用于通用多摄像头人员跟踪的分层聚类和细化技术

Zongyi Li [1†] Runsheng Wang [1†] He Li[1] 魏博浩[1] 史宇轩[1*]
合肥凌[1] Jiazhong Chen[1] Boyuan Liu[1] 李中阳[1] Hanqing Zheng[1]
[1]华中科技大学计算机科学与技术系。
{zongyili,wrsh,he_li,xavid,shiyx,lhefei,jzchen,leobryan,lzy123,zgwxzhq3}@hust.edu.cn

摘要

*近来，多摄像头人员跟踪因其在监控场景中的广泛应用而备受关注。然而，由于视点不同、遮挡严重和光照变化等原因，这项任务充满挑战。为了应对这些挑战，我们提出了一种用于广义多摄像头人员跟踪的新型分层聚类和细化框架。具体来说，我们的框架由两个主要部分组成：分层聚类和分层细化。与直接在多摄像头之间对轨迹点进行聚类相比，我们的分层聚类策略可以逐步将轨迹点分配给正确的目标。然而，聚类和跟踪过程不可避免地会产生不正确的匹配。因此，我们提出了一种分层细化策略来减少这些不正确的匹配，其中包括：摄像机内轨迹点级别细化、姿态细化、时空细化和脸部细化。大量实验表明，我们的方法非常有效，在 2023 年人工智能城市挑战赛赛道 1 中取得了*92% 的 *IDF1，在排行榜上名列第五。*

1. 导言

近来，许多研究人员不断探索各种视频监控场景下多个摄像头之间的人员检索方法，以创造更多实用的人员搜索应用。与基于虚构图像的人员再识别相比，多目标、多摄像头跟踪（MTMCT）是一项更具实用性和挑战性的计算机视觉任务。除了跨摄像头的人员识别外，它还需要在每个摄像头内跟踪多个目标，以及跨摄像头的轨迹关联。有了来自多个摄像头的多个目标的轨迹，MTMCT 就能提供多视角和更多信息的人脸识别。

† 同等贡献。* 通讯作者。

(a) 低分辨率的小轨迹。 (b) 高分辨率的小轨迹。 (c) 不同视角的小轨迹。

图 1.对单个摄像机内的小轨迹进行聚类。不同分辨率和不同视点的同一特征的小轨迹可以通过聚类结果进行汇总。

对不同目标进行有针对性的分析，促进公共安全系统的发展和实用性。一般来说，MTMCT 的流程可归结为以下几个步骤：1) 行人检测；2) 单摄像头跟踪；3) 人员 ReID 特征提取；4) 跨摄像头关联。具体来说，第一步是定位监控视频每一帧中的人员位置。第二步，使用最先进的多目标跟踪（MOT）方法对第一步检测到的人员进行短期跟踪，以获得短期小轨迹。第三步，利用深度人员再识别模型提取每个轨迹子的人员再识别（ReID）特征。最后一步是利用第三步获得的小轨迹的 ReID 特征，并根据提取的小轨迹特征的相似性将不同摄像头的小轨迹关联起来。

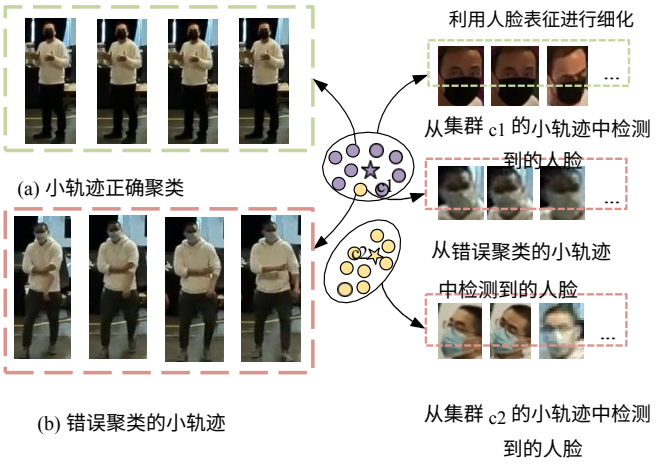
然而，MTMCT 任务存在几个问题：首先，MTMCT 应适用于不同的监控场景。值得注意的是，随着元诗歌的发展，在虚拟监控场景中追踪人员是

值得探讨。在 AICity 2023 的第一赛道中，考虑了不同室内监控场景下的 MTMCT 任务，包括真实监控场景和虚拟监控场景。

其次，单个摄像头内的 MOT 只能捕捉到短期的小轨迹，将目标与单个小轨迹联系起来不够稳健，因为这容易受到视角、光照和投入图像分辨率差异的影响。在真实的公共场所，如办公室和超市，同一身份经常在同一摄像头内反复出现和消失。这表明同一身份在同一摄像头下有多个跟踪点。如图 1 所示，图中显示的轨迹点是不同的位置和视角。直观地说，如果我们首先将

通过镜头内聚类分析同一目标的小轨迹，聚合后的特征对于跨镜头关联具有更强的鲁棒性。因此，我们提出了**分层聚类**作为聚类策略。由于镜头内的小轨迹聚类比镜头间的小轨迹聚类更容易，我们的方法首先利用提取的人物 ReID 特征对镜头内的小轨迹进行镜头内聚类，以获得所有镜头内目标的更健壮的表示。随后，我们将镜头内聚类特征作为相应目标的鲁棒性人物 ReID 表示，并进一步利用聚类特征关联跨镜头目标。

最后，我们注意到，由于单个图像会受到遮挡和光照差异的影响，因此无法保证小轨迹中帧的质量。此外，还有许多错误聚类的小轨迹。因此，我们提出了**分级细化方法**，包括小轨迹级细化和集群级细化。在小轨迹级细化中，我们会过滤掉与视频级小轨迹特征不相似的低质量帧。在集群级细化中，我们引入人脸表征作为另一种有效的生物识别线索，用于细化聚类结果。如图 2 所示，由于错误聚类的小轨迹与 c_1 的 ReID 特征相似，因此被聚类到 c_1。至于人脸表征，它与 c_1 差异很大，而与 c_2 相似。具体来说，我们首先通过选择与相应聚类不相似的小轨迹来提取可能被错误聚类的小轨迹。然后，我们将选出的小轨迹与人脸表征联系起来。



