

T/GDAF

广东省公共安全技术防范协会团体标准

T/GDAF ×××—20××

人像识别系统数据集制作与使用规范

Specification for the production and use of a dataset for portrait recognition systems

(征求意见稿)

2023-09-01

××××—××—××发布

××××—××—××实施

广东省公共安全技术防范协会 发布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 数据集技术要求 1

 4.1 数据来源 1

 4.2 数据规模 2

 4.3 数据安全保密性 3

5 数据集管理规范 3

 5.1 数据采集 3

 5.2 数据标注 3

 5.3 数据存储 5

 5.4 数据使用 6

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由广东省公安厅办公室科技信息化处提出。

本文件由广东省公共安全技术防范协会归口。

本文件起草单位：

本文件主要起草人：

人像识别系统数据集制作与使用规范

1 范围

本文本适用于人像识别系统数据集的制作与管理使用，从数据来源、规模、标注、安全保密性规定了人像识别系统的数据集技术要求，从数据采集、数据标注、数据存储规范使用管理，以确保测试结果准确可靠，为系统的开发和优化提供有力的支撑。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 26238—2010 信息技术 生物特征识别术语

GA/T 893—2010 安防生物特征识别应用术语

GA/T 1325—2017 安全防范人脸识别应用视频图像采集规范

3 术语和定义

GA/T 893—2010和GB/T 26238—2010界定的以及下列术语、定义和缩略语适合于本文件。

3.1

人像识别 vision Human visual recognition

通过人脸信息，人体体态等人体特征判断人物的过程。

3.2

标准数据 standard data

标准数据是测试结果的参考对比标准。

3.3

测试数据 test data

由数据集取出的用于测试的部分数据。

4 数据集技术要求

人像识别系统数据集为测试人员提供充分的测试数据选择。测试人员根据人像识别系统具体要求，从已完成的人像识别数据集中选择对应的数据去完成更充分的测试。

4.1 数据来源

人像识别数据应来自于真实的数据集，包括不同场景、不同状态、不同形态条件下的人像图片。

4.2 数据场景

场景数据要求如下：

- a) 数据要具有多样性,覆盖多种自然环境和人文环境,应考虑天气、光照、场地、周围人群密度等多种条件;
- b) 数据尽量从交通枢纽、购物中心等场景中获取。

4.2.1 数据状态

- a) 环境状态:日照、阴雨、雾霾;
- b) 人员状态:化妆、素颜;
- c) 肤色状态:不同种族人群,例如白种人、黑种人、黄种人等;
- d) 特殊状态:双胞胎、多胞胎、整容状态。

4.2.2 数据形态

- a) 人脸数据形态
 - 1) 人员标准证件照片;
 - 2) 俯拍、侧面 10-60 度干扰状态照片。
干扰类型1:戴口罩、戴墨镜、戴帽子、打伞以及四者各类组合形态;
干扰类型2:阴雨天气、日照天气、雾霾天气三种环境类型。
- b) 人体数据形态
 - 1) 行人正面标准步行照片;
 - 2) 俯拍、侧面 10-60 度正面步行状态干扰照片;
干扰类型1:戴口罩、戴墨镜、戴帽子、打伞以及四者各类组合形态;
干扰类型2:阴雨天气、日照天气、雾霾天气三种环境类型;
 - 3) 行人背面标准步行照片;
 - 4) 俯拍、侧面 10-60 度背面步行状态干扰照片。
干扰类型3:戴帽子、打伞以及两者组合形态;
干扰类型4:阴雨天气、日照天气、雾霾天气三种环境类型。

4.3 数据规模

人像识别系统数据集的数据数量应足够大,能够覆盖不同的情况和场景,以充分测试系统的准确性。应包括但不限于种类数量如表1:

表1 数据规模

序号	干扰因素数量	样品数量	人数
1	0 个干扰因素	不低于 20 万张	不低于 20 人
2	1 个干扰因素	不低于 300 万张	不低于 20 人
3	2 个干扰因素	不低于 500 万张	不低于 20 人
4	3 个干扰因素	不低于 200 万张	不低于 20 人
5	4 个干扰因素	不低于 150 万张	不低于 20 人
6	5 个干扰因素	不低于 120 万张	不低于 20 人
...

4.4 数据安全保密性

- 4.4.1 确保数据来源可靠，以免收集到虚假信息而做出错误的决策；
- 4.4.2 确保收集的数据不会被未经授权的人或机构获取和利用，防止数据泄露和滥用；
- 4.4.3 确保数据的完整性和准确性，以确保所做出的决策和行动是基于正确的数据。

5 数据集管理规范

5.1 数据采集

5.1.1 采集方式

- a) 监控视频采集：对于公共场所监控的视频流，可以通过监控视频流进行人像数据采集；
- b) 手机拍照采集：可以让用户使用手机拍照采集人像数据；
- c) 现场采集：在公共场所、活动现场等地方设置采集设备，对来往行人进行人像采集；
- d) 网络爬虫采集：在网络上搜索人物信息和图片，并进行自动化采集；
- e) 合作伙伴数据合作：与其他相关单位和组织进行数据合作，获取更全面的数据信息。

5.1.2 数据授权

数据采集时需要注意保护用户隐私，遵守法律法规的相关规定，确保数据的合法性和安全性。遵守《中华人民共和国个人信息保护法》、《保障个人资料：跨境资料转移指引》等，同时签订肖像使用授权许可协议。

