

本文引用格式：卢杏坚,杨丹妮,焦泽昱.基于改进 YOLOv5 的科技项目评审过程人员行为分析方法[J].自动化与信息工程, 2024,45(6):64-72.

LU Xingjian, YANG Danni, JIAO Zeyu. Analysis method of personnel behavior in the technology project review process based on improved YOLOv5[J]. Automation & Information Engineering, 2024,45(6):64-72.

基于改进 YOLOv5 的科技项目评审过程 人员行为分析方法*

卢杏坚 杨丹妮 焦泽昱

(广东省科学院智能制造研究所/广东省现代控制技术重点实验室, 广东 广州 510070)

摘要: 为提高科技项目评审会议组织实施的质量, 规范评审过程中参会人员的行为, 提出一种基于改进 YOLOv5 的科技项目评审过程人员行为分析方法, 实时分析评审会议现场监控视频数据, 识别参会人员的违规行为。首先, 基于改进的 YOLOv5 构建监控视频小目标检测网络, 在 YOLOv5 主干网络中融合 TCANet 注意力机制, 获取评审会议现场监控视频数据中重点关注的目标区域, 并在其头部网络增加了特征图上采样处理, 将上采样得到的特征图与主干网络中的浅层特征图进行融合, 实现评审会议现场中手机、名片等小目标的检测; 然后, 提出参会人员行为分析算法, 通过人体目标跟踪网络模型实时跟踪参会人员的移动轨迹, 建立区域属性与专家位置域联合的时空关联关系判别式, 识别参会人员与专家接触、攀谈等违规行为。实验结果表明, 该方法对评审会议现场中手机、名片小目标的检测准确率为 0.657, 相比于 YOLOv5m, mAP 提升了 0.196; 参会人员的跟踪准确率 Rank-1 达到 0.938, 图像处理帧率为 21 F/s, 能够准确识别参会人员接触、攀谈行为, 对评审会议现场人员行为智能化管理具有重要意义。

关键词: 科技项目评审; 人工智能; 行为分析; 目标检测; 目标跟踪

中图分类号: TP3

文献标志码: A

文章编号: 1674-2605(2024)06-0010-09

DOI: 10.3969/j.issn.1674-2605.2024.06.010

开放获取

Analysis Method of Personnel Behavior in the Technology Project Review Process Based on Improved YOLOv5

LU Xingjian YANG Danni JIAO Zeyu

(Institute of Intelligent Manufacturing, Guangdong Academy of Sciences/
Guangdong Key Laboratory of Modern Control Technology, Guangzhou 510070, China)

Abstract: To enhance the quality of organizing and implementing technology project review meetings and to regulate participants' behavior during the review process, a behavior analysis method based on an improved YOLOv5 is proposed. This method enables real-time analysis of surveillance video data from review meetings to identify participants' violations. First, an improved YOLOv5-based small-object detection network is constructed for monitoring video data. By integrating the TCANet attention mechanism into the YOLOv5 backbone network, the model focuses on key target areas within the surveillance footage of review meetings. Additionally, the head network incorporates an upsampling process, where the upsampled feature maps are fused with shallow feature maps from the backbone network to achieve detection of small objects such as mobile phones and business cards in the meeting environment. Next, a participant behavior analysis algorithm is proposed. Using a human target tracking network model, the system tracks participants' movement trajectories in real time. A spatiotemporal correlation model is established by combining regional attributes with the spatial domain of expert locations, enabling the detection of participant behaviors, such as interactions and

conversations with experts, which may constitute violations. Experimental results demonstrate that the method achieves a detection accuracy of 0.657 for small objects like mobile phones and business cards, with a *mAP* improvement of 0.196 compared to YOLOv5m. The participant tracking accuracy reaches 0.938, with an image processing frame rate of 21 frames per second (F/s). This approach effectively identifies participant behaviors such as contact and conversation, making significant contributions to the intelligent management of participant behavior during review meetings.

Keywords: technology project review; artificial intelligence; behavioral analysis; object detection; target tracking

0 引言

科学技术是第一生产力,我国高度重视科学技术,在政策、资金方面均给予大力支持。为规范科学技术活动,营造风清气正的科研环境,我国制定了《科学技术活动违规行为处理暂行规定》等法律法规,确保科技项目的高质量发展与公平竞争。科技项目评审会议作为评估项目可行性的重要环节,规范参会人员的行为,正是落实《广东省科学技术厅科技业务评审会议现场监督工作规程(试行)》的要求。我国每年都有多批次、大量的科技项目申报,评审会议频次高,参会人员众多。采用人工甄别评审会议现场参会人员使用手机、发放名片、与评审专家接触、攀谈等违规行为十分繁琐。如何快速识别参会人员的违规行为,亟需通过技术手段来解决。

