基于YOLOv7融合注意力机制的灾后搜救机器人人体识别研究

冯鹏博1,2 祁宇明1,2\* 邓三鹏1,2 罗明坤1,2 朱晨1,2

（1.天津职业技术师范大学机器人及智能装备研究院，天津 300222；2.天津市智能机器人技术及应用企业重点实验室，天津 300350）

摘 要：针对复杂城市环境，地震后救援存在发现不够及时、识别率低的问题，提出一种基于YOLOv7的小目标人体检测算法模型SE-YOLOv7来解决这一问题。以YOLOv7作为基础，对复杂城市灾区作为背景做了人体识别数据集，在框架加入残差网络模块SE注意力机制[1]，以提高正样本特征的贡献来更好训练模型权重；结合受灾人体成像具有与正常情况不同的特点，改进先验框[2]长宽的分布密度，将该模型部署在灾后八足救援机器人[3]嵌入式平台上进行灾后环境的人体目标检测实验，对人体残肢识别准确度较voc2012数据集提高了2.3%map，召回率提高2.7%，速度为67.6FPS，满足灾情下救援机器人目标检测的实时性和精度要求。

关键词：YOLO v7；人体识别；注意力机制；灾后搜救；八足机器人

**作者简介：**冯鹏博（1995-），男，河北石家庄人，硕士研究生，研究方向：智能机器人技术及应用,E-mail：892480499@qq.com。

**通信作者：**祁宇明（1979-），男，天津人，副教授，硕导，研究方向为：智能机器人技术及应用，E-mail：284075043@qq.com。

基金支持：全国职业院校教师教学创新团队建设体系化课题研究项目(TX20200104)；天津市科技军民融合重大专项（18ZXJMTG00160）。

引言

地震、火灾、战争导致的灾难具有突发性、破坏性等特点。救灾时效性强，准确性要求高，但救援人员立刻投入救援工作易遇到二次危险状况。为保障救援人员的安全与救援效率，可在灾后第一时间投入救援机器人进入灾害现场，进行灾后复杂环境的信息采集与人体识别救援。

在灾后复杂环境中，救援机器人想要完全代替救援人员自主完成搜救任务，首先通过机器人到达受灾地区边缘的安全区，投放机器人使其自主进入灾区，同时打开摄像机实时采集灾后环境图像，然后根据图像对周围环境解析，最后借助视觉检测算法进行目标检测，已实现救援任务。传统的目标检测模型会使用几种滑动窗口从图像中生成目标候选框，随后在目标候选框中进行特征提取[4]，经过多次卷积、池化后，最后进行特征分类。该模型采用滑动窗口生成候选框，且经过池化后，数据量变大，造成时间复杂度高，不能适应更多样的目标特征，造成召回率偏低，鲁棒性较差。而基于深度学习的的目标识别，可以自适应的从大量图片中学习目标特征，精度与鲁棒性更好。

本文在单阶段目标检测模型YOLOv7的基础上，提出一种人体目标识别的改进模型，首先在原COV2012数据集上增加室外场景数据集，通过数据增强使正样本数量翻倍，使训练模型效果更好。本文针对室外灾难下复杂场景的人体目标识别，故选择了当下效率最高的YOLOv7模型作为基础框架，提出一种注意力机制与多尺度融合的SE-YOLOv7检测算法，提高了复杂背景下对人体检测的准确度，可较好满足紧急情况下的人员搜救工作。

**1 八足机器人硬件方案设计**

八足机器人采用串联腿部结构，使用60kg舵机为动力，3个舵机组成一条腿。单腿具有3自由度，可以完成前进、后退、左转向、右转向等功能，能够实现救援机器人的基础移动功能。一共24个舵机安装在图示位置，通过3个舵机分别带动3各关节，使得机器人具有3自由度。每条腿可以举起形成手-足通用模式，以完成更多姿势与动作，可以穿越复杂障碍物，腿上可搭载摄像头进行实时人体目标识别，满足灾后搜救要求。

|  |  |
| --- | --- |
| 8fa918b5823508f751632823dc983f3 |  |
| 八足机器人1 | 2e7db29743da637eb59f37ed06708f3 |
| 图1八足机器人设计 | |

