计算机科学
COMPUTER SCIENCE

**改进 YOLOv3 网络模型的人体异常行为检测方法**

张红民，李萍萍，房晓冰，刘宏

引用本文

张红民，李萍萍，房晓冰，刘宏. 改进 YOLOv3 网络模型的人体异常行为检测方法[J]. 计算机科学, 2022, 49(4): 233-238.

ZHANG Hong-min, LI Ping-ping, FANG Xiao-bing, LIU Hong. Human Abnormal Behavior Detection Method Based on Improved YOLOv3 Network Model[J]. Computer Science, 2022, 49(4): 233-238.

---

**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)**

基于时空自适应图卷积神经网络的脑电信号情绪识别

EEG Emotion Recognition Based on Spatiotemporal Self-Adaptive Graph ConvolutionalNeural Network

计算机科学, 2022, 49(4): 30-36. https://doi.org/10.11896/jsjkx.210900200

融合快速注意力机制的节点无特征网络链路预测算法

Link Prediction for Node Featureless Networks Based on Faster Attention Mechanism

计算机科学, 2022, 49(4): 43-48. https://doi.org/10.11896/jsjkx.210800276

图神经网络在 Text-to-SQL 解析中的技术研究

Technical Research of Graph Neural Network for Text-to-SQL Parsing

计算机科学, 2022, 49(4): 110-115. https://doi.org/10.11896/jsjkx.210200173

基于数据学习的结构静力学性能预测方法

Prediction Method of Structural Static Performance Based on Data Learning

计算机科学, 2022, 49(4): 140-143. https://doi.org/10.11896/jsjkx.210300238

共享浅层参数多任务学习的脑出血图像分割与分类

Intracerebral Hemorrhage Image Segmentation and Classification Based on Multi-taskLearning of Shared Shallow Parameters

计算机科学, 2022, 49(4): 203-208. https://doi.org/10.11896/jsjkx.201000153

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210300251

# 改进 YOLOv3 网络模型的人体异常行为检测方法

张红民[1]　李萍萍[1]　房晓冰[1]　刘　宏[2]
1 重庆理工大学电气与电子工程学院　重庆 400054
2 中国科学院计算技术研究所　北京 100080

**摘　要**　针对传统视频监控数据量大且复杂、不能及时有效地检测到人体异常行为的问题,文中提出了一种基于 YOLOv3 改进网络模型的人体异常行为检测方法(YOLOv3-MSSE)。该方法基于经典 YOLOv3 网络模型,利用残差模块构建多尺度特征提取网络,提升了对大目标的检测精度;同时,在网络结构不同位置融入注意力机制,对特征图各个通道的特征重要性实现加权处理,有效提高了模型人体异常行为的检测性能。实验结果表明,相比传统 YOLOv3 算法,YOLOv3-MSSE 方法的 mAP 值提升了 20.8%,F1-scores 提升了 11.3%,该方法不仅能够有效地检测出监控场景中的人体特定异常行为,还能较好地平衡检测精确率与召回率之间的关系,比同类方法更适用于实际监控场景下的人体异常行为检测。

**关键词**:神经网络;异常行为;多尺度;注意力机制;残差网络
**中图法分类号**　TP391

## Human Abnormal Behavior Detection Method Based on Improved YOLOv3 Network Model

ZHANG Hong-min[1],LI Ping-ping[1],FANG Xiao-bing[1] and LIU Hong[2]
1 College of Electrical and Electronic Engineering,Chongqing University of Technology,Chongqing 400054,China
2 Institute of Computing Technology,Chinese Academy of Sciences,Beijing 100080,China

**Abstract**　The data of traditional video surveillance is very large and complex,and cannot detect the abnormal behaviors of human in a timely and effective manner. In response to these problems,this paper presents an improved YOLOv3 algorithm (YOLOv3-MSSE) for the detection of human abnormal behavior. This algorithm can improve the detection accuracy of large targets for it is based on the traditional YOLOv3 network model,and a multi-scale feature extraction network is constructed by the residual modules. At the same time,by incorporating the attention mechanism into different positions of the network structure,and the importance of the features in each channel of the feature map can be weighted,which effectively improves the detection performance of the model for abnormal human behavior. Experimental results show that compared with the traditional YOLOv3 algorithm,the mAP of YOLOv3-MSSE is increased by 20.8%,and F1-scores is increased by 11.3%. The proposed algorithm can not only detect the specific abnormal behavior of the human in the monitoring scene effectively,but also can balance the relationship between the detection accuracy rate and the recall rate well. In addition,It is more suitable for the detection of human abnormal behavior in actual monitoring scenarios than similar methods.