近年来,人工智能技术在人体行为分析、目标检测、目标跟踪等图像处理领域取得了重要突破^[1]。同时,评审会议现场基本配备了视频监控系統,能够获取监控视频数据。因此,人工智能和视频大数据技术的综合应用,可为科技项目评审过程人员行为分析提供技术方案。

专家学者们在人员行为分析领域开展了大量的研究。文献[2]引入三维卷积技术,利用视频的时序信息进行学生学习行为的识别和分析,但三维卷积神经网络的计算量较大。文献[3]提出一种融合区域生成网络(region proposal network, RPN)的行为识别算法,通过目标检测机制引导网络关注和学习人体的行为信息,减少了无关信息的干扰,但 RPN 会增加网络的运算时间。文献[4]提出基于 Alphapose 优化模型的老人跌倒行为检测算法,对行人目标检测模型和姿态

估计模型进行加速优化,并通过分析人体的结构化特征识别跌倒行为。文献[5]提出一种扩展的特征金字塔网络,设计了基于特征引用的超分辨模块,为扩展的特征金字塔赋予可靠的细节信息。文献[6]采用高层次的抽象特征作为上下文特征,通过连接上下文特征和小目标特征,增加小目标信息,以便网络能更好地检测小目标。文献[7]基于视频序列中的时间结构,利用目标的运动上下文背景和在多帧上聚合的时空特征来提高检测性能。文献[8]提出基于双光流网络的视频目标检测算法,利用两种不同的光流网络,分别估计位移较近和位移较远的当前帧与近邻帧之间的光流场,进而捕获不同尺度的多帧特征。目前,尚未见人工智能技术应用于科技项目评审过程人员行为分析的相关研究。

本文提出一种基于改进 YOLOv5^[9]的科技项目评审过程人员行为分析方法。通过监控视频小目标检测网络,融合时间上下文聚合网络(temporal context aggregation network, TCANet)注意力机制^[10-11],实现对科技项目评审会议现场中手机、名片等小目标的检测;提出参会人员行为分析算法,通过人体目标跟踪网络模型,实时跟踪参会人员的移动轨迹,建立区域属性与专家位置域联合的时空关联关系判别式,用于识别参会人员与专家接触、攀谈等违规行为,以推动科技项目评审过程的智能化管理。

1 监控视频小目标检测网络

在科技项目评审过程中,参会人员利用手机通信、发放名片等方式与管理人員或评审专家进行非必要的联系是违规行为。由于手机、名片在监控视频中以小目标的形式出现,图像特征不明显,且容易被遮挡,

使人工检测监控视频的准确率、效率均较低。为此，通过在 YOLOv5 主干网络中融合 TCANet 注意力机制，获取科技项目评审会议现场监控视频中重点关注的目标区域，以实现对手机、名片等小目标的检测。

1.1 TCANet 注意力机制

融合 TCANet 注意力机制的目的是将注意力集中在人体和动态变化的前景目标上，既能捕获人体的方向和位置感知信息，也能提取图像通道间的信息。TCANet 注意力机制的输入为 3 通道的图像数据，设图像的宽、高分别为 W, H ，输入的图像数据记为 Da ， Da 分别经过空间坐标信息网络模块和通道信息网络模块来提取特征。TCANet 注意力机制的结构如图 1 所示。

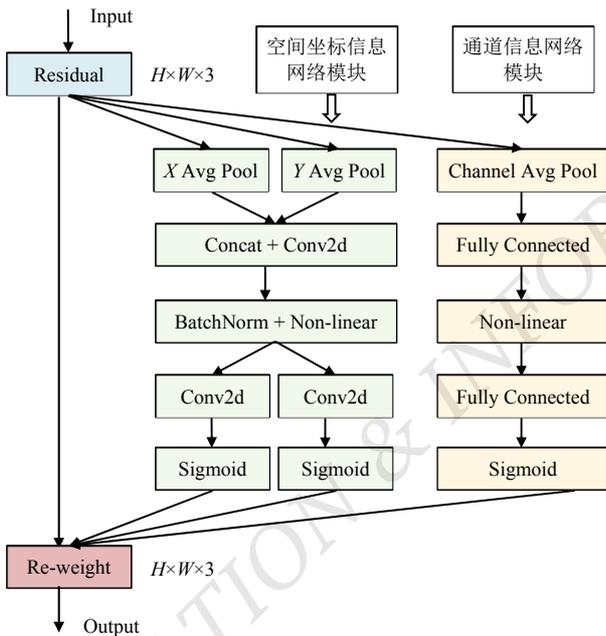


图 1 TCANet 注意力机制的结构

1.1.1 通道信息网络模块

Da 经通道信息网络模块进行压缩操作，提取图像的全局信息，压缩操作的公式为

$$o_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W d_c(i, j) \quad (1)$$

式中： o_c 为第 c 个通道的输出； $d_c(i, j)$ 为第 c

个通道图像位置坐标为 (i, j) 的图像数据，输入图像为 3 通道，即 $c \in [1, 2, 3]$ 。

通过激励操作，获取通道之间的依赖关系。

$$\hat{D} = D \cdot \delta(\hat{o}) \quad (2)$$

式中： \hat{D} 、 D 分别为通道信息网络模块的输出、输入数据， δ 为 Sigmoid 函数， $\hat{o} = T_2(\text{ReLU}(T_1(o)))$ ， T_1 、 T_2 为可学习的线性变换。

