第 49 卷 第 17 期/2022 年 9 月/中国激光 研究论文

中国激光

多尺度特征融合轻量化夜间红外行人实时检测

何自芬*，陈光晨，陈俊松，张印辉**

昆明理工大学机电工程学院，云南 昆明 650500

摘要 针对辅助驾驶中夜间小目标红外行人检测精度低、网络模型占用内存空间大、检测速度难以满足实时检测要求等问题，提出了一种轻量化的夜间红外图像行人检测神经网络 YOLO-Person。首先提出一种以 MobileNetV3 轻量化网络为骨干网络，以多尺度融合目标检测层为预测模块的网络模型，以解决网络模型大、推理速度慢的问题，大幅减少了模型计算量，初步实现轻量化；然后通过在网络中添加空间金字塔池化模块与更小感受野的检测层，增强网络输出特征图的表征能力，解决数据集中行人目标尺度大小不均衡的问题，提高模型的红外行人检测精度；最后应用通道剪枝对模型进行剪枝，减少特征图的通道数，获得最终网络模型 YOLO-Person。通过 Jetson Nano 移动开发平台，在夜间红外图像行人数据集上验证 YOLO-Person 轻量化模型，结果表明：与 YOLOv3 网络模型相比，提出的 YOLO-Person 网络模型更适于移动端的夜间红外行人检测，平均检测精度达到了 92.2%，检测速度由 26 frame/s 提高到了 69 frame/s，模型大小也由 246 MB 减少到了 11.7 MB。

关键词 成像系统；夜间红外行人检测；多尺度融合；MobileNetV3 网络；模型剪枝

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/CJL202249.1709002

1 引 言

随着科学技术的进步，越来越多的先进技术运用到车辆辅助驾驶中，如超声波测距、毫米波雷达、红外夜视系统及行人检测技术等。夜间行车中的红外图像行人检测技术也逐渐成为行人检测领域的研究热点之一。利用红外热成像技术以及基于深度学习的目标检测算法，对于夜间行车中道路上的行人进行实时检测，可以有效增强夜间行车安全性，避免可能的交通事故、人员伤亡以及财产损失，也有助于缓解驾驶员夜间行车压力。

行人目标检测算法可分为传统的检测方法和基于深度学习的检测方法。在传统的行人检测方法中，Ge 等[1]提出使用双阈值分割方法对近距离行人目标进行分割。Nanda 等[2]提出了一种可对低质量红外图像进行检测的方法，通过引入概率模板适应人体形状的变化，减少搜索空间，且通过了现实场景的测试。Xu 等[3]针对行人检测任务中最常遇到的遮挡问题，提出应用信息更加丰富的人体特征部分对行人进行预测，如头部、肩膀，先通过阈值获取感兴趣的区域，再应用支持向量机(SVM)进行分类。

深度学习是如今发展最快的机器学习方法之一，在辅助驾驶行人检测技术中也有广泛应用[4]。为获取更多重点目标细节信息，目前检测领域常应用注意力机制[5]聚焦重点目标，针对关注点投入更多计算资源。刘学等[6]在 SSD 网络中引入压缩和激励(SE)注意力模块，以解决红外行人图像细节少、特征信息不丰富等问题。赵斌等[7]提出一种基于深度注意力机制的多尺度红外行人检测方法，设计多尺度特征金字塔，引入更低层高分辨率特征图，提高小尺度行人检测性能；应用注意力模块生成基于卷积特征的局部显著图，抑制不相关区域的特征响应，突出图像局部特征。但简单叠加注意力机制会导致网络计算复杂度大幅增加，耗费大量计算资源，使得模型推理速度下降。针对图像目标尺度小和特征信息较少等问题，赵亮等[8]提出一种基于语义分割特征的区域卷积神经网络，用基于语义分割特征的最远点采样算法预测点的语义分割类别，以提高采集关键点中前景点的比例，从而提升算法的检测精度。苗壮等[9]提出了一种基于关键点的快速红外目标检测方法，以目标中心作为检测关键点，设计特征融合网络融合多尺度空间信息和语义信息，实现红外目标快速检测。于博等[10]提出一种远红外图像优化检测与分割网络模型，其应用 K-means++聚类算法寻找多尺度预测标记候选尺寸，并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分割。从关键点或局部区域对目标进行检测的方法对小尺度目标有一定效果，但数据集存在目标尺度不均衡问题时会导致模型检测效果不佳。

收稿日期：2021-11-12；**修回日期**：2021-12-08；**录用日期**：2021-12-27

基金项目：国家自然科学基金(62171206，61761024，62061022)

通信作者：*zyhhzf1998@163.com；**zhangyinhui@kust.edu.cn

本文针对事故频发的夜间行车环境，提出了一种夜间红外图像行人检测模型，能够对夜间路面行人进行实时检测。研究成果可用于辅助驾驶领域，能对驾驶员进行预警提醒或主动刹车，降低夜间行车发生事故的概率，为车辆和行人提供更高的安全保障，具有一定的市场应用前景。

2　端到端的 YOLO 神经网络

Redmon 等[11]提出了端到端的 YOLO 神经网络模型。YOLO 把目标检测看作回归问题，用回归思想进行目标预测，将输入图像划分成 $S\times S$ 个单元格，若物体中心落入某一单元格，则由该单元格负责预测该物体，实现了快速高效的目标检测，使得神经网络算法部署到嵌入式开发平台去解决实际应用问题成为可能。但 YOLO 算法使用较多的降采样层导致上文信息大量丢失，因此其对距离近或尺度小的目标检测精度不够理想。且 YOLO 作为单阶段检测方法，预测的边界框数量较少，使得准确率也较低。

针对 YOLO 算法存在的小尺度、近距离目标检测效果差以及边界框数量少等问题，Redmon 等[12]在后续改进中采用更精简的 DarkNet19 主干网络，同时应用锚点框替代回归框，使得 YOLOv2 算法在多种类预测、检测精度和定位精度方面都有较大提升。