1. **改进的人体识别视觉算法研究**

结合本实验数据集的情况，为解决灾后复杂环境下人体识别存在精度低、负样本数量过多、物体遮挡等问题，满足**准确率**与**召回率**需求，本文在YOLOv7基础上增加注意力机制，提出SE-ELAN模块提取主干网络，增强对目标的特征提取能力，使用K-means++算法聚类合适先验框，提高对正样本的检测精度。

网络设定输入层尺寸大小为640×640×3，backbone先使用4次Bconv卷积核[5]提取特征，再依次SE-ELAN与MPconv模块进行循环下采样，得到特征图的分辨率分别为80×80，40×40，20×20，随后在Head层进行一次上采样与一次下采样，使用cat模块将其进行稠密连接，最终提取出3个尺度的特征输出，可以满足不同尺度的目标检测，具体网络如下图2所示。

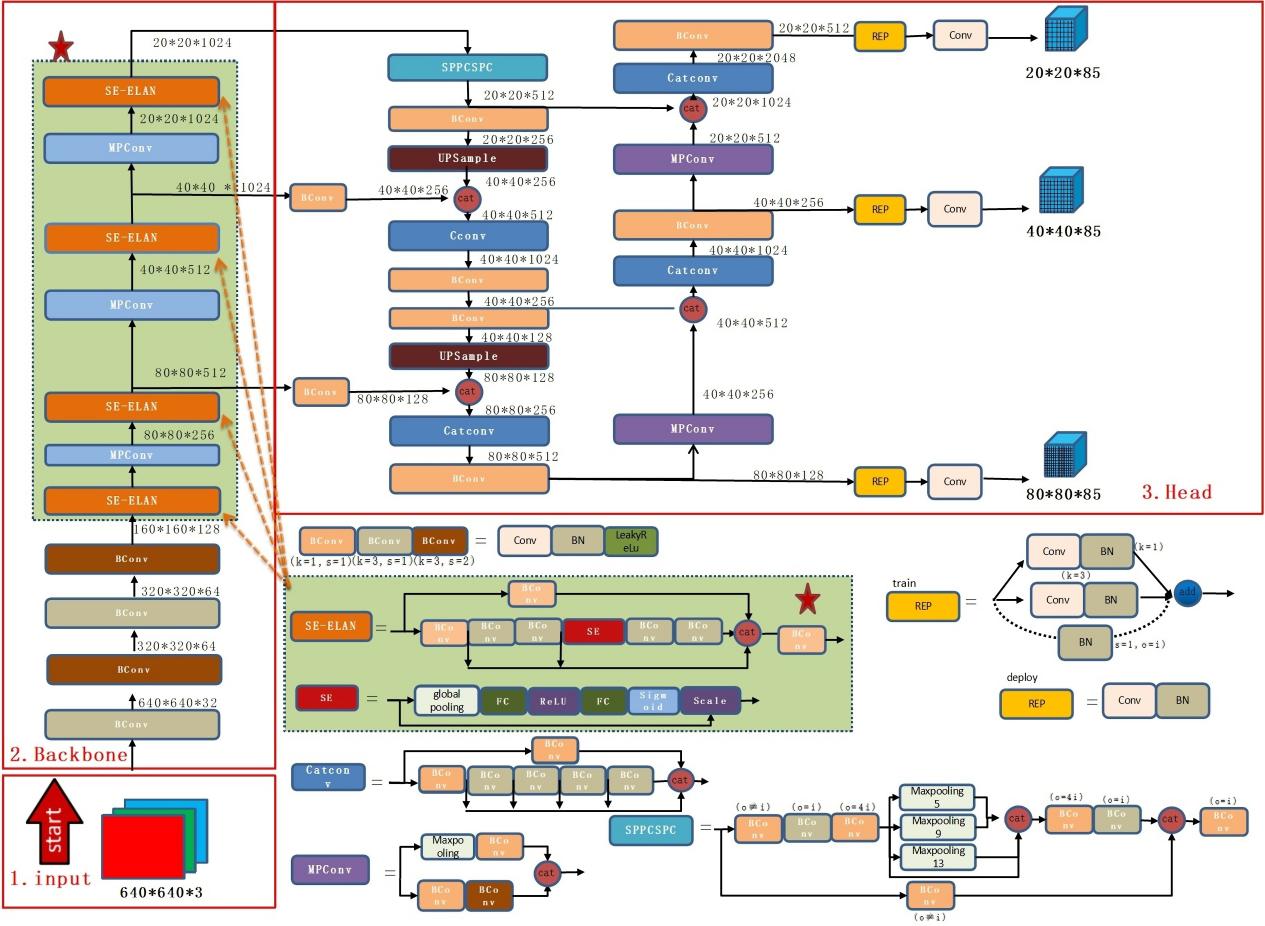


图 2加入SE注意力机制的YOLOv7网络结构

* 1. SE注意力模块

由于下采样增加了网络的感受野[6]，同时导致小目标的特征信息易丢失，针对此问题引入SE注意力机制以增强网络的特征提取能力，SE注意力机制可以解决下采样的卷积池化过程中因通道重要性差异导致的信息损失问题，故引入SE模块增加正样本特征在网络中的重要性。

SE模块结构图由压缩（Squeeze）、激励（Excitation）和Scale三部分组成，如下图3。首先通过squeeze操作，对空间维度进行压缩，对每个特征图做全局池化，平均成一个实数值。该实数具有全局感受野。excitaton操作根据特征通道的相关性生成一个代表各特征通道重要性的权重，同时增加了网络的非线性能力。最后Scale操作将权重通过乘法加权的方式乘到先前的特征上，从事实现提升正样本特征，抑制负样本特征这个功能。

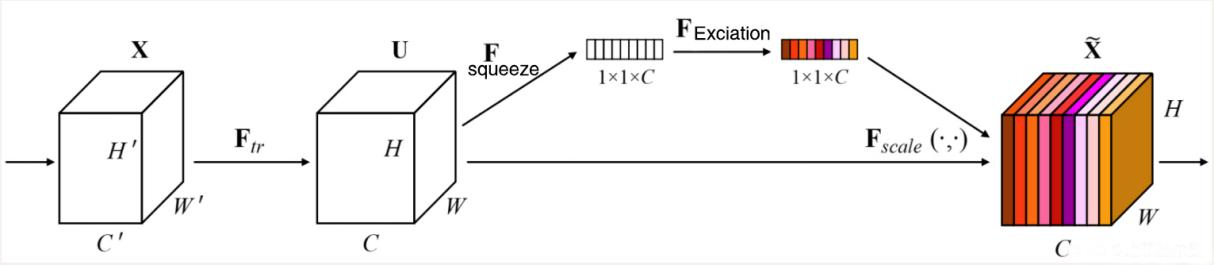


图 3 SE模块

2.2改进后的SE-ELAN