1.1.2 空间坐标信息网络模块

首先，利用 $(H, 1)$ 和 $(1, W)$ 的池化层，沿着图像的 X 、 Y 坐标方向对每个通道进行编码，得到高度为 h 的第 c 个通道输出 $o_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} d_c(h, i)$ ，宽度为 w 的第 c 个通道输出 $o_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} d_c(j, w)$ 。

TCANet 注意力机制能够捕获图像沿着一个空间方向的依赖关系，同时保留图像沿着另一个空间方向的精确位置信息，从而更准确地定位感兴趣的目标。

然后，通过 Concat 操作级联 $o_c^h(h)$ 和 $o_c^w(w)$ ，再经过 1×1 的卷积进行变换 F_1 ，生成空间信息在 X 、 Y 坐标方向的特征图 f 。

接着，沿空间维度将 f 切分为 2 个单独的张量 $f^h \in \mathbb{R}^{3/r \times H}$ 、 $f^w \in \mathbb{R}^{3/r \times W}$ ，其中 r 为下采样比例系数。

最后，利用 2 个 1×1 卷积变换张量，分别得到空间坐标信息网络模块和通道信息网络模块的输出结果：

$$\begin{aligned} g^h &= \delta(F_h(f^h)) \\ g^w &= \delta(F_w(f^w)) \end{aligned} \quad (3)$$

将空间坐标信息网络模块和通道信息网络模块的输出结果进行加权融合，得到 TCANet 注意力机制提取的目标图像特征 F ：

$$F = g^h + g^w + \hat{D} \quad (4)$$

1.2 改进的 YOLOv5 网络

利用 TCANet 注意力机制替换 YOLOv5 主干网络中的 BottleneckCSP 模块，通过多个 TCANet 注意力机制堆叠，提取空间坐标信息和通道信息，定位重点关注的目标区域，获取更多需要关注的细节信息。

在 YOLOv5 头部网络增加了一层上采样层，用于对特征图进行上采样处理，以放大特征图；同时，将

上采样得到的特征图与主干网络的第一个 TCANet 注意力机制提取的浅层特征图进行融合，以获得更丰富的特征信息，从而实现对科技项目评审会议现场监控视频中手机、名片等小目标的检测。改进的 YOLOv5 网络结构如图 2 所示。其中，虚线区域表示增加的上采样层和特征融合过程。

