



图 9 连通区掩模

Fig. 9 Connected area mask

3 实验准备、过程与结果分析

3.1 实验数据准备及预操作设置

为验证方法的有效性, 本文使用全球最大热成像厂商 FLIR 公司 2018 年 8 月发布的自动驾驶热成像数据集, 同时本文作者在研究过程中也购买了热成像摄像机并制作了数据集, 以验证网络对不同远红外图像的通用性。

FLIR 热成像数据集提供带有注释的远红外图像和无注释的 RGB 图像集, 用于目标检测网络的训练与测试。通过安装在车辆上的热成像摄像机和可见光相机获取数据集。该数据集总共包含 14452 张带注释的远红外图像, 其中 10228 张图像来自短视频, 4224 张图像来自连续的 144 s 视频。所有视频均于 11 月至 5 月在美国加利福尼亚州圣巴巴拉的道路和高速公路上拍摄, 拍摄环境为晴天, 拍摄时间为白天和夜晚。用 FLIR Tau2 (13 mm $f/1.0$, 45° HFOV 和 37° VFOV) 获得热图像。摄像机位于一个相距 1.9 ± 0.1 inch (2.54 cm) 的机箱内。使用 FLIR 专有软件通过 USB3 视频捕获图像, 图像大小为 512 pixel \times 640 pixel。大多数 10228 张热图像以每秒两张的速率被采样, 本文只使用数据集中的行人类别且对其中的训练与测试集合进行了重新分配。

本文数据集提供带有注释的远红外图像, 在汽车内部后视镜后方的位置对场景进行采集, 目前共有 1000 张图像, 均来自于每段连续半小时的视频采集数据, 一共 30 段。所有视频均于 2018 年 5 月在大连海事大学校内及周边的道路上拍摄, 拍摄环境为夜间的晴天与雨天。使用 FLIR One Pro3 (14 mm, 50° HFOV 和 38° VFOV) 热成像摄像机拍摄, 采集帧率为 8.7 Hz, 图像大小为 1080 pixel \times 1440 pixel, 红外镜头分辨率为 160 pixel \times 120 pixel。使用 LabelImg 工具对图像进行标注并自动生成 xml 文件, 标注的类别

为人(person)。本论文发表后我们将会及时地把该数据集在 GitHub 上无偿公开,并作后续补充,以供各位科研人员研究与使用。

本文使用 Nvidia GTX1060 6 G 显存的移动版 GPU 进行实验,处理器型号为 Intel i7-7700HQ,运行内存 8 GB,实验环境为 Ubuntu16.04 + Cuda9.0+Cudnn7.0.5+Pytorch0.4.1。实验对比方法均为开源代码。

3.2 实验过程及结果

图 10 为算法在 FLIR 数据集得到的实例分割结果与人工标注结果对比,图 11 为算法在本文数据集得到的实例分割结果与人工标注结果对比。为有效地准确评判分割所得结果,上述两种对比实验的分割结果均基于图像真值标定框。图 12 为算法最终的检测分类及分割结果。结果展示均为未经训练的测试集图像。



图 10 FLIR 数据集实例分割效果图。(a)实例分割结果;(b)人工标注结果

Fig. 10 Instance segmentation effect diagrams for FLIR dataset. (a) Instance segmentation results; (b) manual labeling results



图 11 本文数据集实例分割效果图。(a)实例分割结果;(b)人工标注结果

Fig. 11 Instance segmentation effect diagrams for our dataset. (a) Instance segmentation results; (b) manual labeling results