SE模块的位置也会影响网络的表征能力，如果将注意力机制模块放入Head层，则很难将融合后的feature map中区分出关键特征与无用特征，反而会使冗余特征变得更加重要，降低了网络的特征提取能力。在backbone中的浅层网络中有着更多图像原始的细节特征，更有利于对小目标人体识别[7]。随着网络加深，一些区域纹理信息也会丢失，因此将SE注意力模块嵌入ELAN的残差单元中，先将提取的特征进行注意力机制来突出关键特征，弱化冗余信息，在进行残差操作进行深入提取，最后合并输出。

2.3先验框的优化设计

选择合适的先验框对提升网络优化效果有很大提升。因为灾难下的人体被物体遮挡导致人体所占比例与常态不同，原YOLOv7是基于COCO数据集得到的初始先验框不能很好满足检测需求，故需要进行自己的先验框尺寸选择。而K-means算法是基于随机选取初始值，聚类效果不稳定，为降低误差本文选择K-means++[8]算法，对灾后人体标签进行聚类，生成9组不同尺寸的Anchor。聚类结果如表1所示。

表 1 K-means++生成的先验框

|  |  |  |
| --- | --- | --- |
| 特征图 | 感受野 | Achor |
| 80×80 | 小 | (12,16),(20,36),(41,27) |
| 40×40 | 中 | (34,73),(72,55),(69,142) |
| 20×20 | 大 | (132,101),(182,238),(449,397) |

**3实验与结果分析**

**3.1平台搭载与模型训练**

本文模型训练与测试使用的软件环境与硬件环境配置为：win10操作系统，Pytorch深度学习框架，CUDAv11.1,Cudnnv8.2.0,OpenCVv4.5.5,CPU Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz,32GBRAM,GPU NVIDIA GeForce GTX 1060。