**Keywords**　Neural networks,Abnormal behavior,Multiscale,Attention mechanism,Resnet

## 1　引言

在背景复杂的应用场景下,实现准确、实时的异常行为检测已成为智能监控视频领域具有挑战性的任务[1-2]。然而,异常行为的类型极其复杂,在不同的场景下其定义也有所不同。例如,在公路上行驶的车辆被定义为正常,而行驶在操场时则被定义为异常;此外,异常行为相对于正常行为发生的概率一般偏低。鉴于此,目前针对人体异常行为的研究通常是将任何与大多数人体行为不一致的行为定义为异常[3]。Wang 等[4]提出用无监督聚类的方法来提取人体关键帧,通过判断其幅值加权方向角的熵值来判断行为是否异常,但其无法对人体异常行为进行精确定位。Satybaldina 等[5]提出了一个对人体静态和动态手势特征进行分类的系统,以检测 ATM 机前的人是否存在可疑行为,但其检测精度不高,且易受光照条件的影响。值得一提的是,现有大多数人体异常行为检测算法以户外检测为主,这些算法的室内检测效果并不理想。文献[6]提出了一种基于机器学习算法的简明检测框架,通过对校内学生异常行为进行先验性定义,实现了视频监控中异常行为的有效检测,但其检测框架包含的手动参数较多,不利于自训练。Zheng 等[7]同样针对室内人体

到稿日期:2021-03-22　返修日期:2021-06-10
基金项目:重庆市自然科学基金面上项目(cstc2021jcyj-msxmX0525)
This work was supported by the Natural Science Foundation of Chongqing,China(cstc2021jcyj-msxmX0525).
通信作者:张红民(hmzhang@cqut.edu.cn)

异常行为检测和识别的问题，提出了一种多技术混合跟踪的方法，有效改善了障碍物遮挡下的识别精度，但其研究仅局限于单个摄像机的目标对象识别，缺乏对复杂场景下多摄像机下目标对象的识别研究。文献[8-9]则将人体异常行为当作一种目标进行分类，借用 YOLOv3 目标检测网络模型来实现异常行为的检测以及定位，在保证视频检测实时性要求的同时又兼顾了精确率，但大部分基于 YOLOv3 的改进算法更关注精确度本身，忽略了检测精度与召回率之间的平衡问题。

针对现有人体异常行为检测方法中存在的检测精度不高、定位不准确等问题，本文提出了一种基于 YOLOv3 改进网络模型的人体异常行为检测方法。其核心思路是：通过在 YOLOv3 网络结构中添加特征检测层来满足视频监控场景下对大目标的有效检测，同时在该网络结构中融入注意力机制进行训练检测，在网络训练时更加关注特征图的细节信息，使后期模型检测的性能得到明显提升。

## 2 改进 YOLOv3 网络模型的人体异常行为检测方法

### 2.1 改进方法的基本思路

经典 YOLOv3 利用多尺度预测的方法实现了精度与速度的较好平衡，但其在预测时用于提取特征图的最大尺寸为 13×13，相比 SSD[10] 的 1×1 明显偏大，易产生漏检或者误检。而深度残差模块[11]通过构建更深的检测网络得到了更大的输出尺寸，从而可以有效地解决该问题。此外，经典 YOLOv3 模型难以平衡检测精确率和检测召回率之间的矛盾，两者很难同时达到最优，而注意力机制[12]能够很好地学习到图像特征的重要程度并对其进行相应的加强和弱化，可有效提高模型的目标检测性能。

鉴于通道域注意力机制以及残差模块的特性，为平衡精确率与召回率之间的关系，通过对 YOLOv3 网络模型的分析，本文提出了一种人体异常行为检测方法（简称为 YOLOv3-MSSE），其网络结构图如图 1 所示。

