**地铁车站监控视频内容分析与平台实现**

**技术要求**

**（Ver 1.0）**

实验环境及私有数据集

（1）环境：单块GPU，且Nvidia GPU Compute Capacity<=8.6(Geforce RTX30)

（2）私有数据集：输入图片尺寸为1920×1080，包含摄像头8-10个，行人数量500人以内，FPS不小于24帧，视频时长10-15min,拍摄场景为地铁站室内场景；其中，跨摄像机的groudtruth约为5000。

1. **基于多目标跟踪算法的人流量统计评估指标**

乙方在使用公开数据集的情况下，模型评价指标采用MOT17、MOT20数据集进行指标测试。

1.单摄像头多目标跟踪数据集介绍

MOT Challenge 包括 MOT15、MOT16、MOT17以及密集行人跟踪的MOT20 数据集。每个数据集的训练集和测试集都含有官方检测器提供的检测结果，同时训练集也包含手动标注框。

1. MOT17数据集

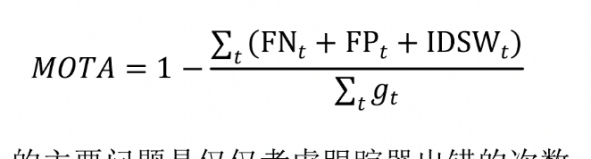
MOT17数据集内容和 MOT16 一致，但是使用了更精确的标注框，同时使用三个检测器得出三种检测结果供研究者使用，包括 11235 帧、1342个行人和 292733 个行人检测框。MOT17数据集输入尺寸在640×480-1920×1080范围内，数据集总共包含14个视频，7 个用于训练 、7个用于测试，视频时长在20s-1min25s之间，人群分布较稀疏，只在MOT17-03、MOT17-04两个子数据集中，人群较为密集，最高人群密度可达到69.7%，含了室内/室外，白天/晚上等多种场景。

1. MOT20数据集

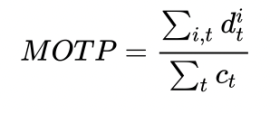
MOT20数据集从3个场景中提取了8份全新的稠密人群序列。这些序列包含了室内/室外，白天/晚上场景。输入尺寸在1173×880-1920×1080范围内，数据集总共包含 8 个视频，4 个用于训练 、4 个用于测试，MOT20 数据集以密集行人跟踪为背景，行人密度极高，最拥挤的视频平均每帧可达 245 人，数据集包含 13410 帧、6869 个行人和2259143个行人检测框。

2.单摄像机多目标跟踪算法评价指标

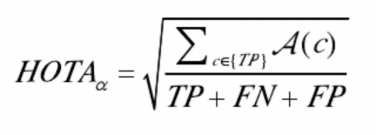
在MOT Challenge评价指标中选用MOTA、MOTP、HOTA、MT、 ML、IDsW、IDF1、FPS等八项指标评估模型，以下为八项指标的概述：

1. MOTA：多目标跟踪的准确度,计算跟踪所有帧中所有目标的误检、漏检和错误匹配，其中FN t、FP t和IDSWt分别是 t 帧时漏检、误检和错误匹配的数量，gt是地面真值目标矩形框的数量。计算公式为

（2）MOTP：多目标跟踪的精度：用来量化检测器的定位精度，𝑑dit代表第i个检测目标与给它分配的真值之间在所有帧中的平均度量距离𝑖𝑐,ct代表在当前帧匹配成功的数目;计算公式为：



（3）HOTA：HOTA是检测评分和关联评分的显性组合，它将精确检测和关联的效果平衡为单一的统一度量。代表更高维跟踪精度；HOTA的分数计算过程由多个部分组成，计算公式为：



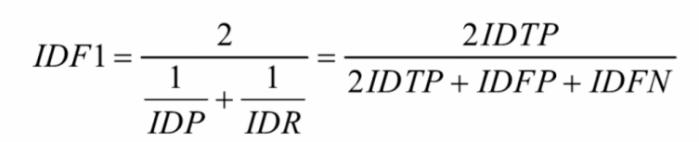
其中A表示数据关联分数。TP 则表示在整个视频中检测目标被正确跟踪的数量。FN 、FP分别是 漏检、误检的数量。

（4）MT:最多跟踪的目标数量,（Mostly Tracked）是指跟踪目标在 80%的时间以上都能够成功匹配的轨迹数量;

（5）ML:最少丢失的目标数量：（Mostly Lost）是指跟踪目标在 20%的时间以下能成功匹配的轨迹数量;

（6）IDsW: 目标身份切换的总次数：即ID改变的次数。

（7）IDF1(Identification F-Score)：IDF1即是指每个目标框中目标ID识别的F值。计算公式：



其中IDP识别精确度是指每个行人框中行人ID识别的精确度，IDR识别召回率是指每个行人框中行人ID识别的召回率。 IDTP是真正ID数， IDFP是假正ID数，IDFN是假负ID数。

