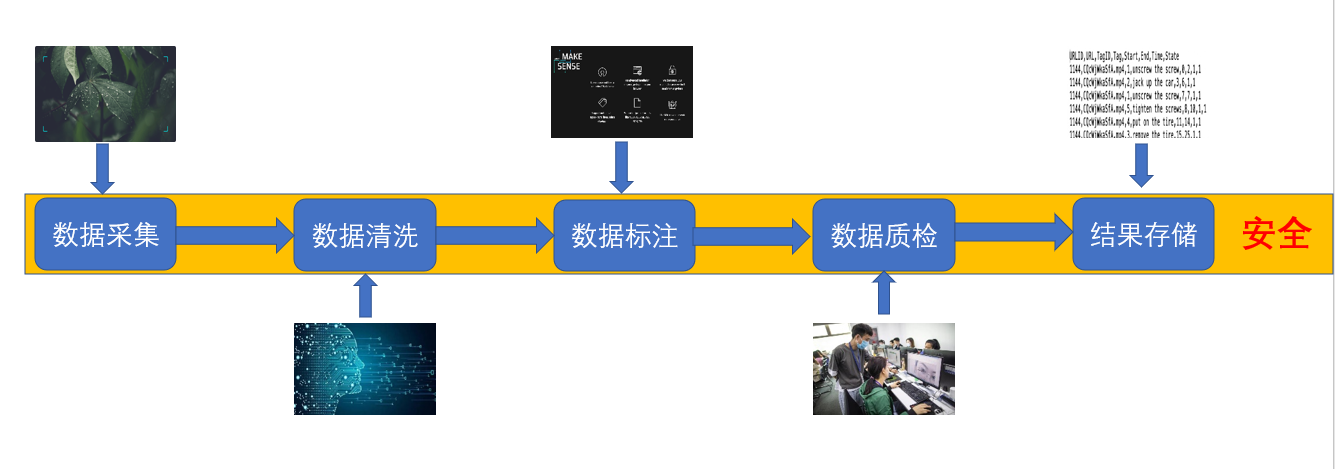
附件1

一、整体解决方案

众所周知，人工智能技术研发，离不开海量的、高质量的数据。伴随着整个产业的发展，尤其是新兴的智能AI体育事业的发展演进，市场对于数据的需求呈指数级增长。近年来，行人检测、轨迹跟踪、动作识别、文本检测、属性识别等计算机视觉分析，已广泛运用到智能体育以及智能AI等领域。但是这些领域的背后，都离不开海量的数据加工和处理。如何有效释放数据价值，成为当前行业面临的一大挑战。

作为数据标注领域的我方，我们有着丰富的数据全流程处理经验，从数据采集、数据清洗、数据标注、结果生产、数据质检、数据安全等流程，有着全方位的处理方法和全局规划。



1）、数据采集

提供本地视频监控设备，支持定时拍照、关键时刻拍照，定点捕捉等多种采集方式，或者从已有的视频数据进行视频抽帧，形成图片数据或者视频文本数据；

2）、数据清洗



通过图片去模糊、图片去重、图片批量裁剪、图片旋转等技术手段完成数据清洗任务。

3）、数据标注

丰富的标注模板，支持图像分类、物体检测、图像分割、文本分类、音频分类、视频分类等丰富易用的标注工具，以及多种不同标注软件，为标注任务提供多种备选和合适的方案。此外提供人机交互协作标注，最高降低90%的标注成本。目前智能标注已支持物体检测、图像分割等数据类型。

4）、数据质检

提供全方位的标注检测技术，第一轮员工自检，后两轮专业质检员质检。为标注精度和准度提供全方位护航。

5）、数据导出。

支持多种脚本导出，如XML、JSON、Mask、txt等。

6）、数据标注技术支撑。

提供三种不同的标注工具，MaskSence、ViBAT、COIN，全方位辅助数据标注。提供完善的数据加密及隔离存储安全技术方案，为数据安全保驾护航。

二、实现方案及关键流程

1、数据标注流程

完成数据标注流程，需要三种不同的角色提供服务，分别是标注员、审核员、管理员。1）**数据标注人员**：标注员负责标记数据；2）**数据质检人员**：审核员负责审核被标记数据的质量；3）**数据管理人员**：管理人员、发放任务、统计工资；

