

2016 年第一届全国网络舆情分析技术邀请赛

评测大纲

视频人脸识别评测

一、 评测定义和任务说明

本项评测的对象任务是：给定的一组目标人物图像，在待测试的视频中判断是否出现目标人物，以及目标人物出现的位置。

二、 评测数据集

1. 目标人物图像

目标人物图像包含约 30 个人的 100 幅图像，用于标识目标人物及其 ID。

2. 测试视频

测试视频包含 11000 段，其中 1000 段视频包含目标人物，每段视频中只包含一个目标人，该目标人可能出现多次。剩余 10000 段视频**不包含**任意目标人物。每段视频长度从 10 秒到 20 分钟不等。平均 1-2 分钟。

三、 参评系统输入输出文件格式

1. 评测数据格式

本次评测的图像为 JPG 格式，视频为 mp4 格式。

2. 参赛程序的输入、输出文件格式

(1) 输入

比赛时，主办方提供：注册图、测试视频数据、注册图列表、测试视频列表
注册图列表：以 txt 格式提供，每个图一行，如下：

```
pepolelist.txt
D:\pepoleimage\image1.jpg 人物 ID
D:\pepoleimage\image2.jpg 人物 ID
...
```

测试视频列表：以 txt 格式提供，给出所有测试视频的绝对地址，和视频的总帧数，其中总帧数为该视频的最大帧号+1（因为是从 0 开始计数）

Videolist.txt	
D:\test\video1.mp4	360
D:\test\video2.mp4	710
...	

(2) 接口

比赛时，参赛方提供可执行程序，XXX.exe, XXX.sh, XXX.py 等格式，主办方工作人员以执行脚本的形式运行程序。若参赛方希望提供 2 组测试结果，则需要提供 2 个对应的可执行程序。

提供的调用接口为：

程序名 注册图列表 测试数据列表 保存测试结果的文件夹名称

举例：

XXX.exe pepolelist.txt Videolist.txt D:\Result

(3) 输出

允许参赛方在规定测试时间范围内提交最多不超过 2 组评测结果。每组评测结果为一个扩展名为 txt 的文本文件，文件名称形式为“单位名称_人脸识别_第 x 组.txt”。检测结果文件中的每一行对应一个测试视频的人脸识别结果（行之间用换行符分割，各项之间用逗号分割），视频帧号从 0 开始计数，具体形式为：

video_name_i,人物 ID,Similarity,n,(s₁, e₁, sim₁),(s₂, e₂, sim₂),...,(s_n, e_n, sim_n)

其中 video_name_i 为第 i 个测试视频的文件名；ID 为该段视频被识别出的身份即目标人物的 ID；Similarity 为该段视频与该 ID 的相似程度，范围为 0-1；n 表示该 ID 在该段视频中出现的次数，(s_i, e_i, sim_i)表示该 ID 第 i 次出现的起始帧、结束帧、以及与该 ID 的相似度，如果 n 为 0，则后面不需要包含任何(s_i, e_i, sim_i)。

例如：“v0003,Jack,0.8,5,(5, 11,0.1),(15, 29,0.8),...,(51, 60,0.4)” 表示结果判断 v0003 这段视频包含目标人物 Jack，相似度为 0.8，该视频中 Jack 出现 5 次，第一次出现为 5-11 帧，相似程度为 0.1，第二次出现在 15-29 帧，相似程度为 0.8，第 5 次出现在 51-60 帧，相似程度为 0.4。

四、 评价指标

人脸识别考察识别的准确率和误识率，参考指标为执行时间。最终比赛名次以 ROC 曲线面积（算法见下）排名。

1. 人脸识别准确率和误识率

给定 Similarity 的阈值t，

正确识别率 = 正确识别出目标人物且信度大于阈值 t 的视频个数/包含目标人物的视频总个数 $\times 100\%$ 。

误识率 = 信度大于阈值 t 且不包含目标人物的视频数/不包含目标人物的视频总数 $\times 100\%$ 。

正确帧识别率 = 正确识别出目标人物且信度大于阈值 t 的视频帧与标注的包含目标人物的视频帧的交并比的平均值。

识别延迟 = 信度大于阈值 t 且正确识别出目标人物的视频的延迟距离的平均值，其中单个视频的延迟帧数 = 结果中第一个信度大于阈值 t 的目标人物出现片段的首个帧号与标注的目标人物首次出现的帧号间差值的绝对值。识别延迟仅在包含目标人物的视频上计算。

2. 整体性能评测

首先针对视频的识别准确率，遍历不同的 Similarity 的阈值 t ，得到正确识别率 vs 误识率的 ROC 曲线，并取 ROC 上误识率 $[10^{-6}, 10^{-1}]$ 之间的曲线下面积 F_{auc}^{video} 作为准确率的评测指标。基于该指标的排名记为 R_{auc}^{video} 。其次，针对帧的识别准确率，遍历不同的 Similarity 的阈值 t ，得到正确帧识别率 vs 误识率的 ROC 曲线，并取 ROC 上误识率 $[10^{-6}, 10^{-1}]$ 之间的曲线下面积 F_{auc}^{frame} 作为准确率的评测指标。基于该指标的排名记为 R_{auc}^{frame} 。最终综合考虑视频的识别准确率与帧的识别准确率，整体识别准确率排名为（ β 取为 0.8）：

$$R_{final} = \beta \cdot R_{auc}^{video} + (1 - \beta) \cdot R_{auc}^{frame}$$

五、 测试步骤

1. 评测步骤：

- 1) 在调试阶段，参赛方通过自带笔记本登录独立的服务器，进行系统部署与环境搭建，并获取调试数据进行调试。调试完成后示意主办方工作人员。
- 2) 在系统测评阶段，主办方工作人员将测试数据（目标人物图像与测试视频数据）拷贝到服务器上，此后参赛方未经允许不得再操作服务器。由主办方工作人员在特定终端上运行执行程序脚本，记录执行完成所用的总时间。

3) 执行完成后, 主办方工作人员运行评测程序, 记录结果。

2. 技术指标评分:

主办方运行人脸识别性能统计工具, 给出识别性能指标, 测试结果形式如下:

第 1 组结果, 视频识别 R_{final} 为____; 识别延迟为 F_{auc}^{delay} ____;

第 2 组结果, 视频识别 R_{final} 为____; 识别延迟为 F_{auc}^{delay} ____;

参赛方选择第__组结果进行最终排名。

比赛名次以 R_{final} 排名。

六、 注意事项

- 1、参赛方需要在指定时间内完成至少一组测试, 否则没有比赛成绩。
- 2、在比赛过程中, 如果程序中断, 可提示主办方工作人员重新运行参赛程序。参评程序执行过程中, 应打印提示程序中断的信息, 方便主办方工作人员及时处理。除了提示程序中断的信息, 禁止打印任何信息到屏幕上。
- 3、建议参赛程序在程序内部记录 log 文档, 记录程序中断时已跑完的测试数据。如果有相关标识, 程序可以继续跑剩余的测试数据。比赛过程中, 不允许通过调试程序来优化结果。