2018 年 Redmon 等[13]又提出了各方面都较为均衡的 YOLOv3 网络，通过借鉴残差网络与特征金字塔网络(FPN)提出了具有更深层结构的 DarkNet53 骨干网络，在预测阶段引入多尺度特征融合模块，大幅提高了模型的小目标预测能力。2020 年提出的 YOLOv4[14]与 YOLOv3 网络相比，仅添加主流优化策略，使最终模型提高了平均检测精度。在具体实施中 YOLOv4 网络在主干网络、损失函数、网络训练以及数据处理等方面都做了优化，使其检测精度有了一定的提高，但在检测速度以及模型权重大小两方面反而不及 YOLOv3 网络。要在车载平台上实现行人实时检测，对检测算法的检测精度、检测速度以及模型大小均提出了更高的要求。因此，本文基于单阶段检测方法进行优化以满足边缘设备部署的相关需求。

3　模型构建

目前小型神经网络的获取主要有两个途径：(1)先使用复杂模型训练，再对训练好的复杂网络模型进行瘦身，从而得到一个小模型；(2)直接对复杂神经网络模型进行优化设计，从而直接得到一个小模型。虽然这两种研究方向的方法与思路不尽相同，但最终的目的都是希望能够在保持模型性能的同时，提升模型检测速度，减小模型大小。这两种模型瘦身方法都有可取之处，但也都有其局限性。第一种方法由于模型剪枝率以及模型本身大小基数的限制，导致模型在剪枝后仍然较大，对移动端部署来说仍然有进一步优化的必要。第二种方法，为了获得较小的网络模型，就需要尽可能少的网络层数与尽可能浅的网络深度，从而使网络的平均检测精度降低，因此其最终的模型在精度方面仍然需要进一步提高，且模型大小对于移动端部署来说仍需要进一步减小。

针对上述问题，本文提出先应用轻量化网络作为骨干网络，实现网络模型的初步轻量化，再对轻量化后的网络模型进行针对性的优化，提高模型检测精度，最后再应用通道剪枝方法，对模型进一步瘦身，最终获得轻量化且高效的模型。

3.1　YOLO-Person 模型结构

如图1所示，为获得轻量化夜间红外行人检测神

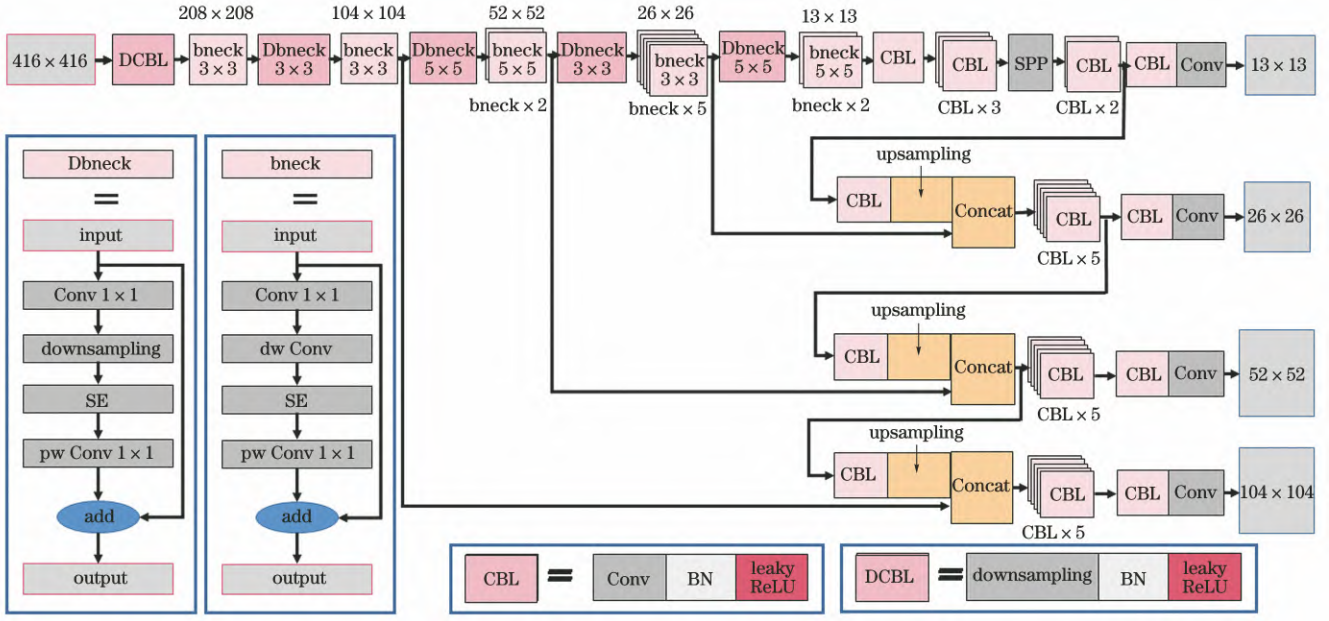

图 1　YOLO-Person 模型网络结构图

Fig. 1　Structure diagram of the YOLO-Person model network

经网络模型，本文在 MobileNetV3 轻量化主干网络的第 1 个 DCBL 层和第 1～4 个 Dbneck 层中添加步长为 2 的卷积操作，以实现特征图不同倍率降采样。在主干网络后添加四个不同尺度和感受野的检测层，分别获取经过主干网络 4、8、16、32 倍降采样后的特征信息，根据不同深度的特征映射对多尺度目标框进行分类和回归。为解决红外图像中行人目标尺度不均衡问题，在主干网络后使用空间金字塔池化（SPP）模块丰富深层特征空间信息，并传递给四个检测层。添加前三个检测层主要针对中尺度和大尺度目标，最后一个层引入 4 倍降采样的小感受野检测层，融合主干网络第二次降采样后的特征信息以提高小目标检测精度，不同尺度检测层感受野大小不同，因此能够获取到多尺度特征信息用于分类和回归，增强模型拟合能力。最后，对该网络模型进行通道剪枝，大大减少模型参数量，加快推理速度，使之适用于移动端的夜间红外行人检测。

3.2 初步轻量化网络模型构建

复杂网络模型往往难以在嵌入式设备或移动端部署应用，一方面是由于复杂模型参数量较大，对移动设备的 GPU 内存有较高要求，另一方面是由于很多的实际应用场景中对网络模型都有低延迟与快速响应的要求，而越复杂的网络模型往往反应速度越慢。为了解决行人检测模型在移动端部署时面临的占用空间大、检测速度慢的问题，本文应用 MobileNetV3 轻量化网络为骨干网络，以多尺度融合目标检测层为检测模块的网络模型，将其命名为 YOLO-MN。

