DOI:10.20033/j.1003-7241.(2026)07-0123-05

一种混合位姿信息与 RGB 图像的多通道图像自动化导播方法

郭　亮

(92941 部队 45 分队,辽宁 葫芦岛 125000)

摘　要:电视台导播部位应用计算机视觉技术时,时常出现图像识别模型错误识别目标的问题,导致自动化导播方法应用效果不佳、输出结果不可靠,只在辅助和特定应用场景下使用,应用范围受限,繁重的播控工作主要依靠人力堆叠完成。为了减轻导播部位的工作压力、提升系统反应速度,提出一种混合位姿信息与 RGB 图像的多通道图像自动化导播方法。构建 RGB+P(位置)的多模态图像优选排序函数,得到多目标与多通道图像的最优匹配,以消除目标被障碍物部分遮挡、多目标特征相似、新物体未纳入训练数据集等多种原因导致的误识误播问题,在测试中实现了 89.975%的排序结果修正比率,提高了输出结果的可靠性,将自动化导播方法的应用范围推广到通用电视直播场景,实现了导播设备的无人化操管。

关键词:导播;计算机视觉;多模态;图像识别;自动化;电视直播

中图分类号: TP391.413;TP317.4;TN957.52　　**文献标志码**: A　　**文章编号**: 1003-7241(2026)07-0123-05

A multi-channel image automatic guidance method combines position information and RGB images

Guo Liang

(45 Unit of 92941 Troops of PLA, Huludao 125000, Liaoning, China)

Abstract: When computer vision technology is applied in the television station′s broadcasting area, there is often a problem of image recognition models incorrectly identifying targets, resulting in poor application effectiveness and unreliable output results of automatic broadcasting methods. They are only used in auxiliary and specific application scenarios, with limited application scope. The heavy broadcasting and control work mainly relies on manual stacking to complete. In order to reduce the work pressure of the guidance section and improve the system response speed, this paper proposes a multi-channel image automatic guidance method that combines pose information and RGB images. A multimodal image optimization sorting function for RGB+P (position) is constructed to obtain the optimal match between multi-objective and multi-channel images, thereby eliminating misidentification and false broadcast problems caused by various factors such as partial occlusion of the target by obstacles, similarity of features among multiple targets, and new objects not being included in the training dataset. In testing, a sorting result correction ratio of 89.975% is achieved, which improves the reliability of the output results. The application scope of the automated guidance method is extended to general television live streaming scenes, enabling unmanned operation of directing equipment.

Keywords: guiding broadcast; computer vision; multimodal; image identification; automation; live television broadcasting

导播是电视直播任务时负责从多路视、音频信号中选择一路或多路播出到指定通道的工作,具有主观性、时效性、随机性等特点,是一个技术门槛低、重复性操作多、工作压力大的人力密集岗位,只有出错时才被观众感知。近年来随着视听采集、传输技术的飞速发展,导播工作需处理的视频、音频信号成倍增长,直播任务中人工导播跟不上节奏、播出慢半拍的问题愈发突出。

导播岗位从业者一直在研究引入计算机视觉技术,采用目标检测算法代替人眼完成导播工作,以实现自动化导播,减少人力投入、提升导播效果[1-3]。但受制于目标检测算法的准确率、通用性和直播任务中极小的容错空间,研制成果多停留在演示、辅助或特定场景专用而推广应用较少,特别是对于高级别的盛会、赛事、演习等,即使数小时的直播过程中只出现 1 次误选切,对于电视台来说就是直播事故。目前实时性较好的 YOLO 系列模型 YOLOv10-X 在 MS COCO 数据集上的 AP^{val} 为 54.4%[4],虽然存在数据集标注错误和缺陷等客观因素带来的影响[5],其工作可靠性距人工识别模式仍有较大差距,当前基于计算机视觉的自动化导播技术未能够实际应用到电视直播任务中。