本文使用原始数据分辨率较高，实验设置input层输入图片大小为640×640×3，模型训练时对图像进行自适应调整策略，完成缩放与填充。网络共进行300次epoch，batch size设置为4，学习率为动态变化，设置为le-2余弦变化衰减至le-3的过程，起步和终止时学习率较低，中间过程学习率较高。

3.2 评价指标

本实验选用候选框与目标真实框的交并比（IOU）的阈值大于0.5来作为标准评测目标检测，AP即平均精度、mAP即平均精度均值以及fps即每秒帧数当做本模型评价指标。P准确率代表正确预测样本占总检测目标比值，R召回率代表正确预测样本占所有目标比值，AP为P-R曲线与坐标轴所围面积。



****

****

****

3.3 消融实验结果

为更好分析SE-YOLOv7中改进模块对人体目标检测的优越性，设计消融实验，将改进后的网络加入原YOLOv7中，与原算法做对照组，具体实验内容如下表2所示。通过表2分析每项改进策略对本模型的贡献，从实验结果得出每个改进策略对模型整体性能有不同提升。

表2 SE-YOLOv7的模块消融实验

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | SE | K-means++ | Algorithm | P / % | R / % | mAP / % | 速率/(frames/s) |
| 1 | × | × | YOLOv7 | 94.3 | 90.2 | 92.1 | 69.7 |
| 2 | √ | × | SE-YOLOv7 | 95.4 | 92.1 | 93.8 | 67.9 |
| 3 | × | √ | SE-YOLOv7 | 95.0 | 92.6 | 93.6 | 69.3 |
| 4 | √ | √ | SE-YOLOv7 | 96.2 | 92.7 | 94.5 | 67.6 |

在模型残差网络中加入SE注意力机制，突出表现了关键特征，为进一步验证该机制优越，引入类激活图，如下图3所示。红色区域对图像分为正类提升贡献最大，蓝色则最小，成功优化了网络的识别性能，提升算法mAP至94.5%。

|  |  |  |
| --- | --- | --- |
| 2022_00008 | 2022_00008 | 2022_00008SE-v7 |
| 2022_00005原yolov7 | 2022_00006# | 2022_00005SE-yolov7 |
| (a)原YOLOv7 | (b)嵌入SE模块类激活图 | (c)检测结果 |

图 3 引入注意力机制前后对比图

4结论

针对YOLOv7模型对人体目标识别出现的小目标遗漏、错检等问题，本文提出一种基于注意力机制与多尺度融合的灾后人体目标检测算法，实现了复杂背景下对人体以及残肢的精准识别。

实验结果表明，本实验所改进模型在自建数据集中mAP值为94.5%,能精准检测出各种灾后各种角度的人体目标，相比于另几种流行模型，本实验的YOLO模型在综合性能有很大优势，满足灾后复杂背景下的人员搜救任务需求。后续应在保证算法精度同时对网络进行轻量化并提升泛化能力，使其应用于各种目标而不仅限于人体的目标识别。

5参考文献

[1]徐沁,梁玉莲,王冬越,罗斌.基于SE-Res2Net与多尺度空谱融合注意力机制的高光谱图像分类[J].计算机辅助设计与图形学学报,2021,33(11):1726-1734.

[2]李丝绦,张荣芬,刘宇红.基于改进先验框和损失的交通标志多尺度检测[J].智能计算机与应用,2021,11(06):56-62.

[3]武慧挺,周海晶,董仕涛.煤矿救援机器人嵌入式实时操作系统设计和分析[J].煤炭技术,2022,41(02):213-215.

[4]陈兵,蒋行国.卷积神经网络用于人脸特征提取[J].现代电子技术,2022,45(18):182-186.

[5]许亭亭. 基于卷积神经网络的手势估计的研究与应用[D].扬州大学,2022.

[6]郑伟,赵金芳,张奕婧,刘帅奇,张晓丹,马泽鹏.基于感受野扩增和注意力机制的U-Net脑肿瘤MR图像分割[J/OL].河北大学学报(自然科学版):1-10

[7]陈普强,郭立君,张荣,赵杰煜.基于全局空间约束块匹配的目标人体识别[J].计算机研究与发展,2015,52(03):596-605.

[8]陈天阳,钱政,荆博,韩妙荃.基于K-means++与ELM的短期风电功率预测模型研究[J/OL].电测与仪表:1-7