数据标记流程通常如下

1）任务分配：假设标注员每次标记的数据为一次任务，则每次任务可由管理员分批发放记录，也可将整个流程做成“抢单式”的，由后台直接分发。

2）标记程序设计：需要考虑到如何提升效率，比如快捷键的设置、边标记及边存等等功能都有利于提高标记效率。

3）数据标注：标注员在规定时间内完成数据标注；

4）进度跟踪：程序对标注员、审核员的工作分别进行跟踪，可利用“规定截止日期”的方式淘汰怠惰的人。

5）质量跟踪：通过计算标注人员的标注正确率和被审核通过率，对人员标注质量进行跟踪，可利用“末位淘汰”制提高标注人员质量。

2、数据采集

数据采集根据采集数据的类型可以分为不同的方式，主要方式有：爬虫、录入、导入、接口等。爬虫的主要目的是获取开放数据，比如网页的内容数据，或者特定行业的公开数据，以此形成标注数据。录入以及导入为我们传输到标注软件上的文本、图片或者视频数据。

3、数据处理

通过图片去模糊、图片去重、图片批量裁剪、图片旋转等技术手段完成数据清洗任务。如过滤清晰度较低的图片，保证数据质量、过滤大量重复的图片，提高关键图片处理效率、批量裁剪图片中的无关元素，提升数据质量、校正采集图片的角度，方便进行下一步的处理。

提供的数字图像处理技术主要包括如下内容：几何处理、算术处理、图像增强等内容。同时在标注之前，形成标注标准文档，标注员需严格按照标注标注标准进行数据标注。

4、标注方法

主要采取MakeSense在线标注工具，ViBAT标注软件，CIOA在线标注软件进行数据标注，具体的操作流程和操作方法见附件二、附件六。

5、数据质检

数据质检旨在对数据集中的图像数据或者文本数据进行质量检测，通过提供客观指标，为对数据集的下一步操作（标注、清洗等）进行参照引导。

整体质检将包括对原图、标注信息两个层面的数据质量信息进行检查，在标注之前，采用人工查验方法检查待标注数据是否合格。在标注完成之后，采用标注员一轮自检以及专业质检员根据形成的标注标准文件，进行两轮质检的检测标注进行数据检查。

6、标注结果输出

标注的结果主要以文本格式输出以及标注Mask形成的RGB图片输出，文本格式主要包含txt文本、json文本、xml文本，RGB图片主要以PS软件形成的图片数据进行保留。同时提供不同的脚本文件，方便不同文本格式之间进行格式转化，如将txt文本转化成json文本。

7、数据安全性解决方案

为[数据处理系统](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E5%A4%84%E7%90%86%E7%B3%BB%E7%BB%9F/8227413)建立和采用的技术和管理的安全保护，保护[计算机硬件](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E7%A1%AC%E4%BB%B6/5459592)、软件和数据不因偶然和恶意的原因遭到[破坏](https://baike.baidu.com/item/%E7%A0%B4%E5%9D%8F/32871)、更改和泄露。一是数据本身的安全，主要是指采用现代[密码算法](https://baike.baidu.com/item/%E5%AF%86%E7%A0%81%E7%AE%97%E6%B3%95)对数据进行主动保护，如数据保密、[数据完整性](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E5%AE%8C%E6%95%B4%E6%80%A7)、双向强[身份认证](https://baike.baidu.com/item/%E8%BA%AB%E4%BB%BD%E8%AE%A4%E8%AF%81)等，二是数据防护的安全，主要是采用现代信息存储手段对数据进行主动防护，如通过[磁盘阵列](https://baike.baidu.com/item/%E7%A3%81%E7%9B%98%E9%98%B5%E5%88%97)、数据备份、[异地容灾](https://baike.baidu.com/item/%E5%BC%82%E5%9C%B0%E5%AE%B9%E7%81%BE)等手段保证数据的安全，数据安全是一种主动的包含措施，数据本身的安全必须基于可靠的加密算法与安全体系，主要是有[对称算法](https://baike.baidu.com/item/%E5%AF%B9%E7%A7%B0%E7%AE%97%E6%B3%95)与[公开密钥密码体系](https://baike.baidu.com/item/%E5%85%AC%E5%BC%80%E5%AF%86%E9%92%A5%E5%AF%86%E7%A0%81%E4%BD%93%E7%B3%BB)两种。

提供完善的数据加密及隔离存储安全技术方案，为数据安全保驾护航。如生产标注前对每位和数据相关的人员签署协议保证数据安全，私有数据交付后即焚，保障每份数据的隐私合法性，严格遵守可靠的生产流程。如提供专业的标注生产环节，在工作期间，禁止使用和标注任务无关的电子产品，严防数据外泄。如提供专业的数据存储系统，防止原始数据以及标注结果丢失和损坏，标注结果每两小时提交一次，防止意外导致标注结果丢失。