直播任务中对于多路图像的优选是较为主观的,没有单一的标准答案,可以见仁见智地呈现不同的风格,对于自动化导播可以通过横向比较多路图像中目标的置信度、

收稿日期:2024-12-20;**录用日期**:2025-02-09

作者简介:郭　亮(1982—),男,工程师,研究方向为视音频信号导播。

引用本文:郭亮. 一种混合位姿信息与 RGB 图像的多通道图像自动化导播方法[J]. 自动化技术与应用, 2026,45(7):123-127. (Guo Liang. A multi-channel image automatic guidance method combines position information and RGB images[J]. Techniques of Automation and Applications, 2026,45(7):123-127.)

大小、位置和图像的灰度均值、平均梯度等指标或选用图像质量评价模型[6]进行排序,但对于图像中是否存在要展示的目标,是对错分明的决定性指标。如只采信 RGB 图像单一模态识别结果,对于目标被障碍物部分遮挡、目标较多且外观一致、出现外观与目标相似且未纳入数据集的新类别物体和多种原因导致的目标模糊等情况,将无法保障稳定的目标识别准确率,限制了自动化导播技术的应用。如果能够引入其他模态的数据,将画幅中无目标的图像提前过滤,再交由图像识别模型进行识别并优选排序,是否能够减少自动化导播的出错概率、提高实际应用效果呢? 基于上述想法,本文构建了一种 RGB+P(位置)的多模态图像优选排序函数,通过位姿、图像静帧检测等信息先验并过滤无目标的图像,降低输出结果的出错概率,再融合 RGB 图像识别结果与镜头、目标的位姿、轨迹、图像静帧检测等模态信息进行优选排序,以提高多通道图像自动化导播方法的应用效果。

1 相关工作

1.1 硬件准备

为摄录目标和摄录设备安装 GPS 定位器或惯性导航仪,实时获取设备与目标的位姿信息。为摄像设备底座加装方位角、俯仰角刻度盘,实时获取设备的方位角、俯仰角信息。需要有线、无线通信网络、时间统一、图像采集、图像传输、图像选切、软件运行服务器等硬件设施。

1.2 衍生类别与数据集

选用实时性较好的图像识别模型,本文选用的是 YOLOv10 开源模型,并根据每次直播任务需求,定制专用的数据集来训练模型,以减少不适用的类别设置和数据集偏差[7]带来的影响,以提高模型的目标识别准确率。在定义数据集的物体类别时,除摄录目标和背景类别外,需考虑摄录目标的衍生类别,即摄录目标在不同摄录条件下所呈现的类别,例如摄录目标是一架飞机,飞机被部分遮挡可能只露一组轮子,飞行时从后方摄录的图像可能是尾焰或尾烟航迹,在红外相机中看到的是光团,距离较远时图像中只是一个点,在夜晚摄录的图像中是机身上的指示灯光,这些在图像识别结果中都不是飞机,但确是飞机在不同摄录条件下图像中的呈现,应结合直播任务的实际情况,在数据集类别分类时增加摄录目标在本次直播任务中涉及的衍生类别,排序时衍生类别范围内的识别物体按照与摄录目标类别的关联度给予一定的得分权重。

1.3 获取位姿信息

实时获取多个目标的位置数据,并统一转换为 CGCS2000 地心空间直角坐标系坐标,加上时间戳后(time, X_T, Y_T, Z_T)以组播方式发送至网络;实时获取多台摄录设备的位置、状态数据,包括坐标、方位角、俯仰角和焦距,并根据摄录设备的视场角指标将焦距换算为当前拍摄图像的水平视场角和俯仰视场角,加上时间戳后(time, X_C, Y_C, Z_C, A_C, E_C, A_R, E_R)以组播方式发送至网络。

1.4 获取 RGB 识别信息

实时采集多台摄录设备拍摄的实况图像,通过模型识别图像,得到 RGB 图像中识别出物体的相关信息,包括识别物体的中心点坐标、宽、高、类别、置信度等加上时间戳后(time, X_P, Y_P, W_P, H_P, C_P, P)以组播方式发送至网络。