MobileNetV3 网络[15]是在 MobileNetV1 网络[16]与 MobileNetV2 网络[17]的基础上进一步改进得到的。MobileNetV3 将平均池化前的网络层移除并用 1×1卷积来计算特征图，在保留高维特征的同时，还节约了计算时间成本，比 MobileNetV2 提速了约 11%。SE 注意力模块能够提高重要特征的权重，因此 MobileNetV3 引入了 SE 注意力模块。将 SE 注意力模块放到 MobileNetV2 中最后的逐点（pw）卷积前，先完成 SE 操作，再进行 pw 卷积操作。

Howard 等[15]通过实验发现使用 swish 非线性激活函数替代 ReLU 激活函数能够提高模型的精度，但是 swish 非线性激活函数的 sigmoid 计算成本太高，会增加移动端的耗时，于是对 swish 函数进行了数值近似，提出了 h-swish 函数：

$$\text{h-swish}(x)=x\frac{\text{ReLU6}(x+3)}{6},\tag{1}$$

式中：x 为输入特征值；ReLU6 为激活函数，ReLU6=min[6，max(0，x)]。

该函数在保持精度的情况下带来了很多优势，既可以兼容大多数软硬件框架，又可以避免数值精度的损失，运行效率提高了大约 15%。因此，本文将 MobileNetV3 作为骨干网络来提取图像特征。

为了使网络模型能够充分利用浅层特征，提取丰富的空间信息，以获得更高的检测精度，本文在 MobileNetV3 网络后添加了三个不同大小的目标预测层以获得多尺度感受野，并融合浅层特征与深层特征，丰富了特征图的信息，获得更好的红外行人检测效果。

3.3 空间金字塔池化及多预测层网络模型构建

在一定范围内，网络模型的深度与提取到的特征信息丰富程度以及检测效果成正比关系，但网络模型的平均检测精度除了与网络模型的深度有关，也与特征图的表征能力以及数据集中待检测目标的特点有关。辅助驾驶红外数据集是由车辆在实际的交通环境中行驶时采集的，因此在数据集中会出现近距离处的行人在采集的红外图像中较大，而远处的行人在采集的红外图像中很小的情况，即严重的行人目标尺度大小不均衡现象。针对这一问题，本文通过在 YOLO-MN 网络模型中依次加入 SPP 模块和降采样小目标预测层的方式，丰富特征图的表征能力，获取更小的感受野，从而提高 YOLO-MN 模型的平均检测精度。

本文的 SPP 结构建立在 He 等[18]的 SPP 思想的基础上，是一种能够实现全局特征和局部特征融合，提高特征图表征能力的网络结构，如图 2 所示。该结构由卷积核尺寸为 13×13，9×9，5×5 的三个最大池化层以及一个跳跃连接并行构成。特征图先经过不同尺寸的最大池化层池化，再将池化后的特征图与原输入特征图进行融合并输出。

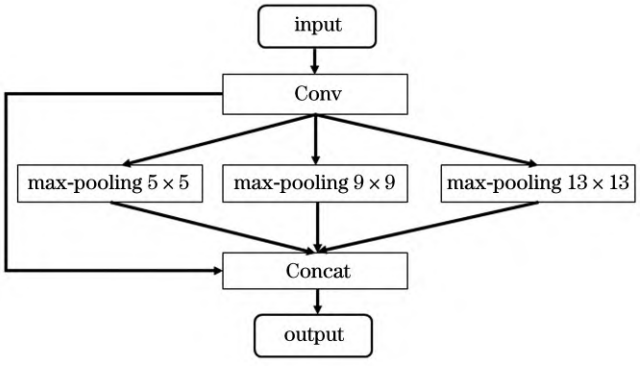

图 2 SPP 模块结构图

Fig. 2 Structure diagram of the SPP module

MobileNetV3 骨干网络会对输入图像进行特征图的提取，提取一个 13×13 的特征图，因此本文将 SPP 模块放置在骨干网络之后、第一个预测层之前，用 13×13 的最大池化层提取全局特征，用 5×5 与 9×9的最大池化层提取局部特征，最终实现了全局特征与局部特征的融合，增强了特征图的表征能力，丰富了深层特征的空间信息，有利于网络模型检测精度的提高。将添加 SPP 模块后的网络命名为 YOLO-MN-SPP。

在该网络模型骨干网络中有 32、16、8 倍降采样三个目标预测层，根据其感受野的不同分别对数据集中

大、中、小不同尺寸的目标进行预测。但由于数据集中行人目标尺度大小严重不均衡，网络容易对小目标红外图像中行人造成漏检与误检。因此，在该网络中引入了一个 4 倍降采样的小感受野目标预测层，通过获得更小感受野的方式进一步提高模型对小目标的检测效果。将 4 倍降采样预测层置于该网络的末端，并使其与骨干网络中第二次下采样的浅层特征进行融合，再把融合后的特征图输入到预测层用于目标预测。将添加小目标预测层后的网络命名为 YOLO-MN-SPP1。

3.4 模型通道剪枝轻量化

YOLO-MN-SPP1 网络模型中仍然存在很多的冗余参数和通道，冗余通道会增加模型计算量，降低模型的反应速度，因此需要对 YOLO-MN-SPP1 模型进行进一步通道剪枝轻量化，剪去冗余的通道数，得到一个适用于移动端部署的、小而高效的网络模型。

模型在剪枝之前需要进行稀疏化训练。稀疏化训练时，网络模型的目标函数为

$$L=\sum_{(x,y)} l\left[f(x,W),y\right]+\lambda\sum_{\gamma\in\Gamma} g(\gamma), \tag{2}$$

式中：(x,y) 是训练的输入坐标；W 是网络中的参数；$\sum_{(x,y)} l\left[f(x,W),y\right]$ 是卷积神经网络的训练损失函数；$g(\cdot)$ 是稀疏惩罚项；选择 L1 正则化用于训练中的自动剪枝；λ 是缩放因子。上述损失函数和常规损失函数的区别为增加了第二项，其中 $g(s)=|s|$，可以提高系数 γ 的稀疏性。