我们的贡献如下

- 我们提出了分层聚类方法，即分别进行摄像机内聚类和摄像机间聚类，以关联各摄像机的身份。
- 我们提出了分级细化方法，包括小轨迹级细化和集群级细化。这两种细化方法分别通过过滤掉低质量的帧来细化小轨迹的特征，并通过提高小轨迹的质量来提高集群的质量。

图 2.利用人脸表征进行簇级细化。(a)
(b) 显示的是一个被正确聚类到集群 c_1 的小轨迹（该小轨迹的真实集群为 c_2），因为它与 c_1 具有相似的外观特征。然而，错误聚类的小轨迹的人脸与从 c_1 检测到的人脸差别极大。

通过利用人脸表征作为另一种有效的生物识别线索，实现聚类结果。

- 我们的方法性能先进，在 AICity 2023 第一赛道中获得第五名。

2. 相关工作

2.1. 人员检测

人员检测是计算机视觉中的一项重要任务，涉及图像或视频中的人员识别和定位。它应用广泛，包括视频监控、自动驾驶、人机交互等。

一般来说，人员检测算法可分为两大类，包括单级检测器 [33, 41] 和两级检测器 [2, 19, 42]。单级检测器，如 YOLO [41] 和 SSD [33]，以实时性能和速度著称，但可能会为了速度而牺牲一些准确性。相比之下，两阶段检测器，如 Faster R-CNN [42]、Mask R-CNN [19] 和 Cascade R- CNN [42]，具有更高的精度和灵活性，但需要更多的计算资源。

由于变换器结构在自然语言处理中的成功应用，基于变换器的检测器（如 DETR [4] 和 Swin Transformer [35]）最近正在蓬勃发展。这些检测器利用视觉变换器将图像视为一系列斑块，利用自我注意机制捕捉远距离依赖关系。因此，它们在物体检测基准测试中表现出了极强的竞争力。

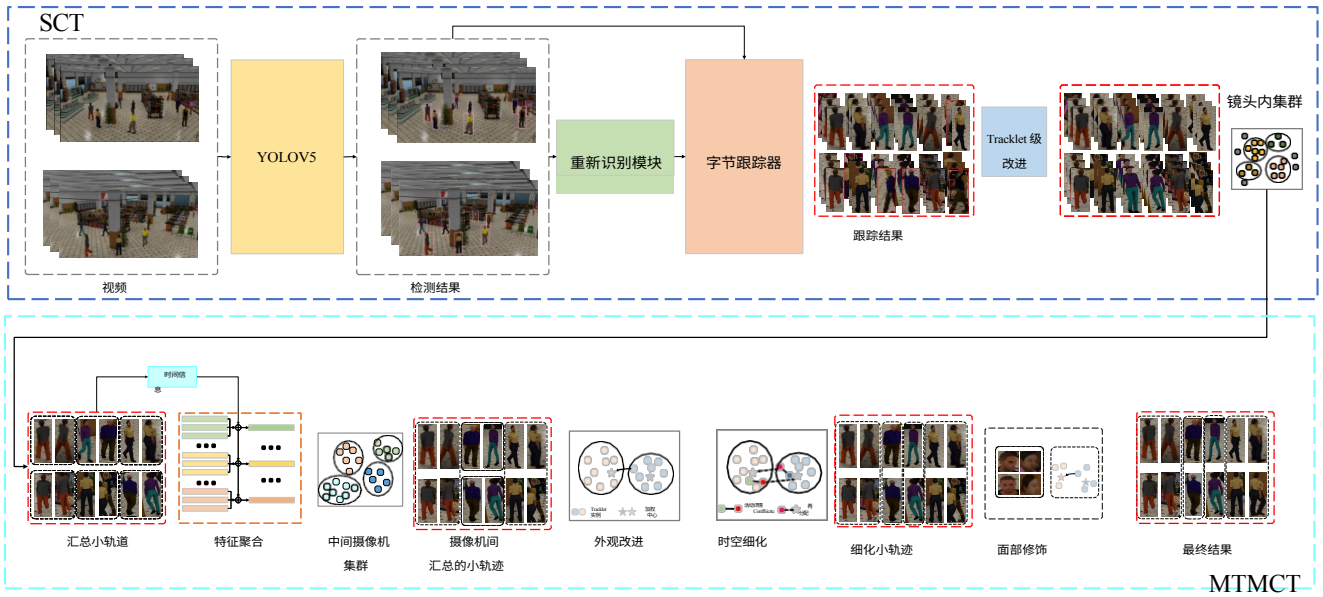


图 3.我们对 MTMCT 系统进行分层细化的流程。我们首先使用 YOLOv5 模型检测所有人员，并使用 ReID 模型提取人员的 ReID 特征。然后将所有边界框和 ReID 特征发送到 ByteTracker，以关联视频中的行人边界框。对小轨迹内和小轨迹间进行分层聚类，以合并单个小轨迹。然后进行小轨迹级细化和摄像机间小轨迹细化，以改善聚类结果。

2.2. 人员重新识别

人员再识别（ReID）[21, 30, 34, 36, 55] 是计算机视觉领域的一项重要任务，尤其是在监控系统中，其目的是匹配不同摄像头下的行人。过去几年中，通过自监督学习和基于变换器的重新识别等先进技术，人员重新识别技术取得了显著进展。