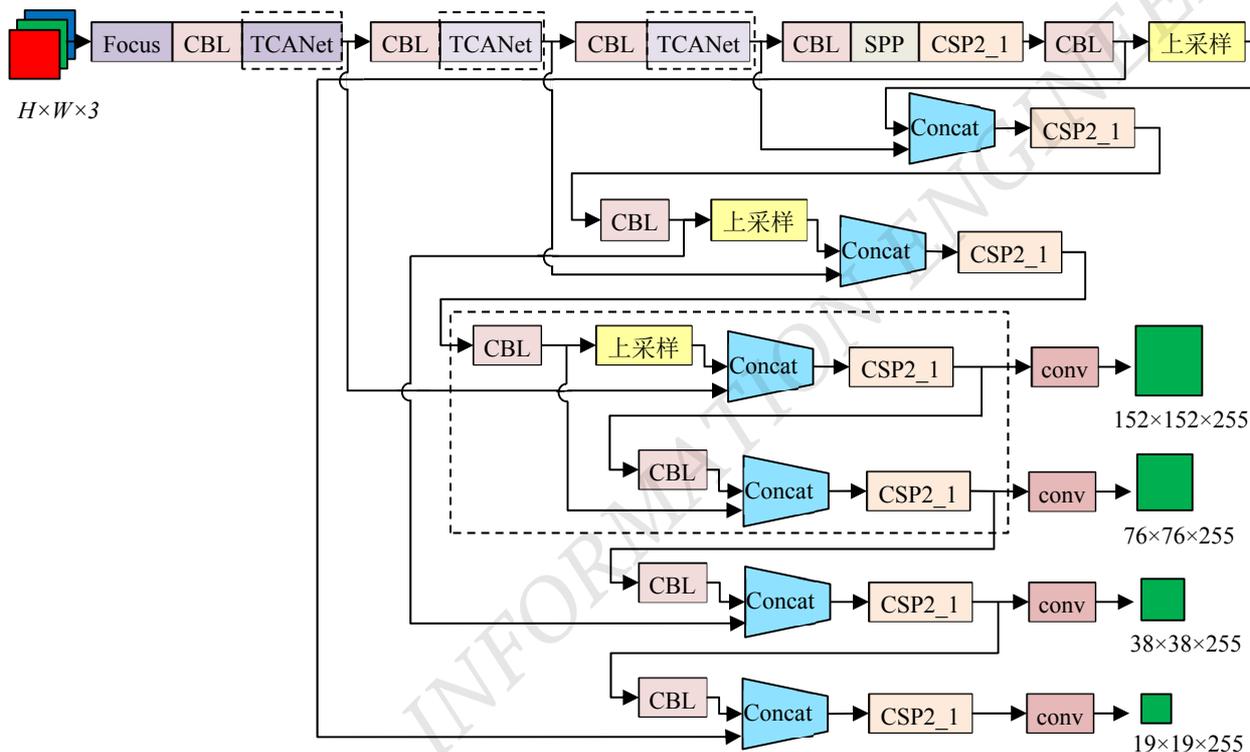


图 2 改进的 YOLOv5 网络结构

2 参会人员行为分析算法

在科技项目评审过程中，会议现场严格限制参会人员在规定时间内随意进出，并禁止参会人员在非指定区域与评审专家接触、攀谈等违规行为。参会人员行为分析算法利用人体目标跟踪网络模型，实时跟踪参会人员的移动轨迹；同时，通过区域属性与专家位置域联合的时空关联关系判别式进行参会人员行为分析，识别参会人员与专家接触、攀谈等违规行为。

2.1 人体目标跟踪网络模型

人体目标跟踪本质上是一个人体重识别 (person re-identification, ReID) 任务。本文构建的人体目标跟踪网络模型利用改进的 YOLOv5 网络，获得评审会

议现场监控视频中的人体目标检测框；人体目标检测框经过滤预处理后，输入到包括实例批次归一化 (instance batch normalization, IBN)^[12]模块和 Non-local 模块^[13]的卷积神经网络中，以提升模型对不同环境和人体目标尺寸的鲁棒性。其中，IBN 模块可提升模型在测试数据与训练数据差别较大时的泛化能力；Non-local 模块引入了注意力机制，用于捕获图像中远距离的依赖关系，以实现视觉场景的全局理解。本文采用基于部分的卷积基线 (part-based convolutional baseline, PCB)^[14]行人重识别 (ReID_PCB) 模型作为基线模型，输出人体目标图像的特征。人体目标跟踪网络模型结构如图 3 所示。