1.5 数据同步与对齐

虽然采用了时统设备,但因网络传输延迟差不一致、数据帧率不一致、数据帧不对齐、数据间断、设备内部时钟不准、算力不足产生的处理延时等问题导致的计算误差,是影响排序结果精度的主要因素。需要对数据进行同步对齐处理,一是统一数据的帧率和起始帧的时间,对数据进行帧间插值处理以补全缺失或错位的数据帧,消除多目标、多台摄录设备的位姿和状态数据的帧率不一致、不对齐问题;二是建立数据不采信机制,设置数据间断时间阈值,在数据间断时间超过阈值后不进行插值处理;三是校正位姿数据的时间戳,设置时间校正阈值,比对当前时统时间与数据时间戳的差值,超出阈值则修正数据的时间戳;四是校正模型识别结果的时间戳,根据摄录设备位姿数据的时间戳、编解码处理时延和传输时延,估算并修正模型识别结果的时间戳,解决摄录图像不显示时间码信息和编解码处理后图像帧头信息丢失导致的每帧图像时间真值无法获取的问题。缓存处理后的数据,设置缓存容量和数据清除机制,用于为后续计算提供数据。

1.6 位姿信息校验

依次遍历目标、摄录设备位姿数据,逐个计算得到目标与摄录设备间的摄录距离 d,见式(1)。

$$d = \sqrt{(X_T - X_C)^2 + (Y_T - Y_C)^2 + (Z_T - Z_C)^2} \tag{1}$$

式中,(X_T, Y_T, Z_T) 是目标坐标,(X_C, Y_C, Z_C) 是摄录设备坐标。

构建摄录距离偏离计算公式,见式(2)。

$$S_1 = 1 - \frac{d_i}{\max(d_1, d_2, \cdots d_n)} \tag{2}$$

式中,n 是摄录设备总数,$i = [1, 2, \cdots n]$ 是摄录设备的 id 号,d_i 是式(1)得到的该摄录设备与指定目标的距离。

根据目标的宽 W_T、高 H_T 和摄录设备当前的水平视场角 A_R、俯仰视场角 E_R,得到在图像中目标的宽、高与图像宽、高的比值 w、h,见式(3)。

$$w = \frac{2}{A_R} \times \arctan\frac{W_T}{2d}, \quad h = \frac{2}{E_R} \times \arctan\frac{H_T}{2d} \tag{3}$$

式中,A_R、E_R 是摄录设备的水平视场角、俯仰视场角,W_T、H_T 是目标的宽、高。

设置摄录图像中目标显示大小阈值,以过滤掉摄录距离过远或过近的图像,得到摄录距离校验值 T_{dis},见式(4)。

$$T_{dis} = \begin{cases} 0, r_{min} \leqslant \min(w,h) \text{ and } \max(w,h) \leqslant r_{max} \\ 1, r_{min} > \min(w,h) \text{ or } \max(w,h) > r_{max} \end{cases} \tag{4}$$

式中,r_{min}、r_{max} 是目标显示大小阈值的最小值、最大值。

依次遍历目标、摄录设备位姿数据,采用站心坐标转

换公式逐个将目标坐标(X_T,Y_T,Z_T)转换为以摄录设备坐标(X_C,Y_C,Z_C)为中心点的测量坐标系坐标(E,N,U),计算得到目标在测量坐标系中坐标的方位角 A_T、俯仰角 E_T,见式(5、6)。

$$A_T=\begin{cases}\pi/2, N=0 \text{ and } E>0\\ \pi\times 3/2,\ N=0 \text{ and } E<0\\ 0,\ N\geqslant 0 \text{ and } E=0\\ \pi,\ N<0 \text{ and } E=0\\ \arctan\dfrac{E}{N},\ N>0 \text{ and } E>0\\ \arctan\dfrac{E}{N}+2\pi, N>0 \text{ and } E<0\\ \arctan\dfrac{E}{N}+\pi,\ N<0 \text{ and } E\neq 0\end{cases} \tag{5}$$

$$E_T=\begin{cases}0,\ U=0\\ \dfrac{\pi}{2},\ E=N=0 \text{ and } U>0\\ -\dfrac{\pi}{2},\ E=N=0 \text{ and } U<0\\ \arctan\dfrac{U}{\sqrt{E^2+N^2}},\ \sqrt{E^2+N^2}\neq 0 \text{ and } U\neq 0\end{cases} \tag{6}$$