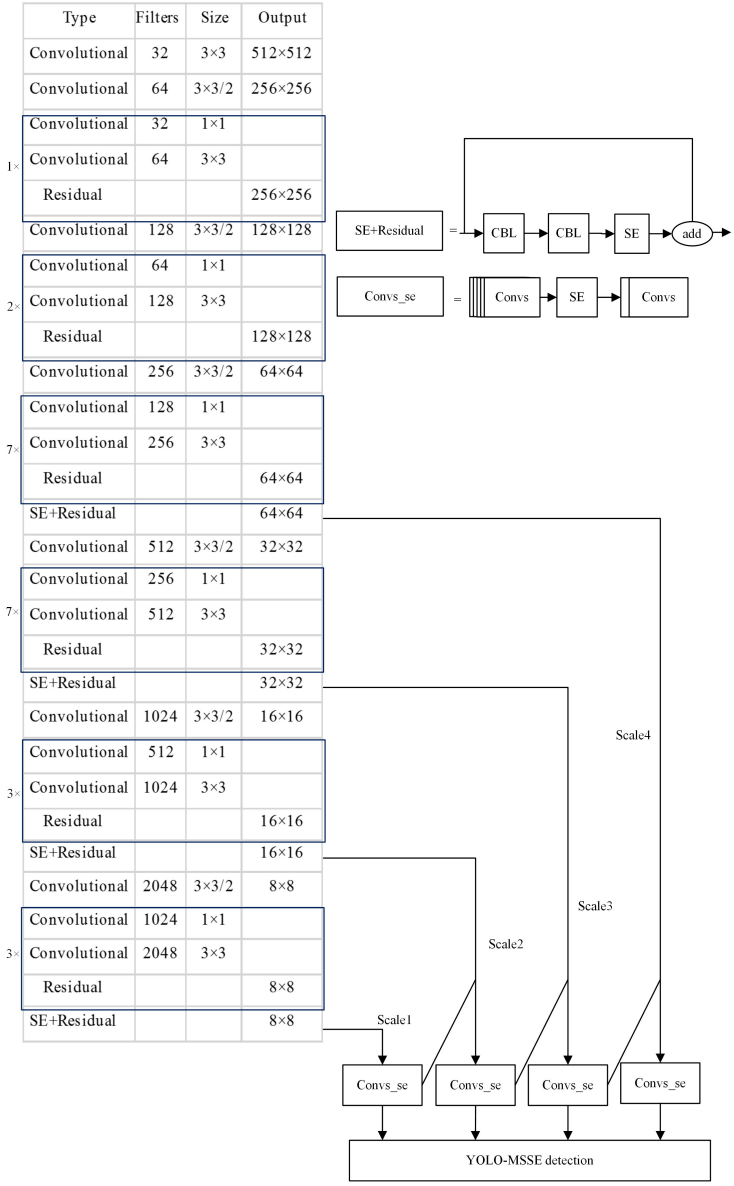



图 1 YOLOv3-MSSE 的网络结构图

Fig. 1 Network structure diagram of YOLOv3-MSSE

首先，通过构建更深的网络层使得改进后的网络模型可以对更大目标进行精确检测；其次，通过添加通道域注意力机制来学习图像特征的分布，以加强有用特征信息的权重，削弱无用信息的干扰

### 2.2 残差模块(ResNet)构建更深的检测网络

经典 YOLOv3 的主干网络由大量的 ResNet 构成，该模块能大幅度提高网络模型的训练速度以及模型性能，且不会给网络本身增加额外的参数和计算量。

YOLOv3-MSSE 对经典 YOLOv3 网络模型进行改进，其沿用经典 YOLOv3 主干网络的残差部分，在该主干网络之后额外再增加一个残差模块。模型对目标图像进行 64 倍下采样，得到 8×8 的特征图以满足对大目标的有效检测，加上 YOLOv3 原有的 3 个特征层组成新的具有 4 个特征尺度的特征金字塔，即分辨率分别为 64×64，32×32，16×16，8×8。

### 2.3 通道域注意力机制改变特征信息权重

注意力机制的本质就是权重的分布，其核心步骤就是从大量的信息中选取与特定研究目标有关的关键信息，通过细致学习这些特征以提高原网络模型的检测性能[13]。具体做法是：通过神经网络训练提取图像特征，将特征点的注意力集中在人群区域，以有效减少背景信息的干扰。

如图 2 所示，通道域注意力机制的实现需要经过 3 个步骤，首先对 $H\times W\times C$ 大小的特征图 $U$ 进行特征压缩，通过对卷积层的输出进行全局池化，得到一个 $1\times1\times C$ 大小的具有全局感受野的特征图，其计算式为：

$$Z_C=F_{sq}(\boldsymbol{U}_C)=\frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=1}^{W}u_C(i,j) \tag{1}$$