自监督学习技术（如对比学习）已成为在没有大量标记数据的情况下进行 ReID 表征学习的成功方法[24]。这些技术简化了大规模数据集上 ReID 模型的训练。现有的 自监督学习方法可分为三类。首先，生成式自监督学习方法（generative self-supervised learn- ing）[8, 11, 25, 40]的目的是生成合成样本，然后进一步扩大训练数据，提高 ReID 模型的泛化性能。其次，对比性自我监督学习 [5,7,9,17,18,48] 的目的是通过将同一样本的嵌入式数据与不同的数据增量拉近，同时将其他样本的嵌入式数据推远，从而训练编码器。最后，广告式自我监督学习 [12, 16, 27, 28] 的目的是通过训练生成器生成虚假样本，并通过判别器将它们与真实样本区分开来。目前，对比式自我监督学习已在计算机视觉领域占据主导地位。

基于 CNN 的技术在 ReID 领域占据主导地位已有数年之久。不过，纯变换器模型的受欢迎程度正在上升。例如，TransReID 模型[21]是第一个有效利用 Vi- sion 变换器进行人员和车辆 ReID 的模型，取得了最先进的成果。其他许多研究也尝试利用变换器来聚合 CNN 主干网的特征或信息。例如，[29, 44, 54] 将变换器层集成到 CNN 主干网中，以聚合层次特征并对齐局部特征。此外，对于视频 ReID，[34, 55]利用变换器来聚合外观特征、空间特征和时间特征，以学习人物轨迹子的判别表示。

结合上述两种方法，TransReID-SSL [36] 进一步研究发现，在现有的自监督学习（SSL）方法和网络架构中，采用 Trans- former 架构的 DINO [6] 算法获得了最好的 ReID 性能。

2.3. 单机跟踪

单摄像头跟踪（SCT）是计算机视觉的一个子领域，旨在跟踪从单摄像头捕获的视频序列中物体的运动[10]。目前有两种 SCT 算法。第一类遵循 "逐个检测跟踪 "范式 [1, 3, 13, 38, 51, 53, 56]，第二类称为 "联合检测跟踪 "范式 [1,3,13,38,51,53,56]。

将物体检测与 ReID 结合在一个网络中 [31, 45, 49, 57, 59]。

通过检测进行跟踪的方法，如 SORT [1] 和 DeepSORT [51]，首先使用深度检测模型检测物体，然后通过相邻帧之间的数据关联获得目标的轨迹。由于物体检测技术不断改进[15,19,35,41,42]，这些方法多年来一直在 SCT 任务中占据主导地位。联合检测跟踪方法，如在检测框架中加入外观嵌入或运动预测的方法，能以较低的计算成本获得相当的性能。然而，这些方法在优化不同组成部分之间的竞争方面面临挑战，最终制约了它们的跟踪性能。

2.4. 多摄像头人员跟踪

根据上述任务的结果，多摄像机人员跟踪的首要目标是在不同摄像机之间建立一系列跟踪链。为了增强其流水线功能，一些研究成果 [22, 23, 26, 39] 加入了关于摄像机设置的外部信息。为了防止不可行的跨摄像头转换，[23, 32, 46]利用了场景拓扑，而[22, 32, 46]则考虑了摄像头的邻接性。在文献[22]中，摄像机的移动方向被用来确定摄像机转换的可行性，而摄像机特定区域则被用来确定轨迹出现在多个摄像机中的可能性。此外，在相关文献[23, 26, 43, 47]中，聚类方法也被认为是解决这一任务的有效方法。

3. 方法

3.1. 概述

我们提出的多目标多摄像头跟踪（MTMCT）框架如图 3 所示。该框架由四个主要部分组成：人员检测、人员重新识别、单摄像头跟踪和多摄像头跟踪。整个过程可归纳如下：(1).使用人物探测器从每个摄像机视图中获取人物边界框。(2).使用预先训练好的 ReID 模型，从每个边界框中提取人物 ReID 特征。(3) 利用单摄像头跟踪模型，为每个摄像头生成单摄像头跟踪子。(4) 对小轨迹特征进行聚类，以关联摄像机内的小轨迹。(5) 采用聚类方法关联摄像机间的轨迹点。(6) 利用外观、空间-时间和脸部约束对相机间小轨迹进行再筛选。

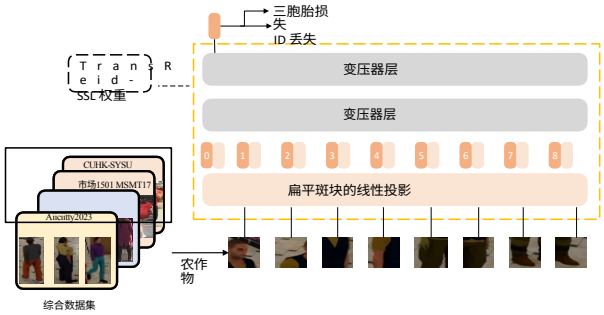


图 4.ReID 模型的流程。我们通过 TransReID-SSL 预先训练的权重初始化 Tran- sReID 模型，并在一个组合数据集上对 ReID 模型进行微调。

至关重要。人员检测阶段需要精确且无遗漏的行人检测盒。为了有效地检测更多行人，我们采用了 YOLOv5 作为检测器。总之，YOLOv5 是一种高精度、高效率的物体检测算法，已被广泛应用于各种领域。

在比赛场景中，既有真实的行人图像，也有虚拟的行人图像。由于预训练的 YOLOv5 模型已经在 COCO 数据集上实现了出色的人员检测性能，因此

3.2.人员检测

人员检测是跨摄像头跟踪的第一步，也是关键一步；因此，使用可靠的检测器是

我们使用它来检测真实世界场景中的人员。对于虚拟场景，我们使用 AICity2023 提供的虚拟数据集，从头开始训练 YOLOv5 检测器。我们忽略其他类别，只检测场景中的行人，并应用 NMS 去除重复检测框。通过 YOLOv5 检测器，我们获得了相应视频中行人的检测框和置信度分数。

