深度学习平台在汽车智能战略中的应用

上汽集团云计算中心

上海帆一尚行科技有限公司

上汽集团战略背景

电动化

智能网联化

国际化

共享化

没有智能的制造是没有灵魂的

具体业务场景介绍

自动驾驶感知

零部件质量检测

智能座舱

生产线智能化

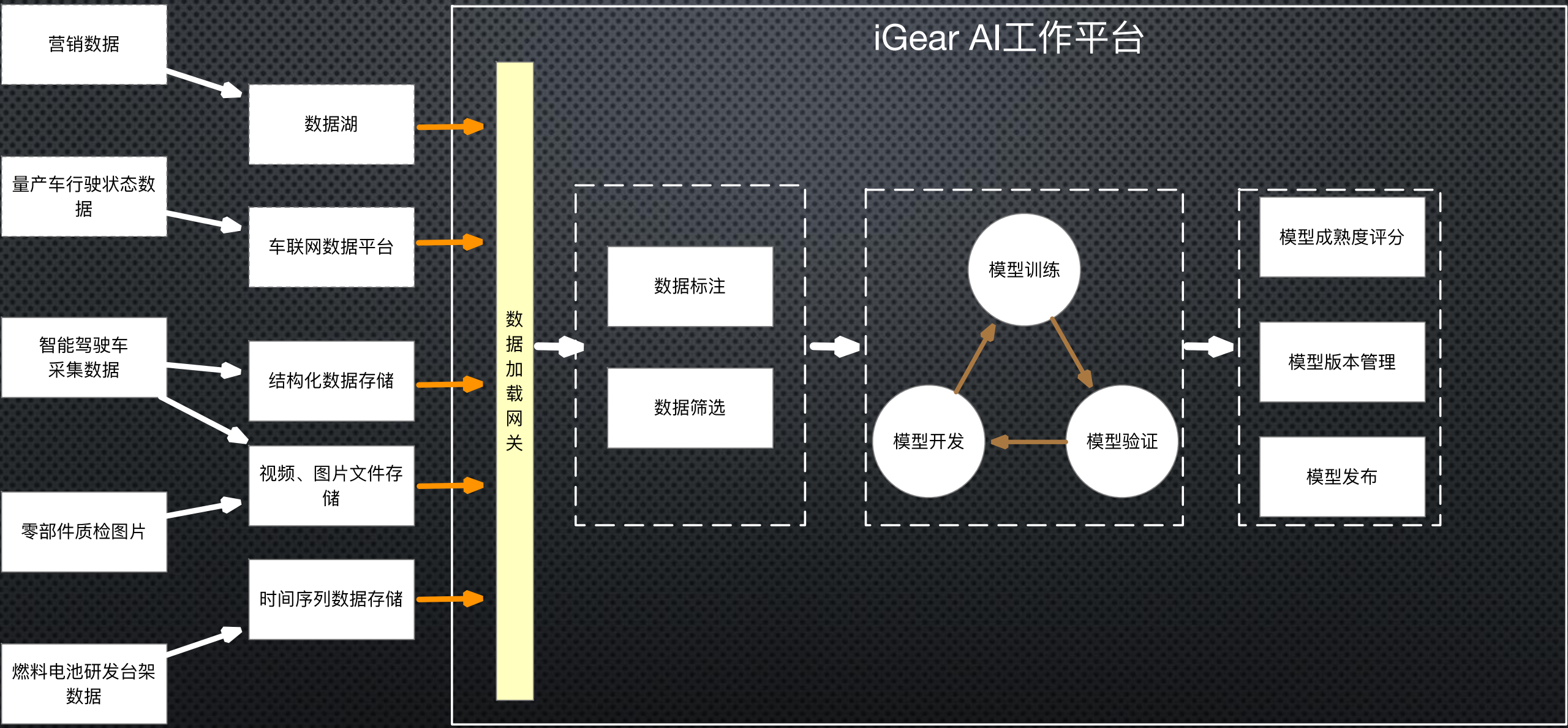
我们该做什么，怎样做？

需要一个机器学习平台来开发智能化应用

iGear AI的一般流程

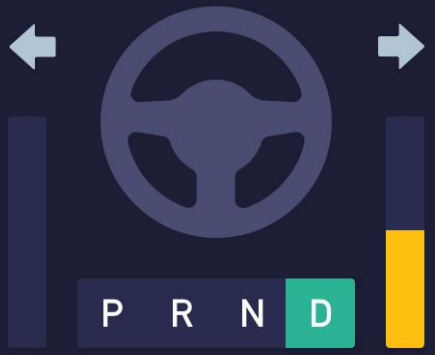
交付一个产品的速度，不如响应变化的速度重要。

好的系统是用出来，不是计划出来的


营销数据