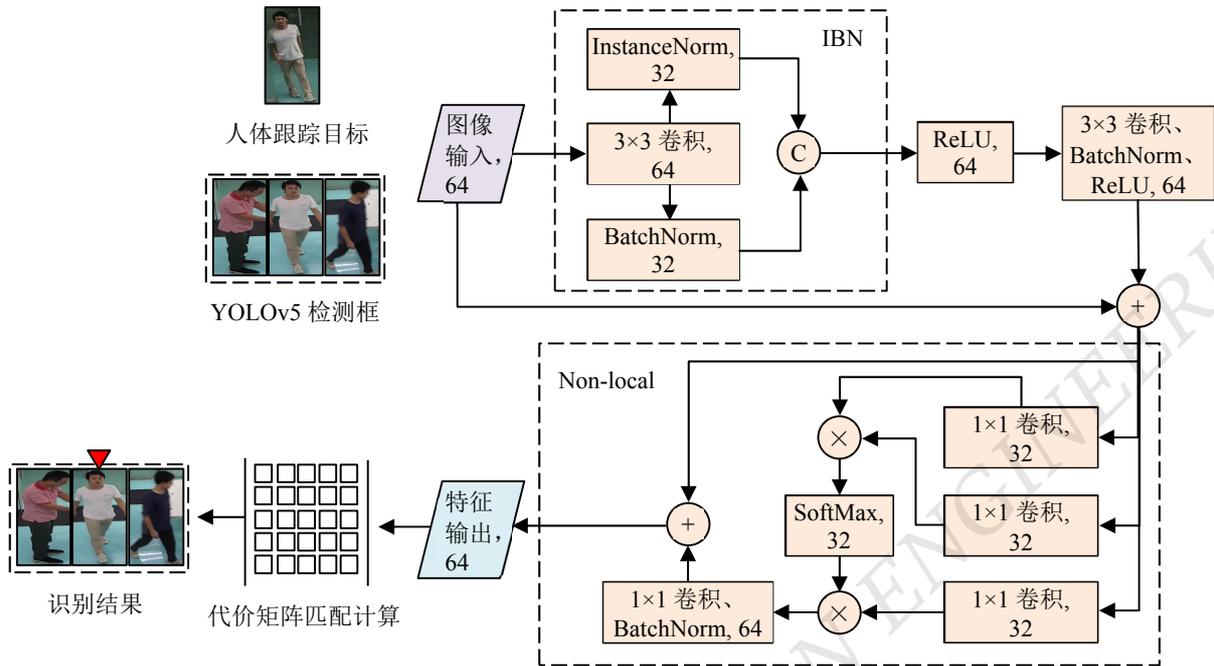


图3 人体目标跟踪网络模型结构

在图3中, 32、64 分别为各层输出特征图的通道数。通过计算两帧之间人体目标特征的余弦距离及过滤后的人体目标检测框的交并比, 可以获得代价矩阵; 利用匈牙利算法对代价矩阵进行匹配计算, 可获得跨帧的人体目标; 将前一帧的跟踪框与当前帧的人体目标信息进行关联, 并利用卡尔曼滤波器来预测人体目标在下一帧的位置, 从而确认其是否处于被跟踪状态; 若人体目标处于被跟踪状态, 则分配相应的跟踪 ID, 进而实现评审会议现场监控视频中人体目标的连续跟踪。

2.2 参会人员行为分析

科技项目评审会议现场通常划分了专家落座区、项目人员区、等候区等。参会人员需到对应的区域参与项目评审活动, 并禁止跨区域活动。基于参会人员的移动轨迹, 建立区域属性与专家位置域联合的时空关联关系判别式, 通过分析参会人员的空间特征和时间特征, 统计其在专家位置域的停留时间, 识别其与专家接触、攀谈等违规行为。

区域属性与专家位置域联合的时空关联关系判别式为

$$D(a, b) = \int_T f_a(t) \cdot g_b(t) dt \quad (5)$$

式中: $D(a, b)$ 为参会人员 a 与专家 b 的行为关联度, $f_a(t)$ 为参会人员 a 在时间 t 的空间状态 (如位置信息), $g_b(t)$ 为专家 b 的空间状态, T 为行为分析的时间区间。

通过设定的阈值对 $D(a, b)$ 进行判断。当 $D(a, b)$ 超过阈值时, 利用改进的 YOLOv5 网络检测参会人员的移动轨迹上是否出现名片、手机, 并以此为依据识别参会人员的违规行为。参会人员行为分析示意图如图4所示。

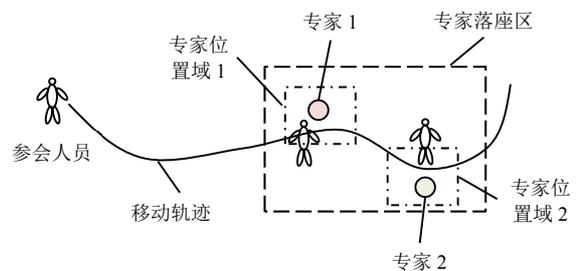


图4 参会人员行为分析示意图

3 实验与分析

实验硬件环境: Intel(R) Xeon(R) CPU E5-2650 v4 @2.20 GHz, 内存 64 GB, GTX1080Ti GPU。软件环

境：64 位 Ubuntu18.04 操作系统，OpenCV4.7、TensorFlow2.3、PyTorch1.7 等工具，开发语言 Python3.7。

实验数据来源于 1080P 监控摄像头拍摄的模拟评审会议现场的监控视频，视频内容包括参会人员、专家、桌椅、名片、手机等目标；参会人员向专家派发名片、专家使用手机、参会人员与专家接触等行为。

3.1 小目标检测网络性能评估实验

设置训练循环数 epochs 为 200 次，批尺寸 (batch size) 为 16 个，初始学习率为 0.001，循环学习率为 0.1，图像分辨率为 640×640 像素，预训练网络为 YOLOv5m。性能评价指标分别为 mAP@0.5、loss、准确率 (P)、帧率 (FR)。其中，mAP@0.5 表示所有目标类别的 IoU 阈值在 0.5 时的平均检测精度，可以反映算法对不同类别目标的检测精度。

3.1.1 消融实验

通过消融实验，验证 TCANet 注意力机制、上采样层对 YOLOv5m 网络性能的影响，实验结果如表 1 所示。

表 1 小目标检测网络消融实验结果

| 网络 | mAP@0.5 | P | loss | FR/(F/s) |
|---------------------|--------------|--------------|--------------|--------------|
| YOLOv5m | 0.335 | 0.430 | 0.024 | 55.55 |
| YOLOv5m+TCANet | 0.383 | 0.478 | 0.021 | 43.48 |
| YOLOv5m+TCANet+上采样层 | 0.531 | 0.657 | 0.010 | 33.33 |