计算摄录设备的方位角 A_C、俯仰角 E_C 分别与 A_T、E_T 的角度差,即摄录设备光轴与目标在水平和俯仰方向上的角度差 A_S、E_S,设定目标在摄录设备光轴右方、上方为正值,并使 A_S、E_S 的绝对值均小于 π,见式(7)~(9)。

$$A_S=\begin{cases}A_T-A_C,\ A_C\leqslant\dfrac{\pi}{2}\text{and } E_C\geqslant-\dfrac{\pi}{2}\\ A_C-A_T-\pi,\ A_C>\dfrac{\pi}{2}\text{ or } E_C<-\dfrac{\pi}{2}\text{ and } A_T\leqslant\pi\\ A_C-A_T+\pi,\ A_C>\dfrac{\pi}{2}\text{ or } E_C<-\dfrac{\pi}{2}\text{ and } A_T>\pi\end{cases} \tag{7}$$

$$A_S=\begin{cases}A_S+2\pi,\ A_S<-\pi\\ A_S-2\pi,\ A_S>\pi\end{cases} \tag{8}$$

$$E_S=\begin{cases}E_T-E_C,\ A_C\leqslant\dfrac{\pi}{2}\text{ and } E_C\geqslant-\dfrac{\pi}{2}\\ E_C+E_T-\pi,\ A_C>\dfrac{\pi}{2}\text{ or } E_C<-\dfrac{\pi}{2}\text{ and } A_T\leqslant\pi\\ E_C+E_T+\pi,\ A_C>\dfrac{\pi}{2}\text{ or } E_C<-\dfrac{\pi}{2}\text{ and } A_T>\pi\end{cases} \tag{9}$$

建立摄录图像左顶点为原点,右方、下方为正值,图像右下顶点坐标为(1,1)的二维坐标系,以匹配图像识别结果,将角度差换算为此坐标系的坐标点 (x,y),见式(10)。

$$x=\frac{A_S}{A_R}+0.5,\ y=0.5-\frac{E_S}{E_R} \tag{10}$$

将 x、y 值与数据时间戳、目标 id、摄录设备 id 等存入缓存中。构建目标的光轴偏离计算公式,得到距离 S_2,见式(11)。

$$S_2=1-\max(|2x-1|,|1-2y|) \tag{11}$$

既当 $S_2=1$ 时目标中心点与光轴重叠时,$S_2\geqslant 0$ 时目标中心点在摄录设备的可视角度内,当 $S_2\leqslant D_{min}$ 时目标整体全部位于摄录画幅外,D_{min} 值见式(12)。

$$D_{min}=-\max\left(\frac{2}{A_R}\times\arctan\frac{W_T}{2d},\frac{2}{E_R}\times\arctan\frac{H_T}{2d}\right) \tag{12}$$