其中，$\boldsymbol{U}_C$ 表示 $U$ 中第 $C$ 个二维矩阵，$u_C(i,j)$ 为第 $i$ 行第 $j$ 列的像素特征，$F_{sq}$ 为特征压缩操作，$Z_C$ 代表压缩后 $C$ 通道的特征。

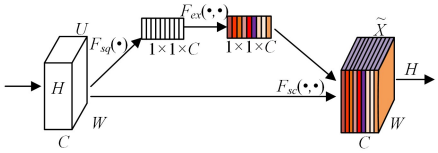



图 2 通道域注意力机制模型的结构图

Fig. 2 Structure diagram of channel domain attention mechanism model

其次，采用两个连续的全连接神经网络对特征压缩得到的具有 $C$ 个通道的特征图 $Z$ 进行非线性变换，得到一个具有权重系数的 $1\times1\times C$ 的卷积核 $S$。其压缩方式如式(2)所示：

$$S=F_{ex}(Z,W)=\sigma(W_2\delta(W_1Z)) \tag{2}$$

其中，$W$ 为一个 $1\times1\times C$ 的卷积核，用于学习网络权值；$F_{sc}$ 为乘法操作；$\sigma$ 为 sigmoid 函数；$\delta$ 为 Relu 激活函数，$W_1\in R^{C/r\times C}$，$W_2\in R^{C\times C/r}$，$r$ 为缩放参数。

然后，再对输入特征图 U 进行加权操作，即将其结果作为权重乘到输入特征上，得到一个加权的特征图 $\tilde{X}$ 作为下一层网络的输入。

$$\tilde{x}_C=F_{sc}(u_C,S_C)=S_C\cdot u_C \tag{3}$$

其中，$\tilde{x}_C$ 为 $C$ 通道对应的特征图，$\tilde{X}=[\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_C]$，$F_{sc}(u_C,S_C)$ 表示特征图 $u_C\in R^{H\times W}$ 与标量 $S_C$ 的对应通道的乘积。

为进一步提升模型的性能，分别在相应的残差层和特征提取层之间添加子通道域注意力机制，其添加方式如图 1 所示。YOLOv3-MSSE 的网络层能够自主学习特征图对应的每个通道的重要性参数，即学习每个通道的权重。最后，将参数反馈到输入特征图以实现特征图的重要性加权。加权后的特征图强化了检测目标，带有更为丰富的特征信息，有效提高了后期模型检测的准确性。

### 2.4 改进网络模型的损失计算

网络模型损失主要由边界框坐标误差损失、置信度误差损失和分类误差损失 3 部分组成[14]。

(1)边界框坐标误差损失

边界框坐标损失由中心坐标误差和宽高坐标误差组成，其计算式如式(4)所示：

$$lbox=\lambda_{coord}\sum_{i=0}^{S\times S}\sum_{j=0}^{B}1_{ij}^{obj}-(2-w_i\times h_i)[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2+(w_i-\hat{w}_i)^2+(h_i-\hat{h}_i)^2] \tag{4}$$

其中，参数 $1_{ij}^{obj}$ 表示第 $i$ 个网格的第 $j$ 个锚框是否负责这个检测目标，若负责则 $1_{ij}^{obj}=1$，否则等于 0。根据文献[15]定义参数 $\lambda_{coord}=5$ 以增加边界框坐标的损失。

(2)置信度误差损失

置信度误差损失函数包括存在对象的检测框的置信度误差、不存在对象的检测框的置信度误差两个部分，其计算式如式(5)所示：

$$lobj=-\sum_{i=0}^{S\times S}\sum_{j=0}^{B}1_{ij}^{obj}[\hat{c}_i\log c_i+(1-\hat{c}_i)\log(1-\hat{c}_i)]-\lambda_{noobj}\sum_{i=0}^{S\times S}\sum_{j=0}^{B}1_{ij}^{noobj}[\hat{c}_i\log c_i+(1-\hat{c}_i)\log(1-\hat{c}_i)] \tag{5}$$

其中，参数 $1_{ij}^{noobj}$ 表示第 $i$ 个网格的第 $j$ 个锚框不负责该目标的预测，参数 $\hat{c}_i$ 为置信度，其值由单元格的检测框是否负责预测某个对象所决定，如果负责，则 $\hat{c}_i=1$，否则 $\hat{c}_i=0$，根据文献[15]定义参数 $\lambda_{noobj}=5$。