3.3. 人员重新识别

我们的人员再识别模型基于 TransReID-SSL [36] ，该模型已在 LUperson 数据集 [14] 上进行了预训练，并以其能够提取稳健且与领域无关的 ReID 特征而著称。如图 4 所示，为了进一步提高其性能，我们在一个由 Market-1501 [58] 、MSMT17 [50] 、CUHK-SYSU [52] 和 CUHK-SYSU [53] 组成的组合数据集上训练 TransReID [21] 模型。

AiCity2023 数据集。我们使用 TransReID-SSL 预训练模型初始化模型权重，并使用交叉熵损失和三重损失对 256 × 128 尺寸的输入图像进行微调。交叉熵损失函数可表述如下：

$$L_{\text{天花板}}^{\text{天}} = -\frac{1}{N}\sum_{i=1}^{N} y^{i} \log\left(y^{\wedge i}\right) 、 \quad (1)$$

其中，y 是第 i 幅图像的 ID 标签，N 是合并数据集中的图像数量。而三重损失可以是

可表述如下

$$L_{tri} = \sum_{i=1}^{N} \max(m + d(f_i^a, f_i^n) - d(f_i^a, f_i^p), 0)、 \quad (2)$$

其中，d 是 l_2 的距离，f_i^p，f_i^n 是正负距离。
m 是三重丢失的余量。

3.4. 单摄像头人员跟踪

对于单摄像头人物追踪，我们需要将视频中检测到的边界框关联起来，以获得相关的小轨迹。我们使用 ByteTrack 作为跟踪算法，它可以通过将每个检测框与唯一标识关联起来来关联低置信度检测。

如图 3 所示，我们首先使用 ReID 模型提取人物特征，然后将边界框和 ReID 特征输入跟踪模型。跟踪器会考虑移动信息和视觉相似性，为每个检测到的方框分配一个跟踪子 ID。有了跟踪模型，我们就可以将视频中的行人边界框进行关联，从而获得小轨迹。

单摄像头跟踪器级细化 由于某些拥挤的情况，不同身份的人可能会被分配到同一个跟踪器上，从而导致 ID 切换
问题。因此，我们首先计算小轨迹的内部方差，方法是计算小轨迹之间的距离。
单个帧的特征和所有帧的平均特征
在一个小轨迹中。如果一个小轨迹包含不同的身份，由于外观差异，其内方差会很大。因此，如果内方差大于某个阈值，我们就会使用 K-means 算法将小轨迹分割成两个小轨迹，以减少单摄像头跟踪器造成的误差。具体来说，我们将阈值设为 0.3。

单摄像头轨迹点关联 由于 AiCity2023 轨迹点 1 是在室内环境下拍摄的，同一个人在单个摄像头下可能会有多个轨迹点。因此，我们首先对这些小轨迹进行粗略聚类，并将这些小轨迹合并到同一聚类中，将其视为同一个人。为此，我们首先通过平均小轨迹中所有帧的特征来获得小轨迹特征。因此，小轨迹可以表示为 $trac = \{f、 ti、to、c\}$，其中 f 是轨迹特征，ti 和

因此，融合距离矩阵可表述为

$$D = D_{appearance} + \alpha D_{jaccard} + \beta D_{spatial}、 \quad (4)$$

其中，α 和 β 是权重参数。得到融合距离矩阵后，我们执行 DBSCAN 算法
对这些轨迹点进行粗略聚类，并将轨迹点合并为一个群集。

3.5. 跨镜头人物关联

在本模块中，我们将介绍我们的多摄像头追踪框架，该框架将单摄像头追踪汇总的轨迹点作为输入。单摄像头人物追踪采用聚类方法，根据摄像头内跟踪小点之间的融合距离对它们进行分组。同样，我们使用 K-means 聚类算法，根据聚合特征对这些聚合的小轨迹进行分组。由于直接平均小轨迹特征并不考虑每个小轨迹的长度。理想情况下，应赋予跨度较长的小轨迹更高的权重。因此，合并后的镜头内小轨迹特征的权重由 不同的小轨迹时间跨度决定，可表述为：

$$f_k^{sct} = \frac{1}{|I_k|} \sum_{f_i \in I_k} w_i f_i、 \quad (5)$$

其中 $w_i = \frac{log(t_i)}{\sum_{j=0} log(t_j)}$ 是小轨迹的时间权重
to 是人物进入和离开摄像机的时间，c 是摄像机 ID。
那么外观识别

i, l_i 是小轨迹 i 中的图像数量。这种加权方法可以更多关注较长的小轨迹。加权和特征 f_{sct} 用于下面的相机间聚类过程。在对相机内聚合的小轨迹进行聚类算法后，属于同一聚类的轨迹会被分配相同的 ID。此外，聚类不可避免地会导致一些不正确的 ID 分配。为了完善跨摄像头人物关联过程中的聚

小轨迹之间的间距可以用 $_{Dappearance}$ 表示。

此外，我们还计算了 Jaccard 距离矩阵，以结合相邻信息，该矩阵可表示为 $_{Djaccard}$。此外，考虑到同一摄像机中时间段重叠的小轨迹不可能属于同一个群组，我们将这些小轨迹的距离设为 1。

$$D^{i,j}_{\text{空间}} = \begin{cases} 1, \{ti^i, to^i\} \cap \{ti^j, to^j\} != 0 \\ 0 \quad \text{否则} \end{cases} \tag{3}$$

类结果，我们采用了三种细化方法：外貌细化、时空信息细化和人脸细化。完善跨镜头追踪结果。

由于聚类算法强调整体相似性，但忽略了每个小轨迹的长度，而较长的小轨迹应被赋予较大的权重，因此我们首先计算每个聚类的中心特征，然后再计算每个小轨迹的权重。

$_{mtmct}^{k}$ 使用公式 5 中的时间加权和。接下来，我们重新计算每个小轨迹与集群中心点之间的距离，并将每个小轨迹的身份重新分配给与其最近的集群。通过这种改进，我们可以有效地改进小轨迹，将其外观和持续时间考虑在内。

