**数据挖掘导论实验报告**



**实验题目：视频帧过滤处理器**

姓 名： 闻梓棋

年 级： 大三

学 号： 3210100968

指导教师： 张东祥

实验要求

给定一个原始视频以及查询目标图片，如何设计一个二分类过滤器，快速并且准确地筛除不包含对象的视频帧。每个查询对应一个目标对象的图片，并且事先已经标注好包含该目标对象的视频帧集合G二分类过滤器将作用于所有视频帧，最后返回未被排除的视频帧集合P。

贡献总结

设计了一种基于Yolov8+Mobilenet的相似度检测方法；

设计了一种基于搜索树结构优化的抽帧预搜索思路。

设计了多个相似度评价指标（ phash，Mobilenet+phash，Siamese+CNN）

技术方案

**1、场景描述**

针对交通路口视频，给定指定车辆图片进行搜索.筛选出特定车辆出现帧区间。

对于这样的场景，需要指出的是，视频具有几个特点：

1. 车流量较少，即每个场景车辆为2-4辆车，
2. 反光严重，车辆灯光对前景提取会存在一定影响
3. 视频帧连续，本身车辆满足物理限制
4. 存在部分视频帧异常问题。
5. 目标检测时长占比小，每辆车进入到离开占比大概在10s也就是300帧左右。
6. 车辆存在转弯现象，局部特征可能会发生变化

**2、Baseline思路**

Baseline基于前景提取，针对找出的轮廓内图像进行直方图比对，设定比对阈值作为筛选依据。通过Baseline，我们可以看到，整体的思路可以分为两个部分：

1、搜索指定区域图像；

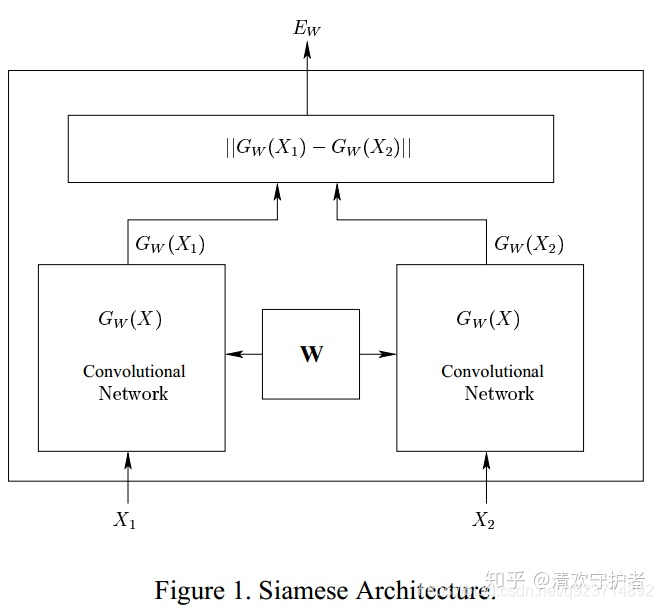
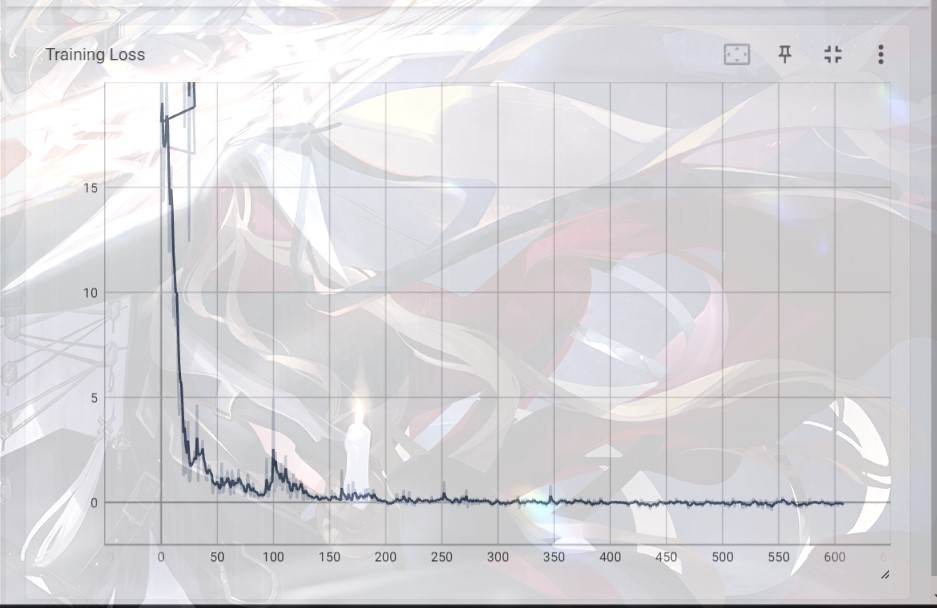
2、对可能的候选区域图像进行相似度比对。

**3、Baseline缺陷**

Baseline为了获取可能的帧，针对整个视频进行了搜索。同时，前景提取由于是传统的视觉识别方式，受到汽车反光的影响，Baseline也并未对其进行很好的滤波处理，导致识别后存在大量的识别噪声，识别准确度也并不高，获取语义信息不明确对两车交汇、车辆转弯的等复杂情况不能有很好的识别情况；并且作为相似度度量的直方图数据在某种意义上也不能对其有很好意义的度量，这些情况导致Baseline的预识别情况非常地差。