(3)分类误差损失

分类误差损失函数使用式(6)计算。当第 $i$ 个网格的第 $j$ 个锚框负责某一个真实目标时，该锚框所产生的检测框才会去计算分类损失函数。

$$lcls=-\sum_{i=0}^{S\times S}1_{ij}^{obj}\sum_{c\in classes}[\hat{p}_i(c)\log(p_i(c))+(1-\hat{p}_i(c))\log(1-p_i(c))] \tag{6}$$

其中，$p_i(c)$ 代表当前目标属于类别 $c$ 的概率。

单个 YOLO 层的损失即为边界框坐标误差损失、置信度误差损失、分类误差损失这 3 部分之和。由于改进后的网络拥有 4 个 YOLO 层，因此网络模型最终的损失是将全部 YOLO 层计算出来的损失值求和，再进行反向传播。

## 3 实验结果与分析

### 3.1 模型训练及参数设置

YOLOv3-MSSE 网络模型实验选择在不同尺度的图像上进行训练与测试。输入图像被统一缩放到 512×512 像素，

通过深度残差网络提取目标特征，并使用 4 个不同尺度的卷积特征图来预测目标的位置坐标以及类型。对于锚点框的选取，采用 $k$-means++均值聚类法，即针对训练数据集的边界框做聚类以选择最合适的锚框。在自制的异常行为数据集上，YOLOv3-MSSE 使用 4 个尺度 12 个锚点，分别为：455×483，409×403，379×196，236×285，125×281，156×410，91×166，198×89，60×295，58×55，47×124，119×83。

在整个训练过程中，为避免训练出现过拟合，将 YOLOv3-MSSE 模型训练 150 个 epoch，其中动量和权重衰减分别配置为 0.9 和 0.0005，批数量设为 16，再通过细分将这 16 张训练图片分 32 次送入训练网络中以减少网络对显存的占用。为使模型更好地收敛，学习率刚开始设置为 0.01，随着迭代次数增加到设定值，学习率依次变为 0.001，0.0001。

### 3.2 模型通用性验证及实验结果

为了验证在通用场景下对于各类物体的目标检测性能，在 PASCAL VOC 数据集上对提出的 YOLOv3-MSSE 方法进行了相关的训练与测试。该数据集共包含 20 种类别，实验选取 VOC2007 测试集、VOC2007 训练验证集以及 VOC2012 训练验证集作为模型训练数据，选取 VOC2012 测试集作为模型测试数据，选择平均精确度（mAP）作为有效性检测指标。其中，5 种同类检测方法在 PASCAL VOC2012 测试集的检测结果如表 1 所列。

表 1 PASCAL VOC2012 测试集下的检测结果

Table 1 Test results in Pascal VOC2012 test

（单位：%）

| category | Faster R-CNN | SSD-300 | SSD-512 | YOLOv3 | YOLOv3-MSSE |
|---|---|---|---|---|---|
| aeroplane | 84.9 | 84.2 | 84.9 | 85.3 | 86.9 |
| bicycle | 79.8 | 76.3 | 82.6 | 84.6 | 84.6 |
| bird | 74.3 | 69.6 | 74.4 | 72.2 | 75.3 |
| boat | 53.9 | 53.2 | 55.8 | 68.7 | 71.5 |
| bottle | 49.8 | 40.8 | 50.0 | 67.3 | 71.1 |
| bus | 77.5 | 78.5 | 80.3 | 85.2 | 88.3 |
| car | 75.9 | 73.6 | 78.9 | 83.9 | 86.2 |
| cat | 88.5 | 88.0 | 88.8 | 84.6 | 86.7 |
| chair | 45.6 | 50.5 | 53.7 | 66.2 | 68.3 |
| cow | 77.1 | 73.5 | 76.8 | 83.9 | 85.3 |
| diningtable | 55.3 | 61.7 | 59.4 | 74.2 | 74.8 |
| dog | 86.9 | 85.8 | 87.6 | 84.8 | 84.7 |
| horse | 81.7 | 80.6 | 83.7 | 83.9 | 88.9 |
| motorbike | 80.9 | 81.2 | 82.6 | 84.3 | 86.4 |
| person | 79.6 | 77.5 | 81.4 | 82.8 | 84.1 |
| pottedplant | 40.1 | 44.3 | 47.2 | 56.2 | 59.2 |
| sheep | 72.6 | 73.2 | 75.5 | 76.7 | 78.3 |
| sofa | 60.9 | 66.7 | 65.6 | 78.1 | 78.2 |
| train | 81.2 | 81.1 | 84.3 | 84.4 | 84.8 |
| tvmonitor | 61.5 | 65.8 | 68.1 | 76.8 | 80.3 |
| mAP | 70.4 | 70.3 | 73.1 | 78.2 | 81.2 |