3.3 实验结果分析

图 10 和图 11 分别展示了适量目标和大量目标情况在算法上的表现,实验均取得了较好的检测分

割结果。图 12 分别展示了几种对比算法的实验效果,可以看出 SVM 得到的边框回归精度差;Faster R-CNN 边框回归较为精准但出现较多漏检;同时

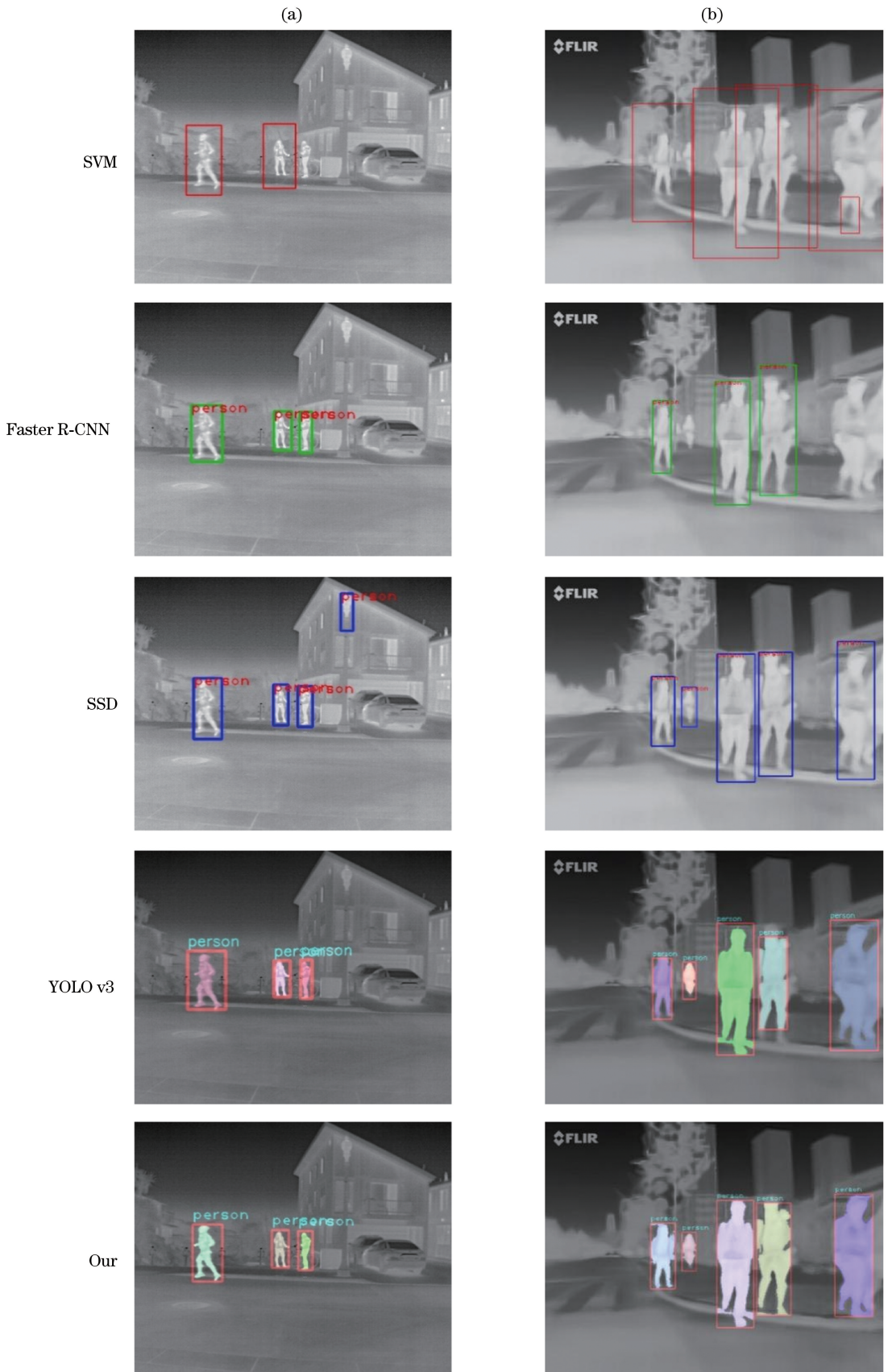


图 12 各算法对比结果。(a) FLIR 数据集;(b)本文数据集

Fig. 12 Comparison of results of algorithms. (a) FLIR dataset; (b) our dataset

SSD 出现了误检;YOLO v3 由于使用默认锚点框大小,因此对行人的细节定位欠缺,在本文数据集中出现了对行人的头部及脚部定位不准的问题;本文算法使用了针对行人的锚点框设计,因此定位准确度有了明显提高。