在对模型完成稀疏化训练得到所有通道对应的 γ 和 λ 后，再进行剪枝操作。本文采用批归一化(BN)层通道剪枝的方法对模型进行压缩。所谓的 BN 层通道剪枝就是用 BN 层的缩放因子 γ 对 BN 层输出通道的重要程度进行评估，γ 的大小说明了所对应的通道的重要程度，γ 值越大说明该通道越重要，对 γ 较小的通道进行裁剪，将不重要的通道从模型中移除，提高模型的检测速度，减小模型的存储空间。

模型通道剪枝示意图如图 3 所示。图 3 左侧为剪枝前结构，右侧为剪枝后的模型结构，ith 和 jth 分别表示第 i 个卷积层和第 j 个卷积层，C_{in} 表示第 i 个卷积层的第 n 个通道。由原始模型图可知，C_{i2} 和 C_{i4} 对应的缩放因子 γ 值较小，因此需要将这两个通道进行修剪，保留较大缩放因子对应的通道。通过此方法获得最终的轻量化模型 YOLO-Person。

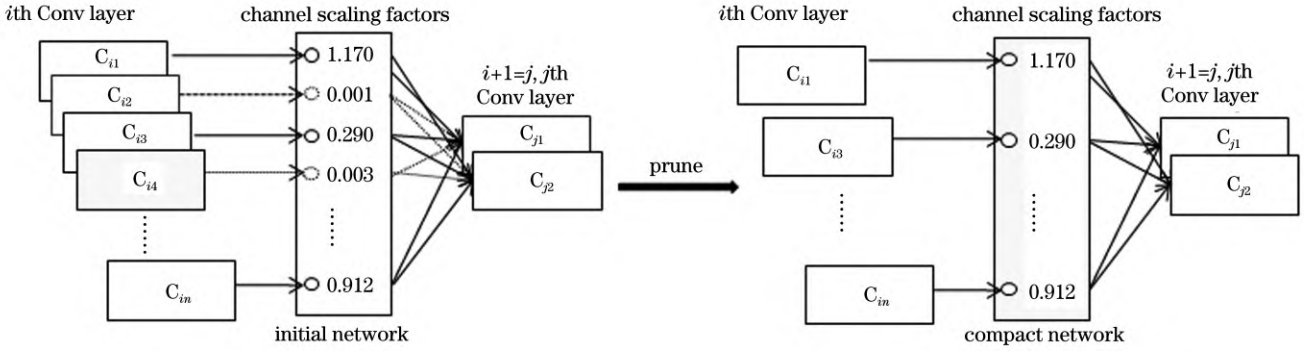

图 3 BN 层通道剪枝示意图

Fig. 3 Schematic diagram of channel pruning on the BN layer

4 实验结果讨论与分析

4.1 数据集建立

本文数据集采用的是 FLIR Thermal Starter 辅助驾驶红外数据集，其相较于自建数据集来说更具有权威性，得到的结果也更为客观。通过对 FLIR Thermal Starter 辅助驾驶红外数据集进行整理发现，原数据集中图像取帧间隔较短，导致相邻帧间图像的相似度很大，因此需要对原数据集进行筛选，增大取帧间隔，剔除相似度大的图像，避免后期训练过拟合。又由于主要针对夜间行车时的行人进行检测，故剔除白天所拍摄的图像，只保留夜间环境下拍摄的红外行人图像。经过筛选，最终得到了 2054 张图像。

进一步分析数据集发现，数据集中行人的姿态各异，在图像上表现出不同的特征。某些姿态数目较少，例如被遮挡的人和骑自行车的行人，导致不同类型的特征数量间存在不均衡，将会对模型最终的平均检测精度造成很大的影响。为了降低这一影响，提高模型的平均检测精度，采用姿态扩充的方法对数据集中相对较少的行人特征类型进行扩充。其过程如图 4 所示。

根据图 4 所示过程，从一张图像中对需要增加的行人姿态特征进行提取，再将提取到的行人姿态特征添加到原图或者数据集的其他图像中，最后将融合后的图像保存到数据集中。融合后的图像如图 5 所示。

通过对数据集中数量较少的行人姿态特征进行扩充，最终数据集图像数量达到了 2254 张。扩充后的数据集将为后续模型的训练提供更丰富的训练样本，提高模型的平均检测精度和鲁棒性。

实验前将数据集中的 2254 张图像进行划分和标注，其中 1352 张图像作为训练集，451 张图像作为验证集，451 张图像作为测试集，分别用于训练、验证和测试模型。然后使用 YOLO-mark 软件进行标注，将正常行走的行人、坐姿行人、骑自行车的行人以及骑摩