**4、改进尝试**

在改进尝试上，自己为了更好获得车辆的语义信息，检测上采用了开源的Yolov8模型[[1]](#footnote-1)，其较高的准确性得到了非常不错的检测结果；

同时，针对图像的相似度特征比对，选取了传统的Phash+HammingDistance作为相似度计算，以及通过在VEID相关数据集[[2]](#footnote-2)上训练相应的Siamese Network通过CNN提取特征向量，通过全连接层计算相应相似度。但是经过测试发现，似乎一开始三者的测试结果都不是很好。传统的PHash特征比对由于截图ROI区域不同，相似度检测特征提取并不明显；而后面两个方案由于采用了网络，计算时间开销巨大。自己认为这可能跟实验数据集过小，车辆特征提取不充分，网络泛化能力不够强有关。

图表 1 Siamese network结构图

图表 2 Siamese network 训练过程图

但在继续尝试时发现，Mobilnet(Tensorflow提供的预训练模型)已经能够比较好地找出相似图，这是超出意料的，故最后还是换用了Mobilenet。

**5、最终改进方案**

针对Baseline存在的如上问题，自己做出了如下几点改进：

1. 每60帧进行一次检测（2s），取代逐帧扫描。由于汽车在图中是连续运动的，其本身情形其存在的语义信息可由当前帧拓展至前后10s，所以要获取特定车辆的存在信息，并不需要逐帧扫描，只需要通过预搜索确定其可能存在的帧区间即可，再针对高可能性的存在帧进行进一步搜索。这样，能够大大减少进行相似度比对和图像选择的计算开销，实现计算速度优化；在保证查询图片的存在下，选取其中相似度比对程度最高5个的局部轮廓所在帧，选取其前后共600帧，作为查询结果。
2. 针对特定汽车识别过程中语义特征发掘不明显，换用Yolov8,针对每帧中存在车辆进行识别框检测，根据检测到ROI部分图片再进行相似度比对，检测效果有大幅度提升；
3. 为得到更好的相似度比对结果，自己采用Mobilenet模型提取图像特征，通过计算针对图像特征的特征向量余弦距离得到相似度。

**6、实验结果**

第一次实验结果（不同相似度比对方案）如下：“—”表示效果太差无法测量。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **实验指标** | **Baseline** | **Yolov8+phash** | **Yolov8+Mobilnet** | **Yolov8+SiameseCNN** |
| 准确率/precision | -- | -- | 0.00169 | 0.226 |
| 召回率/recall | -- | -- | 0.000816 | 0.453（3min，special） |
| 时间 | 2775.251533985138 s | 1295.1996388435364 s | 12471.809068918228 s | 28679.39446759224 s |

第二次实验结果如下

实验时间：1151.044.8338985443 s（除校车外）

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **照片序号** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **平均** |
| 准确率/precision | 0 | 0 | 0 | 0 | 0.1378 | 0.0383 | 0.0994 | 0.092 | 0.1986 | 0.1118 | 0.06779 |
| 召回率/recall | 0 | 0 | 0 | 0 | 0.8558 | 1 | 1 | 1 | 1 | 0.7534 | 0.56092 |

经过优化后，能够实现Baseline一半的查询的时间，并且召回率得到了大幅度提升，但准确率仍然不是很高——需要进一步对搜索得到帧区间进行细分判别。但由于自己时间原因，没有时间进行下一步优化了，故暂且把这个实验结果作为阶段性的检测结果。

**7、实验优化方向**

虽然没能够完全针对目前实验结果做进一步优化，但后续的优化方向有了大致思路：

1、针对得到的区间进行进一步搜索，利用汽车移动连续性，可以考虑采用KNN的前景提取器测试这一段的汽车移动提取情况，利用卡尔曼滤波器+匈牙利匹配算法，通过计算IOU（重叠区域占比），得到车辆连续运动的区间；

2、采用VeRi[[3]](#footnote-3)实验数据集，针对相同车辆在不同角度下的拍摄图像，实现汽车的全程跟踪和多角度捕捉。

1. https://github.com/KdaiP/yolov8-deepsort-tracking.git [↑](#footnote-ref-1)
2. D. Zapletal and A. Herout, "Vehicle Re-identification for Automatic Video Traffic Surveillance," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 2016, pp. 1568-1574, doi: 10.1109/CVPRW.2016.195. [↑](#footnote-ref-2)
3. Xinchen Liu, Wu Liu, Huadong Ma, Huiyuan Fu: Large-scale vehicle re-identification in urban surveillance videos. ICME 2016: 1-6 (Best Student Paper Award, Citation=75);

   Xinchen Liu, Wu Liu, Tao Mei, Huadong Ma: A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance. ECCV (2) 2016: 869-884 (Citation=56);

   Xinchen Liu, Wu Liu, Tao Mei, Huadong Ma: PROVID: Progressive and Multimodal Vehicle Reidentification for Large-Scale Urban Surveillance. IEEE Trans. Multimedia 20(3): 645-658 (2018) (Citation=26); [↑](#footnote-ref-3)