Faster R-CNN，SSD-300，SSD-512，YOLOv3 的数据均来源于文献[17]。从表 1 可以看出，YOLOv3-MSSE 对 PASCAL VOC 数据集下 20 个不同检测目标的检测 mAP 值达到了 80.2%，表明其能够适用于通用场景下的物体检测。与 Faster R-CNN（70.4% mAP），SSD-300（70.3% mAP），SSD-512（73.1% mAP），YOLOv3（78.2% mAP）的检测结果相比，YOLOv3-MSSE 的 mAP 值分别提升了 10.8%，10.9%，8.1%与 3.0%。

### 3.3 模型异常行为检测及结果分析

为了验证 YOLOv3-MSSE 在人体异常行为检测中的有效性，将其与经典 YOLOv3 以及在模型中仅添加了注意力机制的方法（简称为 YOLOv3-SE）进行检测结果对比分析。实验采用的数据集来源于公开的监控视频。考虑到人体异常行为的多样性，实验时仅将公共场所下行人跌倒、行人违规翻越栏杆定义为人体异常行为。在网络训练之前对输入图像进行预处理，包括图片数据的图像增强、视频相关帧截取、图像筛选以及标签标定等。其中数据标签共分为两类，即 fall 和 climb；整个数据集中共标定 2546 张图像数据，随机选取 2000 张作为训练数据集，将剩下的作为测试数据集。为了扩充训练数据集，在训练输入端采用数据增强策略，即将训练图像统一缩小至原尺寸的四分之一后再随机选取 4 张图片分别进行组合得到新的扩充图像，与未经处理的图片共同作为模型训练的数据集，并将其输入到网络模型中进行训练。

（1）3 种网络架构训练的 LOSS 函数图对比

实验时通过观察 LOSS 函数图的下降趋势来调整模型参数，并同时根据 LOSS 值确定何时停止训练。3 种网络架构训练的 LOSS 对比曲线图如图 3 所示。从 LOSS 函数图可看出，添加的注意力机制模块和特征检测层并不会影响模型最终的收敛情况，尽管 YOLOv3-MSSE 由于网络层更深使得模型初始损失值较大，但其经过较短时间的训练迅速减少了与 YOLOv3，YOLOv3-SE 损失值的差距。当模型分别训练到 50 个 epoch 时，3 种网络架构的训练损失值都开始缓慢下降，最终达到收敛的速度相差无几，均能较好地收敛在损失值 0.06 左右。