式中,W_T 是目标的宽,H_T 是目标的高,d 是式(1)得到的距离。

考虑摄录效果和误差,设置目标与图像中心的距离阈值 D,以过滤掉画幅中无目标出现的图像,得到目标是否在画幅内校验值 T_{in},见式(13)。

$$T_{in}=\begin{cases}0,\ S_2\geqslant D\\ 1, S_2<D\end{cases} \tag{13}$$

式中,D 是预设的目标与图像中心的距离阈值,S_2 是式(11)得到的距离。

构建位置偏离计算公式,用于表示该物体与指定目标在图像中的位置偏离程度,见式(14)。

$$S_3=1-\frac{\sqrt{((X_p-x)^2+(Y_p-y)^2}}{2} \tag{14}$$

式中,X_P、Y_P 为图像识别结果中物体的中心点坐标,x、y 是式 10 得到的目标在图像中的坐标。

1.7 图像静帧校验

采用图像静帧识别方法[8]得到图像是否静帧的识别结果,以过滤掉静帧或断路的图像,得到图像静帧校验值 T_{fre},见式(15)。

$$T_{fre}=\begin{cases}0,\ \text{未检出图像静帧}\\ 1,\ \text{检出图像静帧}\end{cases} \tag{15}$$

1.8 类别信息校验

依次遍历 RGB 图像[9]中识别出的物体,查看识别结果中每个物体的类别是否在指定目标的衍生类别范围内,以过滤掉图像中未识别出目标的图像,得到类别校验值 T_{cls},见式(16)。构建类别相似度计算公式,见式(17)。

$$T_{cls}=\begin{cases}0,\ \text{在类别范围}\\ 1,\text{不在类别范围}\end{cases} \tag{16}$$

$$S_4=\left(1-\frac{i-1}{m}\right)\times P \tag{17}$$

式中,m 为衍生目标类别集合中的元素个数,$i=[1,2,\cdots,m]$ 是该物体类别在衍生目标类别中按照关联程度排序的序号,P 是识别物体的置信度。

1.9 获取图像识别物体的运动轨迹

读取缓存中目标位置换算得到的对于每台摄录设备图像的坐标数据,通过前后帧数据计算得到目标在图像中运动的速度和加速度,存入缓存 C_1,既 C_1 中的 1 帧数据为[time,目标 id,摄录设备 id,坐标 x_1,坐标 y_1,类别 class,速度 v_1,加速度 a_1]。

结合 DeepSort 目标跟踪算法[10-12]追踪图像识别结果中的物体,得到物体的 id、速度、加速度等,存入缓存 C_2,既 C_2 中的 1 帧数据为[time,设备 id,物体 id,坐标 x_2,

坐标 y_2，宽 w_2，高 h_2，速度 v_2，加速度 a_2，类别 class，置信度 p]。

1.10 轨迹信息校验

设置阈值以过滤数据间断或长度过短的轨迹数据，设置轨迹坐标间断的时间阈值 r_1、截取轨迹长度的时间阈值 r_2，即截取的轨迹的终点为当前时统时间 t，轨迹起点为 $t-r_2$，截取的轨迹数据帧数不小于 $\frac{r_2}{r_1}-1$。

获取缓存 C_1 中，具有同一目标 id、同一摄录设备 id、最新帧时间戳大于 $t-r_1$、连续两帧时间间隔不大于 r_1、时间戳位于 $t-r_2$ 至 t 之间、坐标通过式(13)计算得到的 $T_{in}=0$ 的数据，按照目标和设备 id 分组保存为列表 C_3，见式(18)。

$$C_3=f(i,j,k)=[x(i,j,k),y(i,j,k),v(i,j,k),a(i,j,k),\text{class}(i,j,k)] \tag{18}$$

式中，$i=[1,2,\cdots n]$ 是摄录目标的 id 号，n 是轨迹符合要求的目标总数，$j=[1,2,\cdots,m_i]$ 是摄录设备的 id 号，m_i 是该目标在图像中轨迹每个坐标 $T_{in}=0$ 且间断 $<r_1$ 的摄录设备总数，$k=[1,2,\cdots,u_j]$ 是时间戳的排列号，u_j 是该目标截取数据的总帧数。

获取缓存 C_2 中，具有同一设备 id 和物体 id、最新帧时间戳大于 $t-r_1$、连续两帧时间间隔不大于 r_1、时间戳位于 $t-r_2$ 至 t 之间的数据，按照设备和物体 id 分组并保存为列表 C_4，见式(19)。

$$C_4=f(i,j,k)=[x(i,j,k),y(i,j,k),v(i,j,k),a(i,j,k),p(i,j,k),\text{class}(i,j,k)] \tag{19}$$