针对检测分类结果,本文使用 Pascal VOC 的准确率-召回率方式进行评测。设 Q 为测试样本目标真值总数量,首先将其按置信度值从高至低排序,并匹配对应的真值标签。从最高分开始,若当前预测标定框与真值标定框的交并比大于预设阈值(本文标准值为 0.5),则记为正确预测(TP);若低于阈值,则记为错误预测(FP);其余未被预测到的真值框为 $Q-TP$ (FN)。

$$P_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (18)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}. \quad (19)$$

根据上述两个指标可以构建出准确率-召回率曲线,即 $P-R$ 曲线。根据最新计算法则,对每个召回率值对应的最大准确率值求和后取平均即为当前类别的平均准确率(AP)。图 13 和 14 是本算法与其他算法在两个数据集上的 $P-R$ 曲线对比,表 2 为算法性能对比评价表,其中的检测速度一项表明本文算法在两种数据集上都超过了 24 frame/s,可实现实时检测功能。

表 2 算法性能评价表

Table 2 Algorithm performance evaluation table

Algorithm	Backbone	FLIR dataset		Our dataset		Instance segmentation
		AP	Velocity/(Frame · s ⁻¹)	AP	Velocity/(Frame · s ⁻¹)	
HOG+SVM(64×128)		63.1	<1	65.2	<1	×
Faster R-CNN(512×512)	VGG16	68.5	6.5	72.9	6	×
SSD(300×300)	VGG16	71.6	16	73.3	15	×
YOLO v3(416×416)	Darknet53	74.4	30	76.1	30	×
Our(416×416)	Darknet53	75.3	29	77.6	28	✓

本文使用 IOU 对实例分割的结果进行评判。IOU 是目标检测中广泛使用的概念,目的是计算产生的候选框与真值标记框的交叠率,即交集与并集的比值,最优情况为完全重叠,此时比值为 1。本文将计算标记框改为计算生成的掩模像素点,针对每个目标,计算算法生成的掩模像素点与人工标记的目标像素点的 IOU 来评价实例分割

的准确度。

$$I_{\text{IOU}} = \frac{A_{\text{area}}(W) \cap A_{\text{area}}(U)}{A_{\text{area}}(W) \cup A_{\text{area}}(U)}, \quad (20)$$

式中: W 为程序得到的目标掩模像素点; U 为人工标记的目标像素点。最终在两种数据集上随机各抽取 50 张和 30 张进行人工标记并测试,图 10 与图 11 的交并比如表 3 所示。

表 3 IOU 评价表

Table 3 IOU evaluation table

	The first line	The second line	The third line	Test result for 50 images
FLIR dataset	82.26%	86.05%	80.7%	75%–90%
Our dataset	86.56%	85.68%	82.89%	70%–90%