时空细化 与相机内人物追踪方法类似，我们需要排除异常的人物。

在聚类结果中加入小轨迹，以获得最终准确率的 MTMCT 结果，因为摄像机之间的聚类也可能会影响 MTMCT 结果的准确性。

将同一摄像机中时间间隔重叠的小轨迹进行聚类。具体来说，在得到摄像机间聚类结果后，我们会遍历所有聚类，并将摄像机时间重叠的小轨迹重新分配到不同的聚类中。对于同一聚类中的冲突对，只保留离聚类中心最近的小轨迹，其他冲突小轨迹则根据相似度分配到不同的聚类中，如图 3 所示。

人脸细化 由于视频中有许多低分辨率图像没有人脸，因此在计算聚类相似度时，我们不会直接将人脸表征纳入其中。取而代之的是，我们结合人脸代表来完善摄像头间的聚类结果。人脸细化包括两个步骤：首先，我们使用 MTCNN 模型从所有图像中提取人脸。然后使用 Arc- Face 模型提取相应的人脸特征。其次，我们计算一个集群中 tracklets 的平均人脸特征，作为该集群的人脸表示。由于人脸信息的信息量更大，我们根据人脸表征直接将错误聚类的小轨迹点分配到新的聚类中。具体来说，当某个小轨迹的人脸特征与根据 ReID 特征分配的群组的人脸特征有较大差异时，我们将该小轨迹视为错误聚类的小轨迹。随后，当小轨迹与其最近的人脸集群之间的人脸相似度大于 0.8 时，我们将该小轨迹的 ID 重新分配到最近的人脸集群中。

通过摄像头间聚类和分层细化策略，所有小轨迹都可以分配到一个 ID 标签。

4. 实验

4.1. 数据集

在本赛道中，MTMCT 数据集包含真实数据和虚拟合成数据。该数据集包含 1,491 分钟的视频和 130 个摄像头。视频数据均为 30 FPS 的高分辨率（1920x1080），分为 22 个子集，其中 10 个子集用于训练，5 个子集用于验证，7 个子集用于测试。此外，我们还利用 Market-1501、MTMC17 和 CUHK-SYSU 这三个公共人物 ReID 数据集进行 ReID 模型训练。

4.2. 评估指标

我们使用 IDF1、IDP 和 IDR 作为 MTMCT 评估指标。IDF1 分数是用于评估物体检测模型性能的指标。它衡量的是正确识别检测的比率，同时考虑了地面实况以及假阴性、真阴性和真阳性计数。IDF1 分数是根据 IDFN、IDTN 和 IDTP 的计数具体得出的，可表述如下：

$$IDF1 = \frac{2IDTP}{2\,\mathrm{IDTP} + IDFP + IDFN}. \tag{6}$$

4.3. 实施细节

我们的框架是在配备 24G 内存的 RTX-3090 GPU 上实现的。在人员检测模块中，我们利用在 COCO 数据集上预先训练好的 YOLOv5-l 模型来执行真实场景中的人员检测。在虚拟场景中，我们从头开始在 AIcity track1 数据集上训练 YOLOv5 模型。检测的 IOU 阈值设为 0.3，NMS 阈值设为 0.45。对于人的 ReID 模块，我们在由 Market、MSMT17 和 CUHK-SYSU 组成的组合数据集上训练 TransReID 模型，并将预先训练好的 TransReID 模型作为初始化权重。此外，我们还采用 ByteTrack 进行单摄像头人员跟踪。

4.4. 定量分析

在本小节中，我们将报告对所提出的分层聚类策略、分层细化以及用于提取所有小轨迹的人物 ReID 特征的不同骨干的消减分析。具体来说，我们在 AICity 2023 第 1 轨数据集的测试集上对所提出的聚类和细化策略进行了深入研究，而 ReID 骨架的消减实验则是在由我们自己划分的验证数据集上进行的。

建议的聚类和细化策略的消减结果如表 1 所示。1.令人印象深刻的是，分层聚类（表 1 中的 "相机内聚类 "和 "相机间 "聚类）大大提高了性能。此外，不同的细化策略也能提高性能。具体来说，外观、时空和轨迹级细化分别提高了 2%、1% 和 2%。

表 2 显示了不同 ReID 主干网的消减结果。 显然，TransReID 比基于 Resnet50 的模型表现更好。此外，使用自监督方案对经过训练的 TransReID 进行微调可获得最佳结果。因此，我们选择 TransReID-SSL 作为主干网。

如图 3 所示，我们的方法在 AIcity2023 挑战赛 Track1 中取得了 0.921 的 IDF1，与其他团队相比排名第五。

4.5. 可视化

最终的 MTMCT 结果如图 5 所示。图中每一列代表不同摄像机中的轨迹，而每一行则代表同一摄像机中不同时间的轨迹。例如，ID 17 的人物出现在不同的摄像头中，视角和视线各不相同，但我们的方法也能有效匹配他们在不同摄像头中的轨迹。

此外，我们还展示了集群间分析的结果，如图 6 所示。属于同一集群的小轨迹被分配了相同的 ID。此外，我们的方法还能处理以下情况

镜头间集群	镜头内集群	外观完善	空间-时间完善	轨道级完善	性能 (ID-F1)
✓					0.77
✓	✓				0.82
✓	✓	✓			0.84
✓	✓	✓	✓		0.85
✓	✓	✓	✓	✓	0.87

表 1.对拟议的聚类和细化策略进行的消融研究。

摄像机 1　　摄像机 2　　摄像机 3

图 5.AIcity2023 测试集的最终追踪结果可视化。同一 ID 的人在不同摄像头下用相同颜色标记。

骨干	等级-1	等级-5	等级-10	m A P
Resnet50 [20]	0.74	0.76	0.76	0.68
Resnet50-IBN [37]	0.77	0.82	0.85	0.73
TransReID [21]	0.87	0.88	0.89	0.83
TransReID-SSL [36]	**0.91**	**0.96**	**0.97**	**0.8 8**