式中，$i=[1,2,\cdots n]$ 是摄录设备的 id 号，n 是符合条件的摄录设备总数，$j=[1,2,\cdots m_i]$ 是该设备摄录图像识别结果中物体的 id 号，m_i 是该图像识别结果中轨迹符合要求的物体总数，$k=[1,2,\cdots u_j]$ 是时间戳的排列号，u_j 是该 id 物体截取数据的总帧数。

依次选取 C_3 中单个目标的轨迹数据，再选取 C_4 中对应目标校验值 $T_{in}+T_{cls}=0$ 且同一摄录设备 id 的轨迹数据进行位置修正，因定位设备安装位置和定位误差等因素影响，位置信息获得的目标轨迹不一定是目标中心点的轨迹，故应选择轨迹形态最接近的，消除轨迹中心点距离带来的影响，故采用两条轨迹的坐标均值差作为偏移量修正物体的轨迹，见式(20)。

$$\begin{cases}(x_2)_i=(x_2)_i+\bar{x}_1\bar{x}_2,\\(y_2)_i=(y_2)_i+\bar{y}_1-\bar{y}_2,\\i=[1,2,\cdots n]\end{cases} \tag{20}$$

式中，$\bar{x}_1\bar{y}_1$ 是目标 1 的坐标点均值，$\bar{x}_2\bar{y}_2$ 是物体 2 的坐标点均值，n 是物体轨迹中坐标点总数，$i=[1,2,\cdots n]$ 是坐标点序号。

受处理算力和传输延时等情况影响，目标与物体轨迹数据的帧率不统一且未对齐，故采用 DTW 动态时间规整算法[13]，先分别计算轨迹坐标 x、坐标 y、速度 v，加速度 a 分量的 DTW 距离，再求和得到轨迹的 DTW 距离，见式(21)。

$$D=D_x+D_y+D_v+D_a \tag{21}$$

式中，D_x、D_y、D_v、D_a 是轨迹中坐标 x、坐标 y、速度 v，加速度 a 分量的 DTW 距离。构建轨迹相似度计算公式，见式(22)。

$$S_5=1-\frac{D_i}{\max(D_1,D_2,\cdots,D_n)}+k \tag{22}$$

式中，默认 $S_5=0$，防止因数据间断等原因导致轨迹无法进行相似度计算时，轨迹相似度项缺失影响最终优选函数的计算。n 是物体总数，$i=[1,2,\cdots,n]$ 是物体的 id 号，D_i 是式(21)得到的该物体与指定目标轨迹的 DTW 距离，k 是权重参数，即轨迹相似度排序中最后 1 名的得分，使得 $S_5>0$，因与目标轨迹 DTW 距离较小的物体，未必比轨迹过短或间断的物体更近似目标，建议 k 设为较小值。

1.11 其他信息校验

增加摄录角度、显示尺寸、图像清晰度、图像亮度等校验信息，提升排序结果的合理性，满足个性化优选需求。计算目标在测量坐标系中坐标轨迹的运动方向向量 $\boldsymbol{v}_1$，采用罗德里格公式计算得到测量坐标系中摄录设备光轴的向量 $\boldsymbol{v}_2$，即可计算得到摄录设备拍摄目标的水平和垂直方向的摄录角度 A、E，设置水平和垂直方向的摄录角度范围得到摄录角度校验值 T_{ang}，用于过滤不符合摄录角度要求的图像；通过设置目标显示面积占图像比例范围得到显示尺寸校验值 T_{siz}，用于过滤图像中物体过大或过小的情况，因存在目标遮挡、形变、部分画幅外等情况，显示尺寸得分不应与目标实际尺寸进行比较，故采用面积大为优的方式构建显示尺寸得分公式，得到显示尺寸得分 S_6 = 识别物体宽 × 识别物体高；采用 Brenner 函数计算图像的梯度，通过设置图像梯度阈值得到清晰度校验值 T_{def}，过滤模糊的图像；计算图像的平均灰度，通过设置图像平均亮度阈值得到亮度校验值 T_{lig}，过滤过亮和过暗的图像。