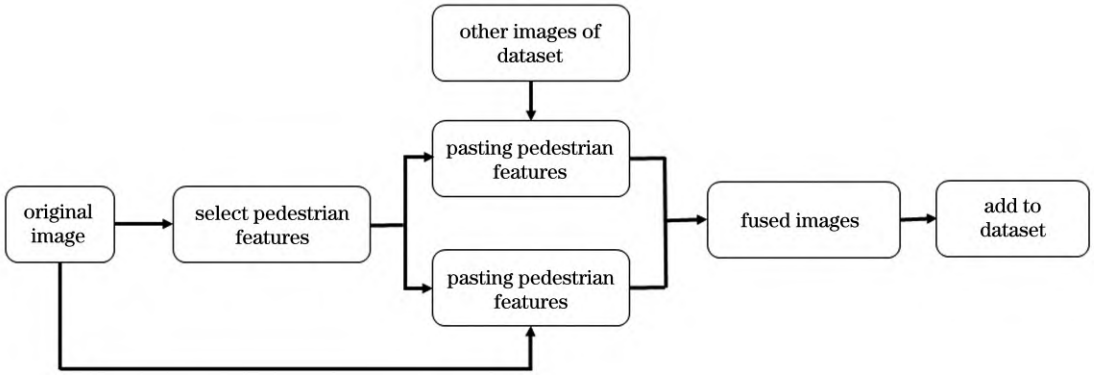

图 4 行人特征扩充流程图

Fig. 4 Flow chart of pedestrian feature expansion

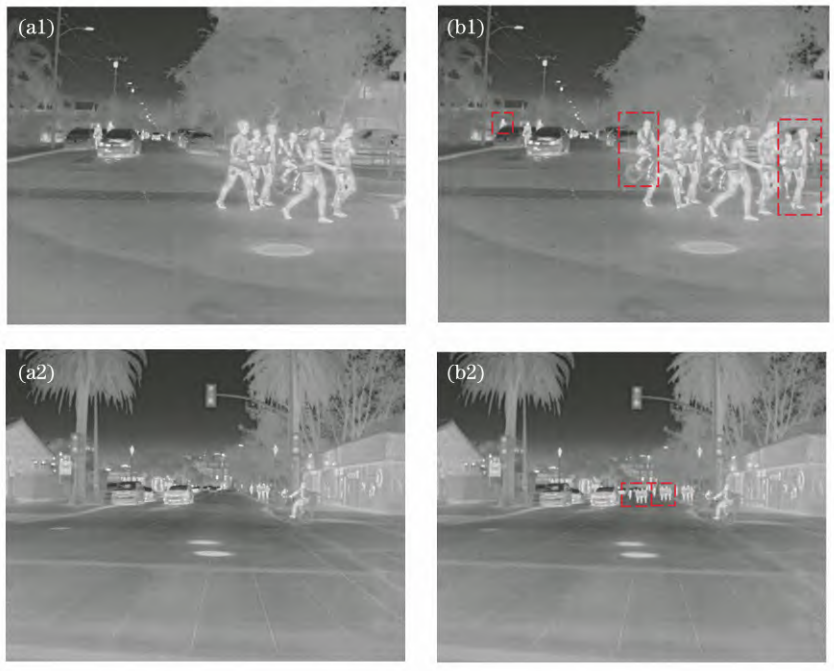

图 5 行人特征扩充。(a1)(a2)原图;(b1)(b2)特征扩充后的图

Fig. 5 Pedestrian feature expansion. (a1)(a2) Original images; (b1)(b2) feature-expanded images

托车的行人统一标注为“人”这一类别。

4.2 实验平台与实验参数

本文以 Ubuntu18.04 环境下 Pytorch 版本 YOLOv3 网络模型和 MobileNetV3 网络模型为基础网络模型。实验平台为一台神舟战神 Z7-KP7EC 笔记本电脑。硬件环境为:运行内存 16 GB,显存 6 GB,i7 处理器,GPU(GTX-1060)。软件环境为:Python3.7,CUDA-10.0 加速环境和 Pytorch1.1 神经网络框架。

网络训练参数的选取主要包括学习率、动量常数、权值衰减系数和损失函数。其中,学习率是影响模型收敛的一个重要参数,较小的学习率会导致模型训练时间过长,而较大的学习率又会导致目标函数收敛困难。本文选取中等初始学习率 0.0001,在训练过程中学习率再分阶段下降,这种学习率的选取方式既能保证模型的收敛速度又能保证模型的收敛效果。动量常数主要用于提高模型收敛速率,本文动量常数取 0.9。权值衰减系数用以调节由于模型复杂度而对损失函数造成的影响,权值衰减系数过大会出现过拟合现象,因此本文网络模型的权值衰减系数取 0.0005。IoU 损失函数是目标检测算法中 bounding box 的评估指标,能够对各种形状间的匹配度进行衡量。CIoU 损失函数在 IoU 损失函数的基础上进行了优化,把目标与锚点间的距离、尺度、重叠率以及惩罚项等多种可能的不利情况都进行了考虑,增加了目标框回归的稳定性,避免训练中出现的发散问题,因此选用 CIoU 损失函数。

4.3 初步轻量化 YOLO-MN 网络模型实验分析

为了得到一个初步轻量化的网络模型,提出了一种以 MobileNetV3 轻量化网络为骨干网络,以多尺度融合目标检测层为检测模块的网络模型 YOLO-MN,将其与 YOLOv3 网络模型分别对本文的夜间红外图像行人数据集进行训练,迭代次数为 600 个 epoch。实验结果如表 1 所示。

表 1 YOLO-MN 模型测试结果

Table 1 YOLO-MN model test results

Network structure	Accuracy /%	Speed /(frame·s^{-1})	Model size /MB
YOLOv3	90.2	26	246
YOLO-MN	89.0	62	95

由表 1 可知，YOLO-MN 网络模型对夜间红外图像行人的检测精度虽然略低于 YOLOv3 网络模型，但也达到了 89%。在模型的检测速度和大小两方面，YOLO-MN 网络模型要远好于 YOLOv3 网络模型，检测速度由 26 frame/s 增加到了 62 frame/s，网络模型的大小也由 246 MB 减小到了 95 MB。通过实验对比可知，本文提出的 YOLO-MN 网络模型符合实验前的预期，基本实现了网络模型的初步轻量化，在保持模型检测精度的同时提高了模型的检测速度，减小了模型大小，有利于模型后期检测精度的进一步提升以及模型剪枝。

4.4 空间金字塔池化及多预测层网络模型实验分析

为了验证 SPP 模块和小目标预测层对夜间红外图像行人检测的有效性，分别应用 YOLO-MN-SPP 模型和 YOLO-MN-SPP1 模型在夜间红外图像行人数据集上进行实验，迭代次数为 600 个 epoch。实验结果如表 2 所示。

表 2 YOLO-MN-SPP 模型测试结果

Table 2 YOLO-MN-SPP model test results

Network structure	Accuracy /%	Speed /(frame·s^{-1})	Model size /MB
YOLO-MN	89.0	62	95.0
YOLO-MN-SPP	90.1	47	99.5
YOLO-MN-SPP1	92.3	39	100.1