表 2.不同骨架的消融研究

等级	团队编号	IDF1
1	6	0.9536
2	9	0.9417
3	41	0.9331
4	51	0.9284
5	**113（我们的）**	**0.9207**
6	133	0.9109
7	34	0.9104
8	82	0.8981
9	151	0.8676
10	38	0.8676

从图 2 中可以看出，不同人物图像之间存在身体重叠。在 TransReID 模型的抗闭塞能力的帮助下，我们能够将部分轨迹点与完整的身体轨迹点进行匹配，如图 6 中的第 3 组所示。

表 3.与赛道上其他队伍的比较1，我们的队伍获得第五名

。

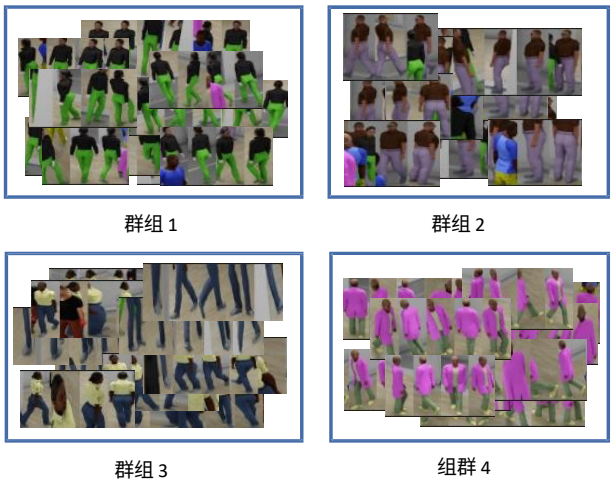


图 6.AIc- ity2023 测试集的最终聚类结果可视化。不同的聚类代表不同的人。

5. 结论

本文提出了一种有效而新颖的 MTMCT 框架，包括人员检测、使用 MOT 算法的单摄像机多目标跟踪、ReID 特征提取、分层聚类和分层细化。我们证明了分层聚类（摄像机内和摄像机间聚类算法依次进行）对室内视频监控场景极为有利，因为在这种场景中，人员会在一个摄像机中反复出现和消失。此外，我们还提出了一系列细化策略，主要包括跟踪器级细化和集群级细化。值得注意的是，我们将人脸表征作为另一种线索来修正错误聚类的轨迹子，进一步提高了跨摄像机关联性能。我们相信，MTMCT 与其他生物识别线索（如人脸和步态表示）的结合值得计算机视觉界在未来进行探索。

致谢

本研究部分得到中国自然科学基金（61972169）、中国博士后科学基金（2022M711251）、国家重点研发计划（2019QY(Y)0202、2022YFB2601802）和湖北省重大科技专项（2022BAA046、2022BAA042）的资助、湖北省重大科技专项（2022BAA046 、2022BAA042 ）部分资助，武汉市应用基础与前沿技术再研究计划（2020010601012182）部分资助，企业重点联合项目（U22B2017）部分资助，武汉市知识创新计划-基础研究。

参考资料

[1] Alex Bewley、Zongyuan Ge、Lionel Ott、Fabio Ramos 和 Ben Upcroft。简单在线实时跟踪。In *2016 IEEE international conference on image processing (ICIP)*, pages 3464-3468.IEEE, 2016.3, 4

[2] Zhaowei Cai 和 Nuno Vasconcelos.Cascade r-cnn：进入高质量物体检测。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154-6162, 2018.2

[3] 曹金坤、翁新硕、拉瓦尔-基洛德卡、庞江苗、克里斯-基塔尼。以观测为中心的排序：Rethink- ing sort for robust multi-object tracking. *ArXiv preprint arXiv:2203.14360*, 2022.3

[4] Nicolas Carion、Francisco Massa、Gabriel Synnaeve、Nicolas Usunier、Alexander Kirillov 和 Sergey Zagoruyko。利用变换器进行端到端物体检测。In *Computer Vision- ECCV 2020: 16th European Conference, Glasgow, UK, Au- gust 23-28, 2020, Proceedings, Part I 16,* pages 213-229. Springer, 2020.2

[5] 马蒂尔德-卡隆、伊善-米斯拉、朱利安-梅拉尔、普里亚-戈亚尔、皮-奥特-博扬诺夫斯基和阿曼德-朱林。通过对比聚类分配无监督学习视觉特征。*Advances in neural information processing systems*, 33:9912- 9924, 2020.3

[6] 玛蒂尔德-卡隆（Mathilde Caron）、雨果-图夫隆（Hugo Touvron）、伊山-米斯拉（Ishan Misra）、埃尔夫-杰古（Herve´ Je´gou）、朱利安-梅拉尔（Julien Mairal）、皮奥特-博扬诺夫斯基（Piotr Bojanowski）和阿曼德-朱林（Armand Joulin）。自监督视觉转换器的新兴特性。In *Pro- ceedings of the IEEE/CVF international conference on com- puter vision*, pages 9650-9660, 2021.3

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Ge- offrey Hinton.视觉表征对比学习的简单框架。In *International conference on ma- chine learning*, pages 1597-1607.PMLR, 2020.3

[8] 陈新雷、范浩琦、罗斯-吉尔希克、何开明。用动量对比学习改进基线。*arXiv preprint arXiv:2003.04297*, 2020.3

[9] 陈鑫磊、何开明。探索简单的连体代表学习。*IEEE/CVF 计算机视觉与模式识别会议论文集*》，15750-15758 页，2021 年。3

[10] Gioele Ciaparrone、Francisco Luque Sa´nchez、Siham Tabik、Luigi Troiano、Roberto Tagliaferri 和 Francisco Herrera。视频多目标跟踪中的深度学习：调查。*Neu- rocomputing*, 381:61-88, 2020.3

[11] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Ben- gio. 使用实nvp的密度估计。*arXiv 预印本 arXiv:1605.08803*, 2016.3