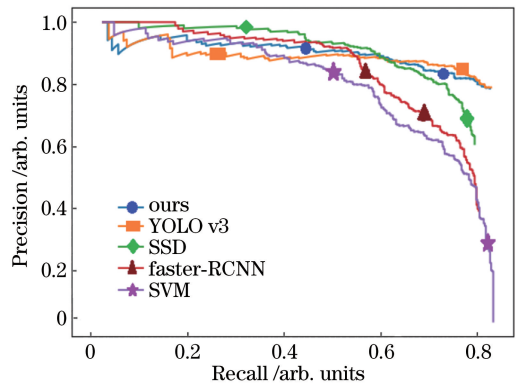


图 13 FLIR 数据集算法的 $P-R$ 曲线

Fig. 13 $P-R$ curves of FLIR dataset algorithm

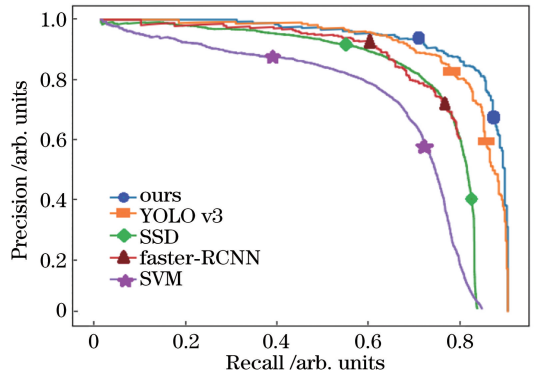


图 14 本文数据集算法的 $P-R$ 曲线

Fig. 14 $P-R$ curves of our dataset algorithm

4 结 论

提出一种红外车载图像行人实时检测与自适应实例分割的方法。基于远红外图像行人目标的特点及属性的研究,使用优化的 K-means++ 算法提取基础锚点框尺寸大小,利用定位框坐标对行人目标进行采样、去噪、最佳阈值分割、连通区域检测等一系列操作,最终获得清晰准确的行人目标掩模图像。实验结果表明,本文算法可有效提取远红外图像行人目标的基础框特征大小,基础框长宽比例符合自然界行人特征,提升了原有算法的精确度。本算法行人目标的实例掩模生成较为精准,且预先无需标注实例图像,有效减少了工作量,平均 IOU 达到 80% 左右,该方法的提出可实现自动化远红外图像实例目标掩模标注,为以后的精准监督训练打下了良好基础。本算法在两种数据集下的检测速度均超过 24 frame/s,达到了实时应用的级别。综上所述,本文算法可有效地对现实场景中的多个行人进行检测与实例分割,稳健性与实用性较强,可应用于自动驾驶、智能安防等多个领域。

后续工作会在此基础上,继续探索车辆等其他目标在远红外图像中的检测与实例分割,并完善本文数据集,优化非目标对分割结果的影响,助力远红外技术在人工智能领域的应用。

参 考 文 献

- [1] Braga-Neto U M, Choudhury M, Goutsias J I. Automatic target detection and tracking in forward-looking infrared image sequences using morphological connected operators [J]. *Journal of Electronic Imaging*, 2004, 13(4): 802-813.
- [2] Corsi C. Infrared: a key technology for security systems [J]. *Advances in Optical Technologies*, 2012, 2012: 838752.
- [3] Cui M Y. Application field and technical characteristics of infrared thermal imager [J]. *China Security & Protection*, 2014(12): 90-93.
崔美玉. 论红外热像仪的应用领域及技术特点 [J]. *中国安防*, 2014(12): 90-93.
- [4] Wang H X, Dong H, Zhou Z Q. Review on dim small target detection technologies in infrared single frame images [J]. *Laser & Optoelectronics Progress*, 2019, 56(8): 080001.
王好贤, 董衡, 周志权. 红外单帧图像弱小目标检测技术综述 [J]. *激光与光电子学进展*, 2019, 56(8): 080001.

- [5] Liu R, Wang D J, Jia P, et al. Overview on small target detection technology in infrared image [J]. *Laser & Optoelectronics Progress*, 2016, 53(5): 050004.
刘让, 王德江, 贾平, 等. 红外图像弱小目标探测技术综述 [J]. *激光与光电子学进展*, 2016, 53(5): 050004.
- [6] Peng Z Y, Wang X J, Lu J. Infrared target detection under hot dome infrared radiation based on visual saliency method [J]. *Infrared and Laser Engineering*, 2014, 43(6): 1772-1776.
彭志勇, 王向军, 卢进. 窗口热辐射下基于视觉显著性的红外目标检测方法 [J]. *红外与激光工程*, 2014, 43(6): 1772-1776.
- [7] Cai W J, Wang L P, Zhang L P. Vehicle detection algorithm based on SLPP-SHOG in infrared image [J]. *Laser & Infrared*, 2016, 46(8): 1018-1022.
蔡文靖, 王鲁平, 张路平. 基于 SLPP-SHOG 的红外图像车辆检测方法 [J]. *激光与红外*, 2016, 46(8): 1018-1022.
- [8] Fan Q S, Fan H B, Lin Y, et al. Multi-object extraction methods based on long-line column scanning for infrared panorama imaging [J]. *Infrared Technology*, 2019, 41(2): 118-126.
范青帅, 范宏波, 林宇, 等. 基于长行列扫描周视红外成像的多目标提取方法综述 [J]. *红外技术*, 2019, 41(2): 118-126.
- [9] Chen Q, Sheng H X, Zhang Z, et al. Moving object detection under infrared light mutation [J]. *Laser & Optoelectronics Progress*, 2016, 53(11): 111005.
陈强, 盛惠兴, 张卓, 等. 红外光照突变下的运动目标检测 [J]. *激光与光电子学进展*, 2016, 53(11): 111005.
- [10] Bertozzi M, Broggi A, Fascioli A, et al. Pedestrian detection for driver assistance using multiresolution infrared vision [J]. *IEEE Transactions on Vehicular Technology*, 2004, 53(6): 1666-1678.
- [11] Zhang L, Wu B, Nevatia R. Pedestrian detection in infrared images based on local shape features [C] // 2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE, 2007: 9738285.
- [12] Suard F, Rakotomamonjy A, Bensch A, et al. Pedestrian detection using infrared images and histograms of oriented gradients [C] // 2006 IEEE Intelligent Vehicles Symposium, June 13-15, 2006, Meguro-Ku, Japan. New York: IEEE, 2006: 206-212.