在表 2 中，由 YOLO-MN 网络模型和 YOLO-MN-SPP 网络模型的对比可知，YOLO-MN-SPP 网络模型的平均检测精度相较于 YOLO-MN 模型提升了 1.1 个百分点，和 YOLOv3 的精度基本相当，达到了 90.1%。在检测速度和模型大小两方面，由于增加了 SPP 模块，导致模型计算量有了一定的增大，因此 YOLO-MN-SPP 网络模型相较于 YOLO-MN 网络模型在检测速度方面由原来的 62 frame/s 降低到了 47 frame/s，在模型大小方面也由原来的 95 MB 增加到了 99.5 MB。虽然增加 SSP 模块导致了模型在检测速度和模型大小两方面不如 YOLO-MN 模型，但相较于 YOLOv3 模型已经有了很大的提升。通过实验分析可知，SPP 模块的效果符合预期，增加了特征图的表征能力，提升了网络模型的检测精度。

由 YOLO-MN-SPP 模型和 YOLO-MN-SPP1 模型对比可知，YOLO-MN-SPP1 模型的平均检测精度相较于 YOLO-MN-SPP 模型提升了 2.2 个百分点，达到了92.3%。在检测速度和模型大小两方面，由于 YOLO-MN-SPP1 模型增加了 4 倍降采样的小感受野检测层，导致模型计算量也有一定程度的增加，因此 YOLO-MN-SPP1 模型检测速度降低到 39 frame/s，模型大小增加到了 100.1 MB。YOLO-MN-SPP1 模型的平均检测精度、检测速度以及模型大小都已经优于 YOLOv3 网络模型，但若要将其部署在移动端，仍然需要进行进一步的瘦身与加速。

4.5 模型通道剪枝实验分析

YOLO-MN-SPP1 网络模型中存在很多的冗余参数和通道，冗余通道会增加模型计算量，降低模型的反应速度，因此需要对 YOLO-MN-SPP1 模型进行通道剪枝轻量化。

对 YOLO-MN-SPP1 网络模型进行稀疏化训练，迭代次数为 1000 个 epoch，稀疏化速率取 0.001。稀疏化训练完成后，对稀疏化生成的模型进行剪枝。尝试不同的剪枝率对模型的剪枝效果，结果如表 3 所示。

表 3 不同剪枝率的剪枝结果

Table 3 Pruning results with different pruning rates

Pruning rate /%	Accuracy /%	Model size /MB	Number of channels
85	92.3	13.6	1450
90	92.3	12.4	967
92	92.2	12.1	774
94	92.2	11.8	580
95	91.2	11.7	485
96	29.5	11.6	387

由表 3 可知：当剪枝率为 95%时，模型的通道数、精度以及模型大小达到较好的均衡，此时模型的精度相较于剪枝前网络模型的精度并没有大幅降低，仍可通过后期的调参训练进行恢复。若剪枝率继续增大到 96%，则剪枝后模型的精度下降严重，由原来的92.3%下降到了 29.5%，难以在后续的调参实验中恢复。综上所述，YOLO-MN-SPP1 模型在保持模型精度条件下的极限剪枝率为 95%。因此本文以 95%的剪枝率对模型进行通道剪枝，将剪枝后的网络模型命名为 YOLO-Person1。

YOLO-MN-SPP1 模型剪枝前后通道数的变化如图 6 所示。图中纵坐标表示通道数，横坐标表示通道所在网络层数。本文的骨干网络采用的是轻量化的骨干网络，已经进行了轻量化处理，因此为了保持模型的精度，此处的通道剪枝主要针对多尺度检测层的多余通道进行剪枝。由图 6 可知，从第 64 层网络层开始进行通道剪枝，模型剪枝前有 9664 个通道，剪枝后通道数降到 485 个，压缩比率为 95%，极大地减少了模型通道数，提高了模型性能。

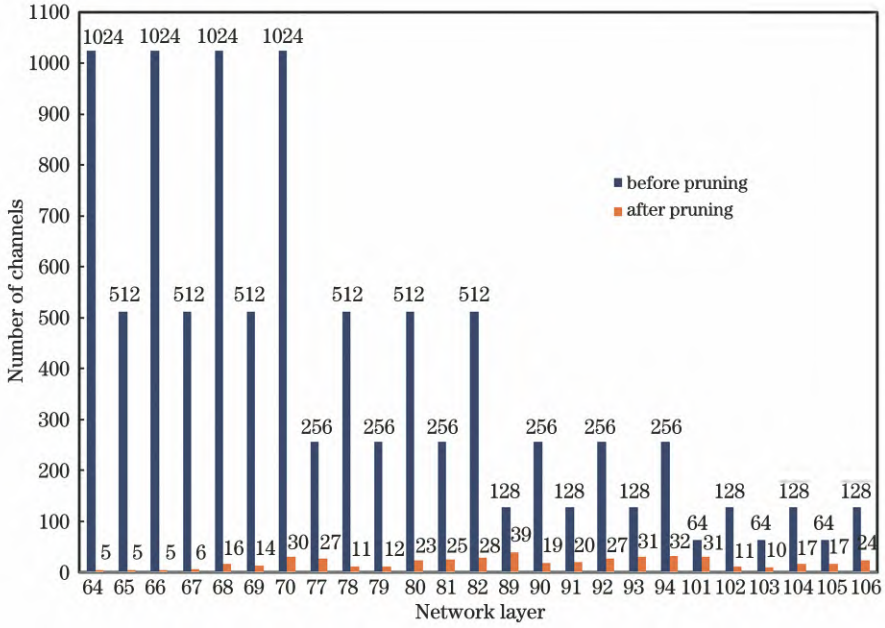

图 6 YOLO-MN-SPP1 模型剪枝前后通道数变化

Fig. 6 Change in the number of channels of YOLO-MN-SPP1 model after pruning

如上所述，YOLO-Person1 在平均检测精度上，由于减去了网络 95%的通道数，模型精度降低了 1.1 个百分点。为了恢复由于剪枝而损失的检测精度，对剪枝得到的模型权重进行微调，训练迭代次数为 600 个 epoch。将微调后的网络模型命名为 YOLO-Person，并与轻量网络 YOLOv3-tiny 网络模型进行对比，结果如表 4 所示。

表 4 YOLO-Person 模型测试结果

Table 4 YOLO-Person model test results

Network structure	Accuracy /%	Speed /(frame·s^{-1})	Model size /MB
YOLOv3-tiny	71.3	83	33.1
YOLO-Person1	91.2	69	11.7
YOLO-Person	92.2	69	11.7