由表 1 可以看出：YOLOv5m 网络引入 TCANet 注意力机制后，mAP@0.5、P 均提高了 0.048，loss 减少了 0.003；同时引入 TCANet 注意力机制和上采样层后，P、loss、mAP@0.5 比其他对比网络均有优势，虽然 FR 有所下降，但 FR 为 33.33 F/s，基本可以满足实时检测的需求。因此，消融实验证明了利用 TCANet 注意力机制可以定位重点关注的小目标区域，以获取更多的细节信息；增加上采样层，可以放大特征图，以获得更丰富的特征进行小目标检测。

3.1.2 与其他目标检测网络对比

为了验证改进的 YOLOv5 网络的有效性，将其与其他典型的目标检测网络 (YOLO 系列、RCNN^[15]家

族、SSD 网络、TPH-YOLOv5^[16]、Deformable DETR^[17]) 进行对比实验。评价指标有 mAP@0.5、mAP@0.5:0.95、FR，实验结果如表 2 所示。

表 2 10 种目标检测网络的对比实验结果

| 网络 | mAP@0.5 | mAP@0.5:0.95 | FR/(F/s) |
|-----------------|--------------|--------------|-----------|
| SSD | 0.194 | 0.126 | 10 |
| Faster-RCNN | 0.332 | 0.170 | 20 |
| Light-RCNN | 0.328 | 0.165 | 15 |
| YOLOv3 | 0.321 | 0.169 | 13 |
| YOLOv4 | 0.345 | 0.176 | 14 |
| YOLOv5m | 0.335 | 0.186 | 55 |
| RefineDet | 0.288 | 0.149 | 14 |
| Deformable DETR | 0.543 | 0.311 | 15 |
| TPH-YOLOv5 | 0.524 | 0.293 | 20 |
| 本文网络 | 0.531 | 0.305 | 33 |

由表 2 可知：本文网络相较于 Deformable DETR，虽然 mAP@0.5 降低了 0.012，但检测速率提升了 120%；且 Deformable DETR 的参数量较大，对存储空间要求高，限制了其实际应用。因此，对比实验证明了本文网络的有效性和优越性，提高了对手机、名片等小目标的检测精度；同时，本文网络基于 YOLO 系列优化而来，技术应用落地具有较好的可实施性。

3.2 人体目标跟踪实验

设置训练 epochs 为 100 次，批尺寸为 64 个，初始学习率为 0.01，循环学习率为 0.1，图像分辨率为 384×128 像素，采用随机梯度下降优化器，动量为 0.85，权重衰减为 5×10^{-4} 。训练数据集为 Market-1501^[18]，评价指标包括首次 (Rank-1) 成功匹配概率、mAP 和 FR。

3.2.1 消融实验

通过消融实验来分析 IBN、Non-local 模块对人体目标跟踪网络模型性能的影响，实验结果如表 3 所示。

表 3 人体目标跟踪消融实验

| 模型 | Rank-1 | mAP |
|----------------------|--------------|--------------|
| YOLOv5 | 0.915 | 0.782 |
| YOLOv5+IBN | 0.923 | 0.795 |
| YOLOv5+IBN+Non-local | 0.938 | 0.816 |

由表 3 可知: 在 YOLOv5 模型的基础上引入 IBN 模块后, Rank-1、 mAP 分别提升了 0.008、0.013; 引入 IBN 和 Non-local 模块后, Rank-1、 mAP 分别提升了 0.023 和 0.034, 表明 IBN 模块可提升模型的泛化能力, 减少测试数据与训练数据差距较大的影响; Non-local 模块可提高网络的注意力, 有助于人体目标特征的提取。

3.2.2 与其他目标跟踪网络模型对比

将本文提出的人体目标跟踪网络模型与目前主流的目标跟踪网络模型 (PABR^[19]、PCB^[14]、SNR^[20]、AlignedReID^[21]、HOREID^[22]) 进行对比实验, 实验结果如表 4 所示。

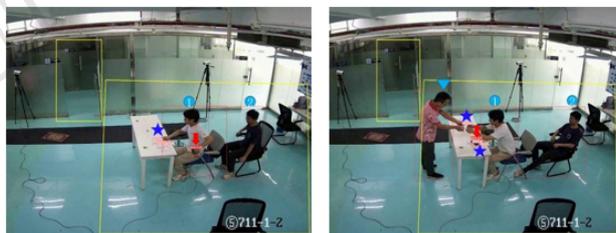
表 4 6 种目标跟踪网络模型对比

| 网络模型 | Rank-1 | mAP | FR/ (F/s) |
|-------------|--------------|--------------|-----------|
| PABR | 0.828 | 0.729 | 7 |
| PCB | 0.917 | 0.801 | 16 |
| SNR | 0.936 | 0.813 | 9 |
| AlignedReID | 0.944 | 0.848 | 11 |
| HOREID | 0.942 | 0.834 | 12 |
| 本文网络模型 | 0.938 | 0.816 | 21 |