量产车行驶状态数据
智能驾驶车采集数据
零部件质检图片
燃料电池研发台架数据
数据湖
车联网数据平台
结构化数据存储
视频、图片文件存储
时间序列数据存储
iGear AI工作平台
数据加载网关
数据标注
数据筛选
模型训练
模型开发
模型验证
模型成熟度评分
模型版本管理
模型发布

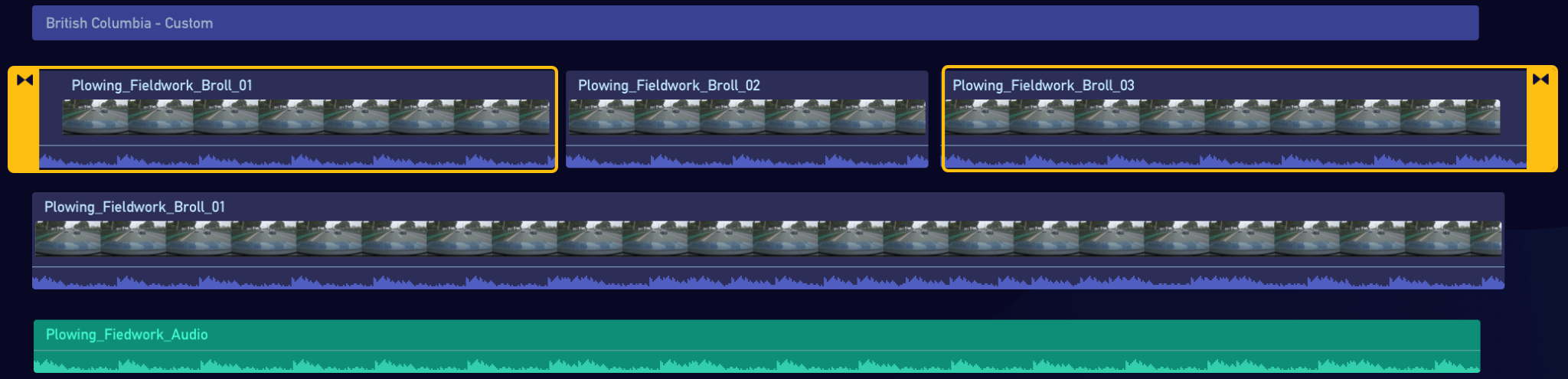
数据筛选

Driver Control



Vehicle Motion

VX :	86.8 [KPH]
AX :	0.1 [m/s2]
YR :	-0.1 [°/s]
Cury :	-0.1 [1/km]
Eng :	1707 [RPM]