1. FPS；是指画面每秒传输[帧数](https://baike.baidu.com/item/%E5%B8%A7%E6%95%B0/8019296)。

单摄像机多目标跟踪算法评价指标选用见下表；

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 评价指标 | MOTA↑ | HOTA↑ | MT↑ | ML↓ | IDsW↓ | IDF1↑ | FPS↑ |
| MOT\_17 | 81% | 64% | 61% | 13% | 2000 | 78% | 30 |
| MOT\_20 | 79% | 63% | 60% | 10% | 1200 | 77% | 17 |
| 私有数据集 | 82% | 65% | 62% | 11% | 5500 | 80% | 24 |

其中箭头向上表示该指标越高效果越好，反之，箭头向下的表示该指标越低越好.

3.多目标跟踪视频人流量统计评价指标

选用召回率、精确率和 F值三项指标；

|  |  |  |
| --- | --- | --- |
| 召回率↑ | 精确率↑ | F值↑ |
| 95% | 95% | 95% |

此外，基于视频的人流量统计结果与真实groundtruth之比要求不小于95%。基于私有数据集的人流量统计识别速率FPS不高于30帧/秒。

1. **跨摄像机多目标跟踪（MTMCT）算法评价指标**

1.数据集介绍

（1）WILDTRACK数据集

WILDTRACK是一个跨摄像机高清监控录像数据集，用于密集的无脚本行人检测，用于多摄像机行人检测，应用于安全、监视、远程人员识别、机器人、自动驾驶和众包。数据集记录了苏黎世联邦理工大学主楼外的学生。这些视频是在“无剧本”、“非演员但真实的环境中”拍摄的。在这个数据集中，数据集使用7个静态定位的高清摄像机记录。7个序列均为分辨率1920×1080像素，以每秒60帧的帧率( fps )记录。包含重叠区域的35min 视频，拍摄场景为室外场景，标注行人总数为313人，每幅图像标注2帧，该数据集只是在每个摄像头下对前3分20秒的数据进行标注，单个摄像头有400张图片，由于一共有7个摄像头，因此图像总数为2800。其中前360帧用于训练集合，其余40帧用于测试。平均而言，每帧捕获23.8人，每个人在30.41帧中被看到。

（2）PETS-2009 S2-L1数据集

PETS数据集包含S0，S1，S2，S3四个子集，S0为训练数据，S1为行人计数和密度估计，S2为行人跟踪，S3为流分析和事件识别。PETS-2009包含7个摄像头采集的有重叠区域的视频，采集对象为英国雷丁大学主校区的学生，摄像头视角较高且均为静态。图像尺寸有两种，分别为768×576、720×576像素；行人总数19人。拍摄场景为校园室外场景；共包含795帧图像，视频帧率为 7帧每秒，PETS-2009 S2-L1为其中的一个序列，该序列既包含了2D标注也包含了3D标注。

2.基于WILDTRACK数据集的跨摄像机多目标跟踪算法评价指标

选用MOTA、IDF1、MT、ML、IDsW、FPS评估模型，指标概述详见上文---单摄像机多目标跟踪算法评价指标。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据集 | MOTA↑ | IDF1↑ | MT↑ | ML↓ | IDsW↓ | FPS↑ |
| WILDTRACK | 97% | 98% | 97% | 1% | 10 | 3.5 |
| 私有数据集 | 85% | 75% | 70% | 5% | - | 10 |

3.基于PETS-2009 S2-L1数据集的跨摄像机多目标跟踪算法评价指标

选用MOTA、MOTP、FPS评估模型，指标概述详见上文---单摄像机多目标跟踪算法评价指标。

|  |  |  |  |
| --- | --- | --- | --- |
| 数据集 | MOTA↑ | MOTP↑ | FPS↑ |
| PETS-2009 S2-L1 | 95% | 90% | 10 |
| 私有数据集 | 85% | 75% | 10 |

1. **关键点检测/异常行为检测评估**

场景划分：

1. 稀疏人群场景：单摄像头范围内人数<=20人
2. 密集人群场景：单摄像头范围内人数>20人

数据集介绍：

1. 公有数据集：
2. MPII人类多人数据集，由3844个训练组和1758个测试组组成，样本数25K，这些组由具有14个身体关键点的高度关节姿势的多个相互作用的个体组成。
3. PoseTrack：多人人体关键点跟踪数据集，包含单帧关键点检测、多帧关键点检测、多人关键点跟踪三个人物，多于500个视频序列，帧数超过20K，关键点个数为15。
4. Subway Entrance and Exit:该数据集包含两个视频段, 一段为地铁入口视频, 时长96 min, 另一段为地铁出口视频, 时长43 min; 包含19种异常行为, 如走错方向、徘徊等. 视频中的每一帧图片具有帧级的异常标签,属于室内场景中的异常行为。（图像分辨率512\*384）
5. 私有数据集：