- [13] Ge J F, Luo Y P, Tei G. Real-time pedestrian detection and tracking at nighttime for driver-assistance systems [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2009, 10 (2): 283-298.
- [14] Ma Y, Chang Q, Hu M F. Research on infrared human detection from complex backgrounds [J]. *Infrared Technology*, 2017, 39 (11): 1038-1044, 1053.
马也, 常青, 胡谋法. 复杂背景下红外人体目标检测算法研究[J]. *红外技术*, 2017, 39(11): 1038-1044, 1053.
- [15] Liu F, Wang S B, Wang X J, et al. Infrared pedestrian detection method in low visibility environment based on multi feature association [J]. *Infrared and Laser Engineering*, 2018, 47 (6): 0604001.
刘峰, 王思博, 王向军, 等. 多特征级联的低能见度环境红外行人检测方法[J]. *红外与激光工程*, 2018, 47(6): 0604001.
- [16] Su X Q, Sun S Y, Ge M, et al. Pedestrian detection and tracking of vehicle infrared images [J]. *Laser & Infrared*, 2012, 42(8): 949-953.
苏晓倩, 孙韶媛, 戈曼, 等. 车载红外图像的行人检测与跟踪技术[J]. *激光与红外*, 2012, 42(8): 949-953.
- [17] Che K, Xiang Z T, Chen Y F, et al. Research on infrared image pedestrian detection based on improved fast R-CNN [J]. *Infrared Technology*, 2018, 40(6): 578-584.
车凯, 向郑涛, 陈宇峰, 等. 基于改进 Fast R-CNN 的红外图像行人检测研究[J]. *红外技术*, 2018, 40 (6): 578-584.
- [18] Xu M, Yu X S, Chen D Y, et al. Pedestrian detection in complex thermal infrared surveillance scene [J]. *Journal of Image and Graphics*, 2018, 23 (12): 1829-1837.
许茗, 于晓升, 陈东岳, 等. 复杂热红外监控场景下行人检测[J]. *中国图象图形学报*, 2018, 23(12): 1829-1837.
- [19] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [20] 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.
- [21] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M] // Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [22] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 26-July 1, 2016, Las Vegas, Nevada. New York: IEEE, 2016: 779-788.
- [23] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 21-26, 2017, Honolulu, Hawaii, USA. New York: IEEE, 2017: 7263-7271.
- [24] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J/OL]. (2018-04-08) [2019-04-16]. <https://arxiv.org/abs/1804.02767>.
- [25] Hartigan J A, Wong M A. Algorithm AS 136: a K -means clustering algorithm [J]. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, 28(1): 100-108.
- [26] Arthur D, Vassilvitskii S. K -means ++: the advantages of careful seeding [C] // *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, January 7-9, 2007, New Orleans, Louisiana. USA: Society for Industrial and Applied Mathematics Philadelphia, 2007: 1027-1035.
- [27] Otsu N. A threshold selection method from gray-level histograms [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62-66.