由表 4 可知, 本文网络模型的 Rank-1、 mAP 与最优的 AlignedReID 网络模型相比, 分别降低了 0.006 和 0.032, 但检测速率提升了 91%, 可以满足评审会议现场实时性的需求, 具有较强的可行性和应用价值。

3.3 识别效果可视化

本文提出的参会人员行为分析算法的识别效果可视化结果如图 5、6 所示。评审会议现场手机、名片小目标的识别效果图如图 5 所示。其中, 下箭头“ \Downarrow ”表示手机, 五角星“ \star ”表示名片, 下三角“ ∇ ”表示参会人员, 圆圈“ \circ ”表示专家, 图 5 中有两位专家。



(a) 监控视频小目标识别效果一 (b) 监控视频小目标识别效果二



(c) 监控视频小目标识别效果三 (d) 监控视频小目标识别效果四

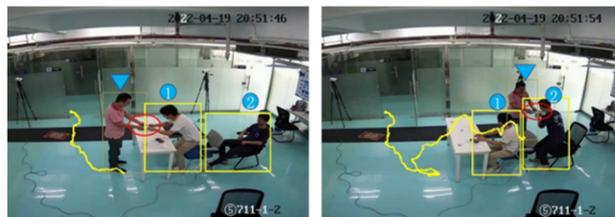
图 5 评审会议现场手机、名片小目标识别效果图

在图 5(a)中, 从监控视频中仅能看到手机的边缘部分, 本文算法可准确识别出专家握在手中的手机。

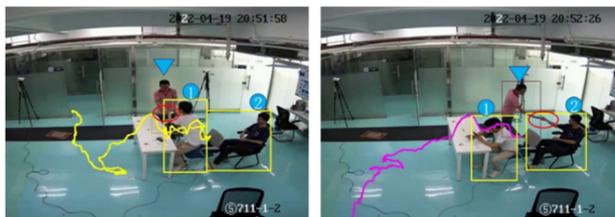
在图 5(b)、(c)中, 参会人员向专家派发被手部遮挡的名片, 本文算法可准确识别桌面上部分重叠的手机与名片。

在图 5(d)中, 本文算法能准确识别参会人员手握的、可视部分较小名片和手机。

参会人员与专家接触、攀谈等违规行为识别的目标跟踪效果如图 6 所示。其中, 矩形框表示专家位置域, 下三角“ ∇ ”表示参会人员, 圆圈“ \circ ”表示专家, 图 6 中有两位专家, 曲线为参会人员的移动轨迹。