数据标注

管理员

标注员

审计员

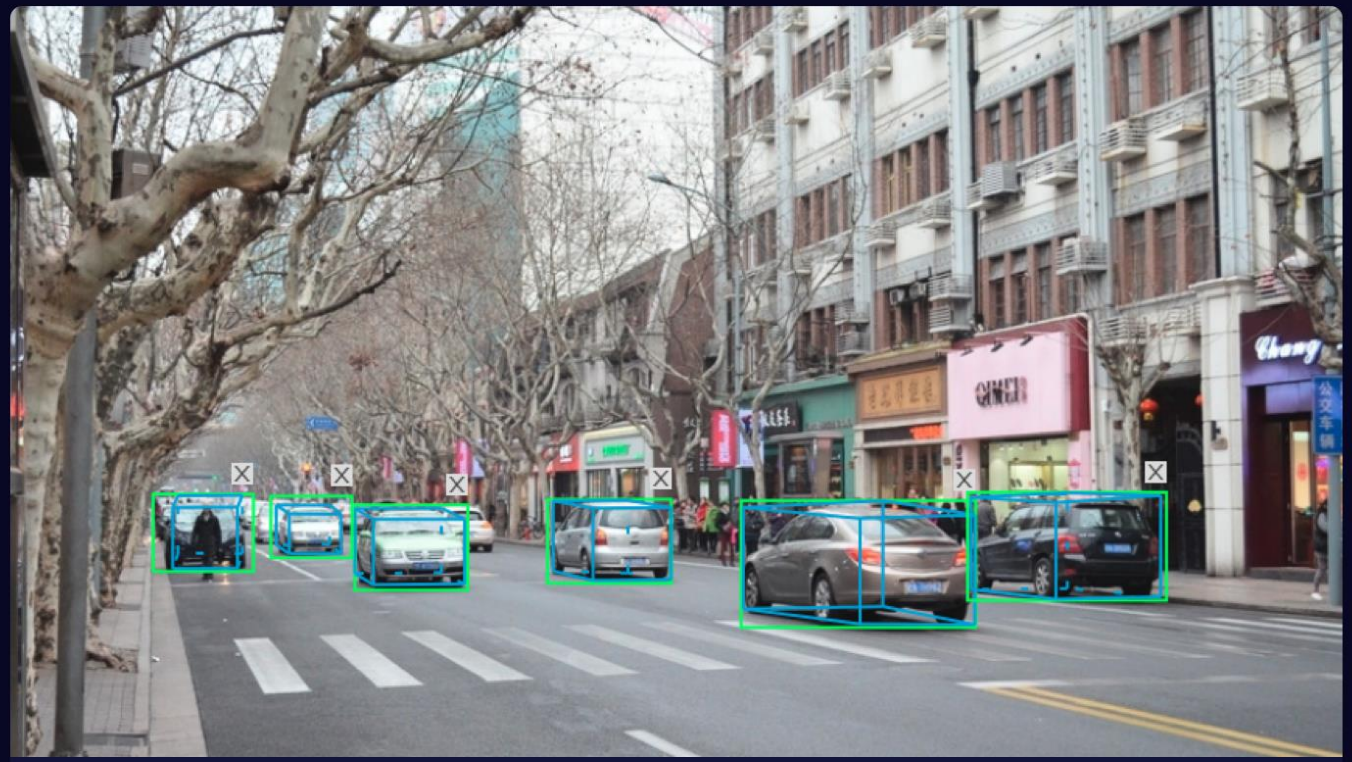
目标追踪 | 3D 标注 | 语义分割

数据标注

管理员

标注员

审计员



目标追踪 | 3D 标注 | 语义分割

数据标注

管理员

标注员

审计员

目标追踪 | 3D 标注 | 语义分割

数据标注

管理员

标注员

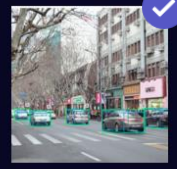
审计员

任务信息

创建任务 2019/04/25 11:27:31

标注完成 2019/04/25 11:35:24

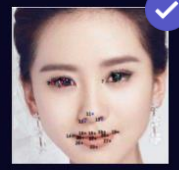
审计完成 2019/04/26 10:01:25

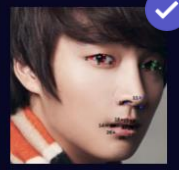

imageF1ile_201 9030…2301.jpg

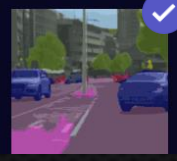
imageF1ile_201 9030…2301.jpg

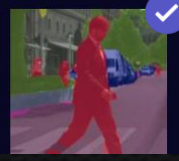
imageF1ile_201 9030…2301.jpg

imageF1ile_201 9030…2301.jpg


imageF1ile_201 9030…2301.jpg


imageF1ile_201 9030…2301.jpg





审计合格

合格线为 95%

任务名称	seamen-3D-pre
任务类型	人脸标注-27点
任务标签	–
任务团队	标注团队1
文件类型	图片
文件数量	7
审计员	余仕墨
截止日期	2019/04/30
错误索引	浏览