[12] Jeff Donahue 和 Karen Simonyan.大规模对抗性表征学习。*神经信息处理系统进展*》，32，2019。3

[13] 杜云浩、宋洋、杨波、赵艳云。Strongsort：让深度排序再次伟大。*arXiv 预印本 arXiv:2202.13514*, 2022.3

[14] Dengpan Fu、Dongdong Chen、Jianmin Bao、Hao Yang、Lu Yuan、Lei Zhang、Houqiang Li 和 Dong Chen。用于人员再识别的无感知预训练。*继续*

ings of the IEEE conference on computer vision and pattern recognition, 2021.4

[15] Zheng Ge、Songtao Liu、Feng Wang、Zeming Li 和 Jian Sun.Yolox：2021 年超过 YOLO 系列。*ArXiv 预印本 arXiv:2107.08430*, 2021.4

[16] Mathieu Germain、Karol Gregor、Iain Murray 和 Hugo Larochelle。制造：用于分布检测的屏蔽自动编码器。*国际机器学习会议*，第 881-889 页。PMLR，2015。3

[17] Jean-Bastien Grill, Florian Strub, Florent Altche´, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Ghesh- laghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *神经信息 处理系统进展*》，33:21271-21284，2020。3

[18] 何开明、范昊琦、吴雨欣、谢赛宁和罗斯-吉尔希克。无监督视觉再现学习的动量对比。*电气和电子工程师协会/计算机视觉与模式识别大会论文集*》，第 9729-9738 页，2020 年。3

[19] Kaiming He, Georgia Gkioxari, Piotr Dolla´r, and Ross Gir- shick.Mask r-cnn。 In *Proceedings of the IEEE international conference on computer vision*, pages 2961-2969, 2017.2, 4

[20] 何开明、张翔宇、任少清和孙健。图像识别的深度残差学习。In *Proceed- ings of the IEEE conference on computer vision and pattern recognition,* pages 770-778, 2016.7

[21] 何书婷、罗浩、王丕超、王帆、李浩、蒋伟。Transreid：基于变压器的物体再识别。*IEEE/CVF 计算机视觉国际会议论文集*》，第 15013-15022 页，2021 年。3, 4, 7

[22] 何宇航、韩杰、于文涛、洪小鹏、魏星和龚一红。通过语义属性解析和跨摄像头小轨迹匹配实现城市规模多摄像头车辆追踪。*IEEE/CVF 计算机视觉与模式识别研讨会论文集*》， 576-577 页，2020 年。4

[23] Hung-Min Hsu, Tsung-Wei Huang, Gaoang Wang, Jiarui Cai, Zhichao Lei, and Jenq-Neng Hwang.基于深度特征再识别和基于轨迹的摄像头链接模型的车辆多摄像头跟踪。In *CVPR workshops*, pages 416-424, 2019.4

[24] Longlong Jing and Yingli Tian.深度神经网络的自监督视觉特征学习：一项调查 *.IEEE patterns analysis and machine intelligence*, 43(11):4037-4058, 2020.3

[25] Durk P Kingma 和 Prafulla Dhariwal。Glow：具有可逆 1x1 卷积的生成流。*神经 信息处理系统进展*》，2018 年第 31 期。3

[26] Philipp Kohl、Andreas Specker、Arne Schumann 和 Jur- gen Beyerer。通过加权距离聚合实现多目标多摄像头行人追踪的 mta 数据集。*IEEE/CVF 计算机视觉与模式识别研讨会论文集*》，第 1042-1043 页，2020 年。4

[27] 古斯塔夫-拉尔森、迈克尔-迈尔和格雷戈里-沙赫纳罗维奇。自动着色的学习表征In *Computer Vision-ECCV 2016：第 14 届欧洲会议，荷兰阿姆斯特丹，2016 年 10 月 11-14 日，论文集，第四部分 14*，第 577-593 页。
施普林格出版社，2016 年。3

[28] Christian Ledig、Lucas Theis、Ferenc Husza´r、Jose Caballero、Andrew Cunningham、Alejandro Acosta、Andrew Aitken、Alykhan Tejani、Johannes Totz、Zehan Wang 等：使用生成式广告arial 网络的照片逼真单图像超分辨率。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681-4690, 2017.3

[29] 李玉林、何剑锋、张天柱、刘翔、张永东和吴锋。多样化部件发现：利用部件感知转换器进行隐蔽人物再识别。In *Pro- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2898-2907, 2021.3

[30] 李宗一、石宇轩、凌海飞、陈家忠、王倩、周丰帆。利用自组队学习进行领域自适应人员再识别的可靠性探索。*美国人工智能学会会议论文集*》，第36卷，第1527-1535页，2022年。
3

[31] Chao Liang, Zhipeng Zhang, Xue Zhou, Bing Li, Shuyuan Zhu, and Weiming Hu.反思多目标跟踪中检测与雷德之间的竞争*IEEE Transac- tions on Image Processing*, 31:3182-3196, 2022.4

[32] 刘冲、张玉琪、罗浩、唐家胜、陈伟华、徐先哲、王帆、李浩、沈屹东。由交叉道路区域引导的城市规模多摄像头车辆追踪。*IEEE/CVF 计算机视觉与模式识别会议论文集*》，第 4129- 4137 页，2021 年。4

[33] Wei Liu、Dragomir Anguelov、Dumitru Erhan、Christian Szegedy、Scott Reed、Cheng-Yang Fu 和 Alexander C Berg。SSD：单枪多箱探测器。In *Computer Vision-ECCV 2016：第 14 届欧洲会议，荷兰阿姆斯特丹，2016 年 10 月 11-14 日，论文集、第一部分 14*，第 21-37 页。Springer, 2016.2

[34] Xuehu Liu, Pingping Zhang, Chenyang Yu, Huchuan Lu, Xuesheng Qian, and Xiaoyun Yang.一个视频胜过三次观看：基于视频的人物再 识别的三叉变换器。*arXiv preprint arXiv:2104.01745*, 2021.3