由表 4 可知，微调后的 YOLO-Person 网络模型的检测速度与模型大小保持不变，但平均检测精度恢复到了 92.2%，与剪枝前网络模型的精度基本相当。微调后的 YOLO-Person 网络模型在各个方面都已经超过了 YOLOv3 网络模型，但若与 YOLOv3-tiny 模型的检测速度相比仍然有一定的差距。YOLOv3-tiny 网络模型是一种通过简化网络结构、牺牲模型精度来追求检测速度的模型，因此其检测精度只有 71.3%，远不及本文 YOLO-Person 网络模型的 92.2%，也难以满足红外图像行人检测的要求。

通过上述剪枝和微调实验可知，模型剪枝能够大幅减少模型的计算量、提高模型的检测速度以及减小模型大小。通过通道剪枝实验与微调实验，最终本文得到的 YOLO-Person 网络模型相较于 YOLOv3 网络模型在检测精度、模型大小以及检测速度三方面都有了较大的优势，更适于部署在移动端平台，满足夜间红外行人目标实时检测的要求。

4.6 YOLO-Person 模型的嵌入式部署验证

2019 年 3 月 NVIDIA 发布的 Jetson Nano 是一款用于小型人工智能网络开发的移动实验平台[19]，可以让研究人员在简易的开发平台上验证神经网络方法的可行性。Jetson Nano 上可以部署各种常见的神经网络模型以及用户自行设计的网络模型，用以实现目标检测、图像分类、语音处理、语义分割等各种功能。通过在 Jetson Nano 移动开发平台上进行部署，进一步验证 YOLO-Person 模型相较于 YOLOv3 网络模型在移动开发平台中的优势。YOLO-Person 模型在 Jetson Nano 移动开发平台上对夜间红外图像行人目标的检测效果如图 7 所示。

图 7 YOLO-Person 模型在 Jetson Nano 上的检测效果

Fig. 7 Detection effect of YOLO-Person model on Jetson Nano

为了在 Jetson Nano 移动开发平台上验证所提方法的优越性，分别测试了 YOLO-Person 模型、YOLOv3 模型以及 YOLOv3-tiny 模型在 Jetson Nano 移动开发平台上的检测精度、模型检测速度以及模型大小，如表 5 所示。

由表 5 可知，本文所提出的 YOLO-Person 夜间红外图像行人检测模型在 Jetson Nano 移动开发平台上的检测速度为 12 frame/s。虽然其检测速度与 GTX-1060 显卡上的检测速度相差较大，但相对于 YOLOv3 模型的检测速度来说优势仍然十分明显，大约是 YOLOv3 模型检测速度的 10 倍，说明 YOLO-Person 模型的反应速度更快，更适合于移动端的部署。本文提出的 YOLO-Person 网络模型在 Jetson Nano 移动开发平台上的检测速度接近 YOLO 系列网络中最快的 YOLOv3-tiny 网络模型，而 92.2%的检测精度远远优于 YOLOv3-tiny 网络模型的 71.3%。综上可知，本文所提出的 YOLO-Person 网络模型的性能优于 YOLOv3 网络模型，更适合在移动端开发环境中部署。

表 5 Jetson Nano 平台测试结果

Table 5 Jetson Nano platform test results

Network structure	Accuracy /%	Speed /(frame·s^{-1})	Model size /MB
YOLOv3	90.2	1.2	246.0
YOLOv3-tiny	71.3	14.0	33.1
YOLO-Person	92.2	12.0	11.7

5 结 论

本文提出了一种轻量化的 YOLO-Person 夜间红外行人检测模型。该模型首先应用 MobileNetV3 轻量化网络作为骨干网络，多尺度融合检测层为预测模块，实现了模型的初步轻量化；然后在网络中添加空间金字塔模块和小感受野检测层，提高了对小目标的检测精度；最后，对模型进行通道剪枝，大大减少了模型参数量，得到最终的轻量化模型 YOLO-Person。在夜间红外图像行人数据集上进行验证，实验结果表明，YOLO-Person 模型检测精度和速度分别达到了 92.2%和 69 frame/s，符合夜间红外图像行人实时检测的要求。将 YOLO-Person 网络模型在 Jetson Nano 移动开发平台部署，得到 12 frame/s 的检测速度，超越了 YOLOv3 并接近 YOLOv3-tiny，进一步证明了所提方法的优越性。在后续的研究中将通过优化

网络结构、增加有效的功能网络层等方法进一步提高模型的检测精度。

参 考 文 献

[1] 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.

[2] Nanda H, Davis L. Probabilistic template based pedestrian detection in infrared videos[C]// Intelligent Vehicle Symposium, 2002. IEEE, June 17-21, 2002, Versailles, France. New York: IEEE Press, 2002: 15-20.

[3] Xu F L, Liu X, Fujimura K. Pedestrian detection and tracking with night vision[J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(1): 63-71.

[4] 王一同, 周宏强, 闫景逍, 等. 基于深度学习算法的计算光学研究进展[J]. 中国激光, 2021, 48(19): 1918004.
Wang Y T, Zhou H Q, Yan J X, et al. Advances in computational optics based on deep learning[J]. Chinese Journal of Lasers, 2021, 48(19): 1918004.

[5] 邹梓吟, 盖绍彦, 达飞鹏, 等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报, 2021, 41(15): 1515001.
Zou Z Y, Gai S Y, Da F P, et al. Occluded pedestrian detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2021, 41(15): 1515001.

[6] 刘学, 李范鸣, 刘士建. 改进的 SSD 红外图像行人检测算法[J]. 电光与控制, 2020, 27(1): 42-46, 59.
Liu X, Li F M, Liu S J. An infrared image pedestrian detection algorithm based on improved SSD algorithm[J]. Electronics Optics & Control, 2020, 27(1): 42-46, 59.

[7] 赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测[J]. 光学学报, 2020, 40(5): 0504001.
Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica, 2020, 40(5): 0504001.

[8] 赵亮, 胡杰, 刘汉, 等. 基于语义分割的深度学习激光点云三维目标检测[J]. 中国激光, 2021, 48(17): 1710004.
Zhao L, Hu J, Liu H, et al. Deep learning based on semantic segmentation for three-dimensional object detection from point clouds[J]. Chinese Journal of Lasers, 2021, 48(17): 1710004.