1.12 RGB+P 优选函数

首先设置数据过滤公式，见式(23)，得到判定值 T_{all}，$T_{all}=0$ 为合格，$T_{all}>0$ 表示该摄像设备未捕获到目标或不符合设定的要求。

$$T_{all}=T_{dis}+T_{in}+T_{fre}+T_{cls}+T_{ang}+T_{siz}+T_{def}+T_{ligV} \tag{23}$$

构建 RGB+P 优选函数，用于多通道图像的优选排序，见式(24)。

$$S=\sum_{i=1}^{6}k_iS_i \tag{24}$$

式中，k_1、k_2、k_3、k_4、k_5、k_6 为权重系数，S_1 是式(2)摄录距离偏离计算结果，S_2 是式(11)光轴偏离计算结果，S_3 是式(14)位置偏离计算结果，S_4 是式(17)类别相似度计算结果，S_5 是式(22)轨迹相似度计算结果，S_6 是显示尺寸得分。

计算优选排序结果过程如下：首先选取 1 个目标，依

据式(23)过滤每个摄录图像中的所有识别物体,然后根据式(24)得到过滤后摄像图像中所有物体对应该目标的排序分值 S,再对每个摄录图像中的物体按 S 值进行倒序排序,最后对所有摄录图像的排序 1 物体 S 值进行倒序排序,得到的排序结果就是该目标的优选排序结果,即该目标的摄录设备拍摄效果排序。依次遍历其他目标进行上述排序处理后,得到的多目标摄录优选排序结果,就是本文的输出结果。

1.13 设置自动化导播策略

设置优选排序范围阈值 r_{num}、最大切换时间阈值 r_{max} 和最小切换时间阈值 r_{min},设置计时器 t_1、t_2,即当排序结果不为空时,设当前播出图像在排序结果中排序号为 i,当 $i=1$ 时,计时器 t_1、t_2 值清零;如 $i \neq 1$ 且 $i \leqslant r_{num}$,计时器 t_2 值清零,累计计时 t_1,如计时时间 $t_1 > r_{max}$ 时发出切换排序 1 图像指令同时计时器 t_1 值清零;如 $i > r$,累计计时 t_2,如计时时间 $t_2 > r_{min}$ 时发出切换排序 1 图像指令。避免反复切换图像造成观看不适。

2 实验分析及结果

除目标和摄录设备外,搭建了由 5 台服务器和 1 套分布式编解码平台等组成的硬件环境,软件采用 PyTorch 框架,共采集 16 路 1 080 p 图像,最大目标处理数量 16 个,最大优选输出排序数量 16 组,软件平均处理帧率为 12 帧/s,实现了多通道图像对应多目标的自动优选排序。因播出图像内容与预期目标不匹配的情况主要依靠人工辨别,在实际任务中难以量化,且不同配置的 RGB+P 优选函数都是建立在 RGB 识别结果的基础上,排序范围只会减少而不能扩大,故采用排序范围被裁剪和最优输出被修正作为测试的正向结果。经实际使用测试,在未加入距离校验值 T_{dis}、摄录角度校验值 T_{ang}、显示尺寸校验值 T_{siz}、清晰度校验值 T_{def}、亮度校验值 T_{lig} 和光轴偏离 S_2、显示尺寸 S_6 等主观设定的条件和计分项的情况下,对比了 RGB+P 优选函数($T_{all}=T_{in}+T_{fre}+T_{cls}$,$S=S_1+S_3+S_4+S_5$)与单模态 RGB 识别($T_{all}=T_{fre}+T_{cls}$,$S=S_4$)的 58 376 668 次排序结果,其中 32 165 373 次两项排序均为空,占比 55.099%,剩余排序不全为空的排序结果中排序修正次数 23 583 661 次,占比 40.399%,其中 22 698 297 次缩减了排序范围,占比 38.882%,21 765 171 次修正了排序 1 的值,占比 37.284%,在不全为空的 26 211 295 次排序结果中实现了 89.975%的排序范围修正和 83.037%的最优输出修正比率,能够有效避免无目标和目标不匹配图像的播出,实现了设计目的。