(a) 违规行为案例一 (b) 违规行为案例二



(c) 违规行为案例三 (d) 违规行为案例四

图 6 参会人员与专家接触、攀谈等违规行为识别的目标跟踪效果图

在图 6(a)、(d)中, 参会人员的移动轨迹逐渐靠近专家, 且短时停留, 结合改进的 YOLOv5 网络, 发现两者手部区域存在名片, 即可判定存在违规行为。

在图 6(b)、(c)中, 参会人员的移动轨迹在专家区

域附近逐渐密集,反映参会人员在专家区域停留的时间较长,结合改进的 YOLOv5 网络,判断参会人员与专家存在接触。

4 结论

为规范科技项目评审过程中参会人员的行为,提高科技项目的管理水平,本文提出基于改进 YOLOv5 的科技项目评审过程人员行为分析方法。利用改进的 YOLOv5 网络检测评审会议现场的手机、名片等小目标;通过参会人员行为分析算法跟踪、识别参会人员接触、攀谈等违规行为。实验结果表明,本文方法能够准确识别评审会议现场中手机、名片等小目标,以及参会人员与专家接触、攀谈等违规行为,对科技项目的智能化管理有重要意义。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] 曹家乐,李亚利,孙汉卿,等.基于深度学习的视觉目标检测技术综述[J].中国图象图形学报,2022,27(6):1697-1722.
- [2] Zhang R, Ni B. Learning behavior recognition and analysis by using 3D convolutional neural networks[C]. 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST). IEEE, 2019:1-4.
- [3] 周波,李俊峰.结合目标检测的人体行为识别[J].自动化学报,2020,46(9):1961-1970.
- [4] 马敬奇,雷欢,陈敏翼.基于 AlphaPose 优化模型的老人跌倒行为检测算法[J].计算机应用,2022,42(1):294-301.
- [5] DENG C, WANG M, LIU L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2021,24:1968-1979.
- [6] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[C]. 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). IEEE, 2021:181-186.
- [7] KIM J, KOH J, LEE B, et al. Video object detection using object's motion context and spatio-temporal feature aggregation [C]. 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021:1604-1610.
- [8] 尉婉青,禹晶,史薪琪,等.双光流网络指导的视频目标检测[J].中国图象图形学报,2021,26(10):2473-2484.
- [9] ZHU Xingkui, LYU Shuchang, WANG Xu, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. The IEEE/CVF International Conference on Computer Vision (ICCV), 2021:2778-2788.
- [10] QING Zhiwu, SU Haisheng, GAN Weihao, et al. Temporal context aggregation network for temporal action proposal refinement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:485-494.
- [11] 崔海朋,姜英昌.基于注意力机制的海上小目标重识别方法[J].机电工程技术,2022,51(7):100-103.
- [12] PAN Xingang, LUO Ping, SHI Jianping, et al. Two at once: Enhancing learning and generalization capacities via IBN-Net [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018:464-479.
- [13] CAO Yue, XU Jiarui, LIN Stephen, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019
- [14] SUN Y, ZHENG L, YANG Y. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.
- [15] MA S, SONG Y, CHENG N, et al. Structured light detection algorithm based on deep learning[C]//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2019,252 (4):042050.
- [16] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [17] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection[J]. arXiv preprint arXiv:2010.04159, 2020.
- [18] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015:1116-1124.
- [19] SUH Y, WANG J, TANG S, et al. Part-aligned bilinear representations for person re-identification[C] //Proceedings of the European Conference on Computer Vision (ECCV), 2018:

402-419.

- [20] JIN X, LAN L, ZENG W J, et al. Style normalisation and restitution for generalisable person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 3143-3152.
- [21] 张勃兴,张寿明,钟震宇.基于多粒度特征融合网络的行人

重识别[J].光电子·激光,2022,33(9):977-983.

- [22] WANG G N, YANG S, LIU H Y, et al. High-order information matter: Learning relation and topology for occluded person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6449-6458.

作者简介:

卢杏坚,男,1987年生,本科,助理研究员,主要研究方向:机器视觉、人工智能。

杨丹妮,女,1998年生,硕士研究生,主要研究方向:机器视觉、人工智能。

焦泽昱(通信作者),男,1991年生,博士研究生,副研究员,主要研究方向:机器视觉、人工智能。E-mail: zy.jiao@giim.ac.cn

~~~~~

(上接第 58 页)

- [5] 王子玉,陈佳星,白博文,等.基于改进 YOLOv5s 的绝缘子缺陷检测方法[J].机电工程技术,2024,53(8):197-200.
- [6] 邢祥胜,刘林,刘晓宇,等.工业机器人火箭喷管冷却通道线激光在位测量系统开发[J].中国测试,2023,49(12):1-7.
- [7] HE Y, SONG K, DONG H, et al. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network[J]. Optics and Lasers in Engineering, 2019,122:294-302.
- [8] HA H, JEONG J. CNN-based defect inspection for injection molding using edge computing and industrial IoT systems[J]. Applied Sciences, 2021, 11(14): 6378.
- [9] NAGATA F, TOKUNO K, MITARAI K, et al. Defect detection method using deep convolutional neural network, support vector machine and template matching techniques[J]. Artificial Life and Robotics, 2019,24(4):512-519.
- [10] LEE M, JEON J, LEE H. Explainable AI for domain experts: A post Hoc analysis of deep learning for defect classification of TFT-LCD panels[J]. Journal of Intelligent Manufacturing, 2022, 33(6):1747-1759.

- [11] CHEN H, HU Q, ZHAI B, et al. A robust weakly supervised learning of deep Conv-Nets for surface defect inspection[J]. Neural Computing and Applications, 2020,32(15):11229-11244.
- [12] 陶显,侯伟,徐德.基于深度学习的表面缺陷检测方法综述[J].自动化学报,2021,47(5):1017-1034.
- [13] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning[C]. Proceedings of the IEEE, 2020,109(1): 43-76.
- [14] 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.
- [15] GAO Y, GAO L, LI X, et al. A semi-supervised convolutional neural network based method for steel surface defect recognition[J]. Robotics and Computer-Integrated Manufacturing, 2020,61:101825.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON.G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012,25: 84-90.

#### 作者简介:

杨志明,男,1974年生,本科,工程师,主要研究方向:智能制造。

黄天仑(通信作者),男,1991年生,博士研究生,高级工程师,主要研究方向:机器视觉。E-mail: huang\_tianlun@126.com

薛艳,女,1989年生,本科,高级工程师,主要研究方向:液压元件、液压系统。

刘家欢,男,1995年生,博士研究生,工程师,主要研究方向:智能检测技术。

钟震宇,男,1971年生,博士研究生,研究员,主要研究方向:深度学习。

张云,男,1981年生,博士研究生,教授,主要研究方向:注塑成型工艺、智能制造。