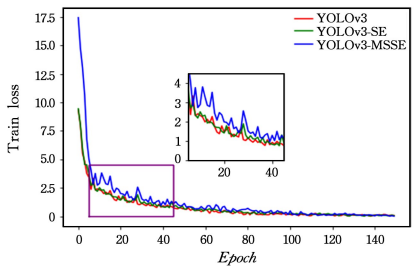


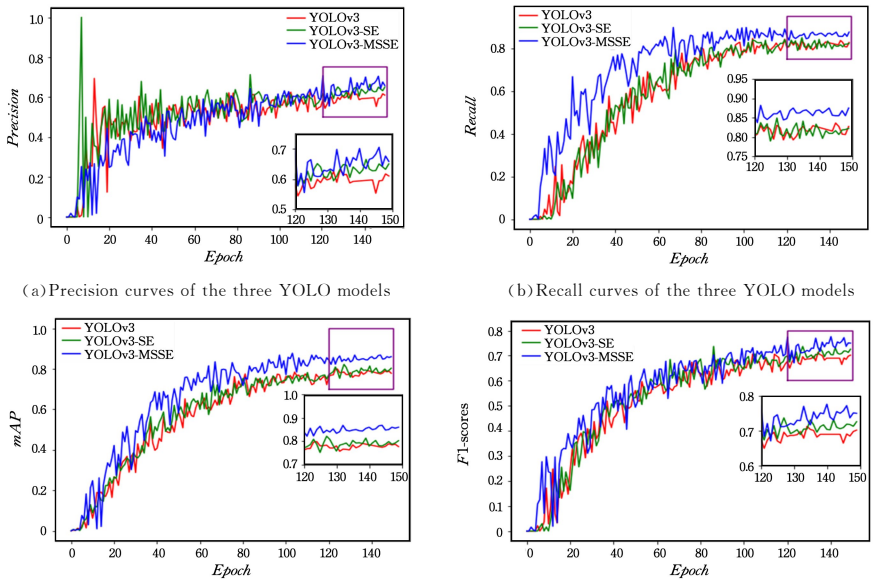

图 3 3 种网络架构训练的 LOSS 对比曲线图

Fig. 3 Loss comparison curve of three network architectures

（2）评价指标对比分析

为了验证 YOLOv3-MSSE 方法的有效性，选取精确率 Precision（P）、召回率 Recall（R）、mAP 以及调和平均数（F1-scores）作为评价指标。

3 种网络架构训练的各类指标曲线图如图 4 所示，图中蓝线代表 YOLOv3-MSSE，绿线代表 YOLOv3-SE，红线代表 YOLOv3。可以看出，各类指标随着迭代次数增加其值也不断增加，蓝线所代表的 YOLOv3-MSSE 网络模型检测的 P，R，mAP 以及 F1-scores 值始终在经典 YOLOv3 网络模型检测相关指标曲线图之上，绿线所代表的 YOLOv3-SE 网络模型检测的 P 以及 F1-scores 曲线也优于经典 YOLOv3 网络模型对应的指标曲线，说明添加的注意力机制、特征检测层模块使得 YOLOv3-MSSE 网络模型的检测效果较经典模型得到了一定的提升，其中同时添加了注意力机制和特征检测层的网络架构的性能提升得最多，检测性能更好。




(a)Precision curves of the three YOLO models

(b)Recall curves of the three YOLO models

(c)mAP curves of the three YOLO models

(d)Recall curves of the three YOLO models

图 4　3 种网络架构训练的各类指标曲线图(电子版为彩色)

Fig. 4　Curves of various indexes of three network architectures

(3)检测效果图对比

为了验证 YOLOv3-MSSE 方法的效果，将 YOLOv3-MSSE，YOLOv3-SE，YOLOv3 这 3 个网络架构在测试集(包含 546 张测试图片)中进行检测实验，随机选取两张检测图片并对其检测结果进行对比，结果如图 5 所示。

(a)翻越栏杆　　(b)行人摔倒

图 5　3 种网络架构的检测结果对比

Fig. 5　Comparion of test results of three network architectures

图 5 中第一行到第三行分别代表 YOLOv3，YOLOv3-SE，YOLOv3-MSSE 模型的检测效果，可以看出，3 种方法均能检测出图 5(a)、图 5(b)中的人体异常行为，其中基于经典 YOLOv3 的检测方法出现了误检，其将站立的人的影子识别为摔倒对象。表 2 列出了 3 种方法对图 5(a)、图 5(b)检测的置信度对比情况，可以看出，YOLOv3-MSSE 的检测置信度均最高，依次为 0.7 和 0.9，远高于经典 YOLOv3 的置信度值。从总体趋势来看，随着模型不同程度的改进，模型检测对象的置信度逐渐提高。

表 2　检测置信度的对比情况

Table 2　Test confidence comparison

| Confidence | YOLOv3 | YOLOv3-SE | YOLOv3-MSSE |
|---|---|---|---|
| picture(a) | 0.40 | 0.40 | 0.70 |
| picture(b) | 0.75 | 0.84 | 0.90 |

表 3 列出了 3 种网络架构的检测指标对比结果。可以看出，同时添加了注意力机制和特征检测层的 YOLOv3-MSSE 网络架构与仅添加了注意力机制的 YOLOv3-SE 网络架构在 P 指标上略低于经典 YOLOv3 网络模型，但其 R，mAP 以及 F1-scores 指标值均优于经典 YOLOv3，相比经典 YOLOv3，改进后的 YOLOv3-SE 和 YOLOv3-MSSE 模型在 R 指标上分别提升了 5.4% 和 19.1%，在 mAP 指标上分别提升了 6.4% 和 20.8%，在 F1-scores 指标上分别提升了 3.2% 和 11.3%，说明 YOLOv3-MSSE 在一定程度上能够有效提高模型的检测性能，并能较好地平衡其检测精确率与召回率之间的关系。