3 结论

本文通过构建 RGB+P 优选函数,增加位置信息校验并融合图像识别结果的处理环节,提高了直播任务中图像目标识别的准确率,减少了因目标局部可见、遮挡、重复、能见度差等情况导致的误报率,响应速度优于传统人工导播工作模式,识别准确率优于单模态的 RGB 图像识别技术,实现避免了输出结果出错的设计目标,提升了导播工作的自动化水平,可完成大部分通用场景的图像自动化导播任务,适应了导播系统的未来需求,为计算机视觉技术在电视台导播部位的应用提供新思路,可供同类任务借鉴参考。局限性在于,该方法对位姿数据的准确性要求较高,在数据丢帧或延时过大时,会出现排序结果清零的情况,既 RGB 图像识别结果被裁剪过度;另外数据未对齐、帧率不一致和图像识别模型的准确率等问题限制了该方法的输出结果精度;还有对图像质量的评价缺少可用的量化指标,优选方法与主观评价的契合度不高。未来工作方向,进一步提高优选结果的准确度和鲁棒性,减少对任一模态数据的强依赖,研究多通道图像融合、多目三维场景还原等多模态数据融合的优选方法;引入与语义匹配的图像识别模型,提高优选结果与导演想法的契合度。

参考文献

[1]陈戈. 智能导播助力 2021 春晚新媒体节目创新——浅析人工智能切换技术的应用[J]. 现代电视技术, 2021(3):35-40.
[2]周杨. 面向体育赛事智能导播的视频时序检测技术[D]. 北京:北京邮电大学, 2023.
[3]李春阳. 全景足球视频自动导播关键技术研究[D]. 北京:北京交通大学, 2020.
[4]Wang Ao, Chen Hui, Liu Lihao, et al. YOLOv10: real-time end-to-end object detection[C]//Advances in Neural Information Processing Systems (NeurIPS). Vancouver: Curran Associates, 2024: 107984-108011.
[5]Singh S, Yadav A, Jain J, et al. Benchmarking object detectors with COCO:a new path forward[C]//European Conference on Computer Vision(ECCV 2024). Cham: Springer Nature Switzerland, 2024:279-295.
[6]Zhang Weixia, Zhai Guangtao, Wei Ying, et al. Blind image quality assessment via vision-language correspondence: a multitask learning perspective[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023:14071-14081.
[7]Liu Zhuang, He Kaiming. A decade's battle on dataset bias: are we there yet? [PP/OL]. arXiv(2024-03-13)[2024-05-18]. https://arxiv.org/abs/2403.08632
[8]郭亮. 图像静帧现象自动识别方法[J]. 自动化技术与应用, 2023, 42(11):19-21.
[9]梁靓. 色块划分算法在数字图像油画化转换中的应用[J]. 自动化技术与应用, 2024, 43(2):53-55.
[10]高海静. 基于细粒度特征提取的数字媒体视频运动目标自适应跟踪[J]. 自动化技术与应用, 2025, 44(8):79-83.
[11]彭炜峰, 罗江华, 张洲亚. 基于注意力胶囊网络的梨树叶部病害图像识别方法[J]. 农机化研究, 2025, 47(10):212-218.
[12]Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing:IEEE, 2017: 3645-3649.
[13]Rakthanmanon T, Campana B, Mueen A, et al. Searching and mining trillions of time series subsequences under dynamic time warping[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 262-270.