[35] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。Swin transformer：使用移位窗口的分层视觉变换器。*计算机视觉国际会议论文集*》，第 10012-10022 页，2021 年。2, 4

[36] Hao Luo, Pichao Wang, Yi Xu, Feng Ding, Yanxin Zhou, Fan Wang, Hao Li, and Rong Jin.基于变换器的人物再识别的自监督预训练。*ArXiv 预印本 arXiv:2111.12084*, 2021.3, 4, 7

[37] 潘新刚、罗平、施建平、唐晓鸥。一举两得：通过ibn-net增强学习和泛化能力。*欧洲计算机视觉会议（ECCV）论文集*》，第 464-479 页，2018 年。7

[38] 庞江淼、邱林禄、李霞、陈皓峰、李琦、特雷弗-达雷尔和余飞雪。多物体跟踪的准密集相似性学习。*IEEE/CVF 计算机视觉与模式 识别会议论文集》*，第 164-173 页，2021 年。3

[39] 钱翊钧、于立军、刘文和、Alexander G Haupt- mann.电力：用于智能城市的高效多摄像头车辆追踪系统。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 588-589, 2020.4

[40] Ali Razavi、Aaron Van den Oord 和 Oriol Vinyals。用 vq-vae-2 生成多样化高保真图像。*神经信息处理系统进展*，32，2019。3

[41] 约瑟夫-雷德蒙和阿里-法哈迪Yolo9000：更好、更快、更强。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263-7271, 2017.2, 4

[42] Shaoqing Ren、Kaiming He、Ross Girshick 和 Jian Sun.更快的r-cnn：利用区域建议网络实现实时物体检测。*神经信息处理系统进展》（Advances in neural information process- ing systems）*，28，2015.2, 4

[43] Ergys Ristani 和 Carlo Tomasi.多目标多摄像头跟踪和再识别特征。In *Proceed- ings of the IEEE conference on computer vision and pattern recognition*, pages 6036-6046, 2018.4

[44] 沈飞、谢毅、朱剑青、朱晓斌、曾焕强。Git：Git: Graph interactive transformer for vehicle re-identification. *arXiv preprint arXiv:2107.05475*, 2021.3

[45] Bing Shuai、Andrew Berneshawi、Xinyu Li、Davide Modolo 和 Joseph Tighe。Siammot：连体多目标跟踪。*电气和电子工程师协会/计算机视觉与模式识别会议论文集》（IEEE/CVF conference on computer vi- sion and pattern recognition）*，第12372-12382页，2021年。4

[46] Andreas Specker、Daniel Stadler、Lucas Florin 和 Jur-gen Beyerer。遮挡感知多目标多摄像头跟踪系统。*IEEE/CVF 计算机视觉与模式识别会议论文集》*，4173-4182 页，2021 年。4

[47] Yonatan Tariku Tesfaye, Eyasu Zemene, Andrea Prati, Mar- cello Pelillo, and Mubarak Shah.使用约束多米 nant 集在多个非重叠摄像机中进行多目标跟踪。*arXiv 预印本 arXiv:1706.06196*, 2017.4

[48] 田永龙、孙晨、本-普尔、迪利普-克里希南、科迪莉亚-施密德和菲利普-伊索拉。什么是对比学习的好视图？*神经信息处理系统进展》（Advances in neural informa- tion processing systems）*，33:6827-6839，2020。3

[49] 王中道、郑亮、刘艺璇、李亚丽和王胜金。实现实时多目标跟踪。In *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XI 16*、第 107-122 页。Springer, 2020.4

[50] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian.弥合领域鸿沟的人物再识别技术In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 79-88, 2018.4

[51] Nicolai Wojke、Alex Bewley 和 Dietrich Paulus。利用深度关联指标进行简单的在线实时跟踪。In *2017 IEEE international conference on image processing (ICIP)*, pages 3645-3649.IEEE, 2017.3, 4

[52] Qiqi Xiao, Hao Luo, and Chi Zhang.边缘样本最小化损失： A deep learning based method for person re identification. *arXiv preprint arXiv:1710.00478*, 2017.4

[53] Fengwei Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan.Poi：利用高精度检测和外观特征进行多目标跟踪。In *Computer Vision-ECCV 2016 Workshops：阿姆斯特丹，荷兰，2016 年 10 月 8-10 日和 15-16 日，论文集，第二部分 14*, pages 36-42.Springer, 2016.3

[54] Guowen Zhang, Pingping Zhang, Jinqing Qi, and Huchuan Lu.帽子：用于人员再识别的分层聚合转换器。*第 29 届 ACM 国际多媒体大会论文集*》（*tional Conference on Multimedia*），第 516-525 页，2021 年。3

[55] Tianyu Zhang, Longhui Wei, Lingxi Xie, Zijie Zhuang, Yongfei Zhang, Bo Li, and Qi Tian.基于视频的人物再识别时空变换器。*arXiv preprint arXiv:2103.16469*, 2021.3

[56] 张逸夫、孙培泽、蒋毅、于冬冬、翁福成、袁泽寰、罗平、刘文宇和王兴刚。Bytetrack：通过关联每个检测框实现多目标跟踪。In *Computer Vision-ECCV 2022: 17th Eu- ropean Conference, Tel Aviv, Israel, October 23-27, 2022*、*论文集，第 XXII 部分*，第 1-21 页。施普林格出版社，2022 年。3

[57] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu.Fairmot：论多目标跟踪中检测和再识别的公平性。*International Journal of Computer Vision*, 129:3069-3087, 2021.4

[58] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jing-dong Wang, and Qi Tian.可扩展的人员再识别：一个基准。In *Proceedings of the IEEE international con- ference on computer vision*, pages 1116-1124, 2015.4

[59] 周欣怡、Vladlen Koltun 和 Philipp Kra¨henbu¨hl.以点为单位追踪物体In *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part IV*, pages 474-490.Springer, 2020.
4