[9] 苗壮, 张湧, 陈瑞敏, 等. 基于关键点的快速红外目标检测方法[J]. 光学学报, 2020, 40(23): 2312006.
Miao Z, Zhang Y, Chen R M, et al. Method for fast detection of infrared targets based on key points[J]. Acta Optica Sinica, 2020, 40(23): 2312006.

[10] 于博, 马书浩, 李红艳, 等. 远红外车载图像实时行人检测与自适应实例分割[J]. 激光与光电子学进展, 2020, 57(2): 021507.
Yu B, Ma S H, Li H Y, et al. Real-time pedestrian detection for far-infrared vehicle images and adaptive instance segmentation[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021507.

[11] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[12] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

[13] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-06-08]. https://arxiv.org/abs/1804.02767.

[14] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-06-03]. https://arxiv.org/abs/2004.10934.

[15] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1314-1324.

[16] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-08-06]. https://arxiv.org/abs/1704.04861.

[17] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.

[18] He K M, Zhang X Y, Ren S Q, 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.

[19] 李玉华, 刘全程, 李天华, 等. 基于 Jetson Nano 处理器的大蒜鳞芽朝向调整装置设计与试验[J]. 农业工程学报, 2021, 37(7): 35-42.
Li Y H, Liu Q C, Li T H, et al. Design and experiments of garlic bulbil orientation adjustment device using Jetson Nano processor[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(7): 35-42.

Multi-Scale Feature Fusion Lightweight Real-Time Infrared Pedestrian Detection at Night

He Zifen*, Chen Guangchen, Chen Junsong, Zhang Yinhui**

Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, Yunnan, China

Abstract

Objective Poor lighting conditions lead to a high accident rate during night driving. In order to reduce the incidence of night traffic accidents, various auxiliary driving technologies such as ultrasonic ranging, millimeter wave radar and visual auxiliary driving are widely used. Infrared thermal imaging technology based on the thermal radiation of object and reflection imaging with certain penetrability is less affected by the weather and light conditions at night. Human targets within the vision field can be accurately captured by infrared thermal imaging technology, which is convenient for

pedestrian detection. In addition, the cost of infrared imaging equipment has been decreased in recent years, making it possible to be mounted on vehicles. Therefore, the fusion of infrared thermal imaging technology and pedestrian target detection algorithm based on deep learning is of great research significance and with a broad market application prospective in vehicle auxiliary driving. In this paper, a pedestrian detection model based on night infrared image is proposed for night driving, which can detect pedestrians on the night road in real time. This study can be applied to the field of auxiliary driving for early warning and active braking provided to drivers, reducing the probability of night driving accidents and providing higher security for vehicles and pedestrians.

Methods Aiming at the problems of low accuracy in infrared pedestrian detection for small targets at night, large committed memory of network model, and the difficulty of real-time detection in auxiliary driving due to the low model detection speed, a lightweight pedestrian detection neural network called YOLO-Person is proposed for night infrared images. Firstly, the MobileNetV3 lightweight network is used as the backbone network, while the multi-scale fusion target detection layer is used as the prediction module to solve the problem of large model size and slow inference speed, which greatly reduces the amount of model calculation and obtains a preliminary lightweight network model. Furthermore, by adding the spatial pyramid pooling module and the detection layer with smaller receptive field in the network, the representation ability is enhanced to solve the problem of unbalanced pedestrian target scale in the dataset and improve the infrared pedestrian detection accuracy. Finally, channel pruning is used to reduce the number of channels in the feature map, and the final network model YOLO-Person is obtained. The lightweight model YOLO-Person is verified on the pedestrian dataset of night infrared images based on Jetson Nano mobile development platform.

Results and Discussions A lightweight model YOLO-Person is proposed for night infrared pedestrian detection (Fig. 1). Firstly, MobileNetV3 lightweight network is used as the backbone network, and the multi-scale fusion detection layer is used as the prediction module. Although the accuracy is reduced by 1.2%, the speed is increased by 34 frame/s, and the model size is reduced by 151 MB (Table 1), which indicates that the lightweight of the night infrared pedestrian detection model is preliminarily realized. Secondly, aiming at the problem of unbalanced pedestrian target scale in dataset, spatial pyramid pooling module (Fig. 2) and small receptive field detection layer are added in the network, through which the accuracy is improved by 3.3%, the speed is reduced by 23 frame/s, and the model size is increased by 5.1 MB (Table 2). Moreover, the model is pruned (Fig. 3) to reduce a large number of redundant channels (Fig. 6). When the pruning rate is 95%, the number of model channels, accuracy and model size achieve balance and optimization (Table 3). In addition, the model is fine-tuned to obtain the final lightweight model YOLO-Person, which reaches the accuracy of 92.2%, the speed of 69 frame/s, and the model size of 11.7 MB (Table 4). Finally, the model is deployed on the Jetson Nano mobile development platform to verify the detection effect (Fig. 7), and the test results of three networks are compared. The lightweight model YOLO-Person gets the best results: the accuracy of 92.2%, the speed of 12 frame/s, and the model size of 11.7 MB (Table 5).

Conclusions A lightweight model YOLO-Person for night infrared pedestrian detection is proposed in this paper. Firstly, MobileNetV3 lightweight network is used as the backbone network, and the multi-scale fusion detection layer is used as the prediction module to achieve the preliminary model lightweight. Secondly, spatial pyramid pooling module and small receptive field detection layer are added to improve the detection accuracy of small targets. Finally, the model parameters are greatly reduced through channel pruning, and the final lightweight model YOLO-Person is obtained. The experimental results show that the detection accuracy and speed of YOLO-Person model reach 92.2% and 69 frame/s, respectively, meeting the requirements of real-time pedestrian detection. The YOLO-Person network model is deployed on the Jetson Nano mobile development platform, where the detection speed of 12 frame/s exceeds that of YOLOv3 and approaches that of YOLOv3-tiny, which further verifies the superiority of the proposed method. By optimizing the network structure and increasing the effective functional network layer, the detection accuracy of the model will be further improved in the future research.

Key words imaging systems; infrared pedestrian detection at night; multi-scale fusion; MobileNetV3 network; model pruning