表 3　3 种网络架构的检测指标对比

Table 3　Comparison of detection indexes of three network architectures

| Network structure | $P$ | $R$ | $mAP/\%$ | $F1$-scores |
|---|---|---|---|---|
| YOLOv3 | 0.920 | 0.603 | 57.3 | 0.728 |
| YOLOv3-SE | 0.908 | 0.657 | 63.7 | 0.760 |
| YOLOv3-MSSE | 0.899 | 0.794 | 78.1 | 0.841 |

**结束语**　YOLOv3-MSSE 方法通过增大检测模型的特征检测层尺寸来实现对大目标的有效检测，提高了模型异常行为检测的精度。同时，该方法在 YOLOv3 目标检测网络结构图中融入了注意力机制，自动学习特征图各个通道的特征权重，再将其结果反馈到输入特征图中，使得网络有选择性地增强信息量大的特征并对无用特征进行抑制，在不引入过多参数的条件下实现了模型检测性能的大幅度提升，对人体异常行为检测处理有着积极的作用。

但是，本文提出的 YOLOv3-MSSE 方法仍有以下不足：

1)不同场景下人体异常行为的定义不同,这使得人体异常行为检测的研究缺乏通用性;2)实验中可用于训练的人体异常行为数据集较少,使得深度神经网络模型训练得不充分。因此,如何扩大人体异常行为数据集,针对不同场景下的人体异常行为训练适应性更强的检测模型,都是下一步值得重点关注的研究内容。

## 参 考 文 献


[1] LENTZAS A, VRAKAS D. Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review [J]. Artificial Intelligence Review, 2020, 53(3): 1975-2021.

[2] XU T, TIAN C Y, LIU C H. Deep Learning for Abnormal Crowd Behavior Detection: A Review[J]. Computer Science, 2021, 48(9): 125-134.

[3] FAN Z, YIN J, SONG Y, et al. Real-time and accurate abnormal behavior detection in videos [J]. Machine Vision and Applications, 2020, 31(7): 1-13.

[4] WANG Y N, LUO J J, WANG D W. An abnormal behavior detection algorithm based on key frame and amplitude histogram entropy [J]. Computer and Digital Engineering, 2019, 47(9): 2281-2285.

[5] SATYBALDINA D Z, GLAZYRINA N S, KALYMOVA K A, et al. Development of an algorithm for abnormal human behavior detection in intelligent video surveillance system [J]. IOP Conference Series: Materials Science and Engineering, 2021, 1069(1): 12046-12054.

[6] WU C F, CHENG Z X, JIANG Y Z. A Novel Detection Framework for Detecting Abnormal Human Behavior [J]. Mathematical Problems in Engineering, 2020, 2020(1): 1-9.

[7] ZHENG H, LIU J F, LIAO M Y. Detection and recognition of abnormal human behavior based on hybrid algorithm under indoor video surveillance [J]. Computer Applications and Software, 2019, 36(7): 224-230, 241.

[8] LI X Q, SUN S L. Research on abnormal behavior detection based YOLO network [J]. Electronic Design Engineering, 2018, 26(20): 160-164, 170.

[9] LUO F B, WANG P, LIANG S Y, et al. Recognition of abnormal crowd behavior based on deep learning and sparse optical flow [J]. Computer Engineering, 2020, 46(4): 287-293, 300.

[10] MA Y D, LUO Z J, NI Z F, et al. Improved SSD algorithm for multi-target detection [J]. Computer Engineering and Applications, 2020, 966(23): 29-36.

[11] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE Xplore, 2016: 770-778.

[12] HU J, SHEN L, ALBANIE S, et al. Squeeze and Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[13] KANG Y, CUI G R, LI H, et al. Software requirement clustering algorithm integrating self-attention mechanism and multi-pyramid convolution [J]. Computer Science, 2020, 47(3): 48-53.

[14] WANG S Y, WANG J J. Research on real time detection method of high density crowd target based on improved yolov3 algorithm [J]. Safety and Environmental Engineering, 2019, 26(5): 194-200.

[15] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE Xplore, 2018: 1086-1091.

[16] ZOU Z W, QIN C. Method of dynamically constructing spatial topic R-tree based on k-means++[J]. Journal of Computer Applications, 2021, 41(3): 733-737.

[17] ZHANG F K, YANG F, LI C. Fastvehicle detection method based on improved YOLOv3 [J]. Computer Engineering and Application, 2019, 55(2): 12-20.

**ZHANG Hong-min**, born in 1970, Ph.D, professor, is a member of China Computer Federation and ACM. His main research interests include image processing and pattern recognition.

(责任编辑:李亚辉)