5.2 数据标注

人像识别数据需要进行标注，标注信息应包括人物的身份、状态等信息，以便对系统进行细致的测试和分析。编号由以下11个字段构成，字段分布见图1：

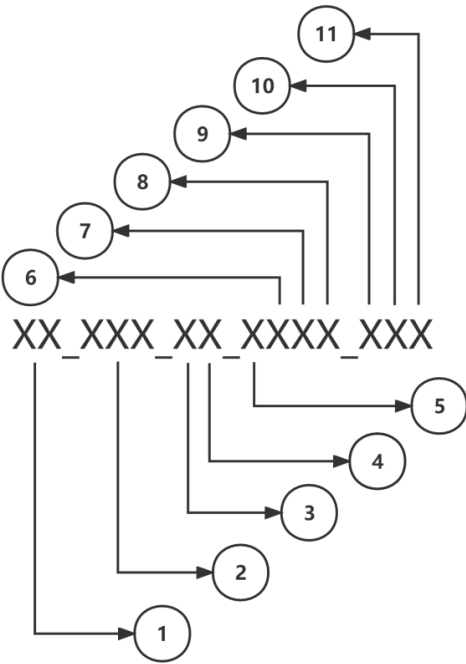


图1 字段分布图

编号字段和标识对应关系见表 2:

表2 编号字段和标识对应表

字段编号	标注	含义
①	ST	标准样
	TE	测试样
②	代码编号	人物代码编号
③	R	雨天
	N	夜晚
	D	白天
	S	晴天
④	u	地铁口
	c	校园、广场
	h	走廊
	f	过马路
	l	沿着正向行走
	r	沿着逆向行走
	g	绿地
	s	车站
	b	商业街
⑤	0	不打伞行走/不戴头套/不戴面具
	1	打伞行走
	h	戴头套
	f	戴面具
⑥	0	不戴帽子
	1	戴帽子
⑦	0	不戴口罩
	1	戴口罩
⑧	0	不戴墨镜
	1	戴墨镜

⑨	X	正光
	Y	逆光
	Z	不区分正逆光
⑩	M	浓妆
	N	素颜
11	S	侧面镜头
	F	正面镜头
	B	背面镜头
/	Z	不区分场景

注：1. 免冠证件照的命名规则为：标准库标识_代码编号_流水号；

2. 人物代码编号按照 26 个英文大写字母顺序排列，用完再增加一位，以此类推，如：

A. B. Y. Z. AA. AB. AY. AZ. BA. BB. ZZ. AAA. AAB.

5.3 数据存储

数据按照难度等级分类存放，数据难度等级按照“☆”个数进行分类，“☆”越多，代表等级越高，数据难度越大。戴口罩、戴墨镜、戴帽子、双胞胎分别为一个干扰因素，打伞、面罩、头套分别为2个干扰因素。根据现有人像识别系统测试数据，整理种类见表3：

表3 难度等级分类表

天气	一般因素	其他因素	难度等级
晴天	墨镜	/	☆
	口罩	/	☆
	帽子	/	☆
	帽子+墨镜	/	☆☆
	帽子+口罩	/	☆☆
	口罩+墨镜	/	☆☆
	帽子+口罩+墨镜	/	☆☆☆
	墨镜	双胞胎	☆☆
	口罩	双胞胎	☆☆
	帽子	双胞胎	☆☆
	帽子+墨镜	双胞胎	☆☆☆
	帽子+口罩	双胞胎	☆☆☆
	口罩+墨镜	双胞胎	☆☆☆
	帽子+口罩+墨镜	双胞胎	☆☆☆☆
	面罩	/	☆☆
	头套	/	☆☆

	打伞	/	☆☆
	打伞+墨镜	/	☆☆☆
	打伞+口罩+墨镜	/	☆☆☆☆
	打伞	双胞胎	☆☆☆
	打伞+墨镜	双胞胎	☆☆☆☆
	打伞+口罩+墨镜	双胞胎	☆☆☆☆☆
雨天	打伞	/	☆☆
	打伞+口罩	/	☆☆☆
	打伞+口罩+帽子	/	☆☆☆☆
	打伞+口罩+帽子+墨镜	/	☆☆☆☆☆
	打伞	双胞胎	☆☆☆
	打伞+口罩	双胞胎	☆☆☆☆
	打伞+口罩+帽子	双胞胎	☆☆☆☆☆
	打伞+口罩+帽子+墨镜	双胞胎	☆☆☆☆☆☆

5.4 数据使用

5.4.1 取样申请

申请人根据具体需求参考数据库清单填写取样申请。取样员根据现有数据情况审核取样申请表，审核通过后数据管理员批准申请。

5.4.2 取样

数据管理员发放对应硬盘的临时密码，设置移动硬盘数据具备只读功能，不可复制和修改。取样员发放文件临时密码，检查数据是否具备只读功能，不可复制和修改。

取样员按照审批通过的取样申请表的内容，根据现有数据使用情况选择目录，制订取样计划表，同时发放使用教程给申请人。数据管理员按照取样计划表将对应数据拷贝到移动硬盘并填写记录数据使用情况，同时设置数据临时密码。

数据服务器配备双重加密机制（数据服务器加密和文件加密），数据服务器加密密码由数据管理员保管，文件加密密码由取样员保管，且数据管理员和取样员不能为同一人。数据每次使用都必须记录数据使用情况。

5.4.3 数据发放

数据管理员发放对应硬盘的临时密码，取样员发放文件临时密码。

5.4.4 数据回收

数据使用完应告知数据管理员，数据管理员查看使用记录，做好登记，删除硬盘的数据。