模型训练

输入数据集，调用GPU集群进行训练，输出神经网络模型

Base on NVIDIA DGX-1　Cluster

发布导出

车

OTA

边缘设备

PC 服务器

iGear SDK and Workflow

01

iGear SDK登录，浏览数据信息

02

开发数据预处理代码

03

开发模型训练任务代码

04

本地构建预处理及训练所需docker镜像

05

iGear SDK提交Wrorkflow至iGear平台

06

iGear SDK查看训练过程及结果

iGear SDK and Workflow

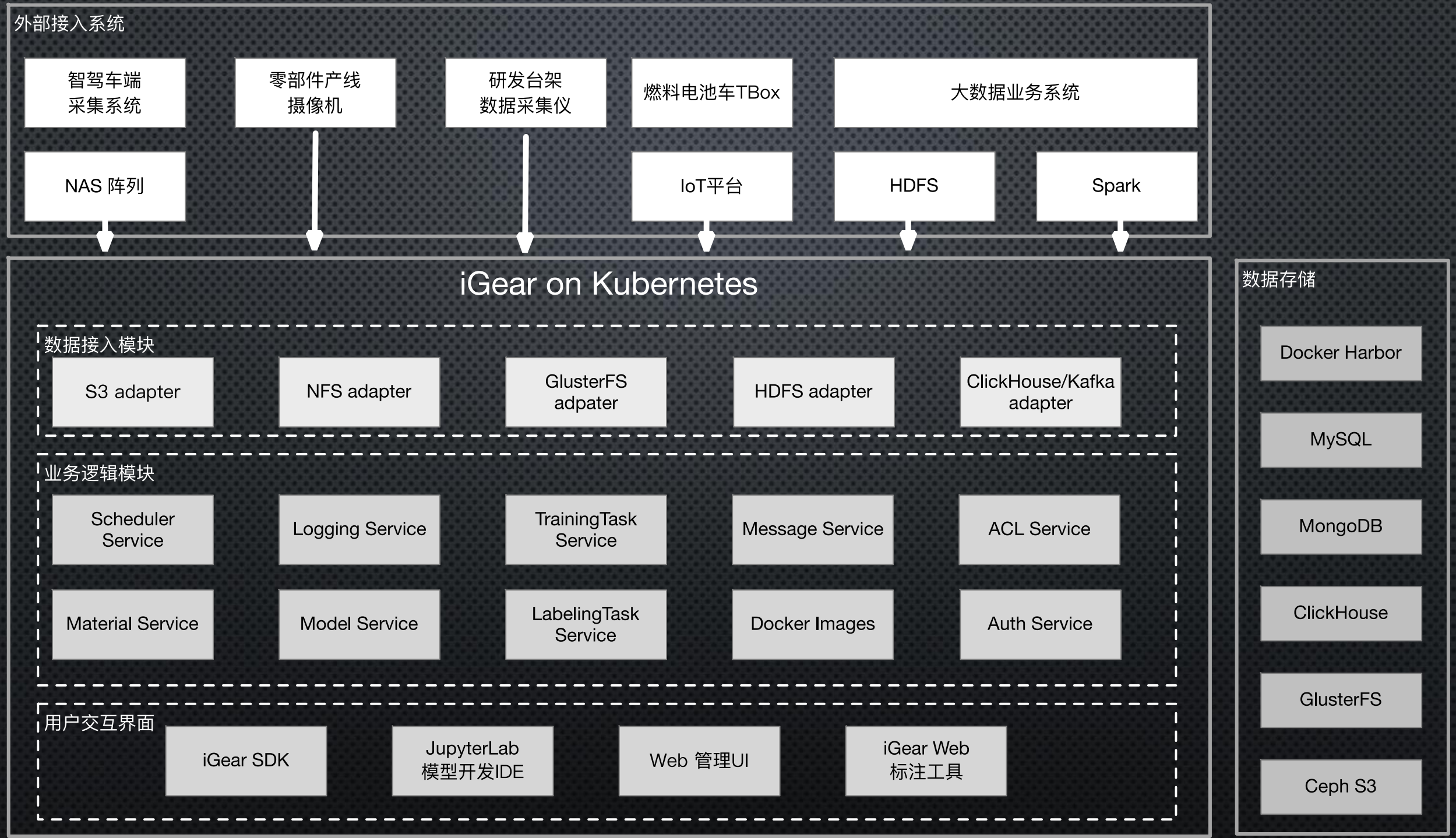
```yaml
task:
  task_name: sdk-11-21-1
  model_name: model-11-21-0
  use_gpu: True
  resource_level: MINMAL
  task_type:
    name: DETECTION
    secondary_task_type: ''
  input:
    dataset_id: 348
  share_type: GROUP
  priority: LOW
  hyper_parameters:
    MAX_ITER: 30
  description: train demo
```

```
ubuntu@qianzhan-se-02:~$ igear login
username: 19921369606
password:
success login
ubuntu@qianzhan-se-02:~$ igear dataset ls
  id  name                size  taskType         shareType    createDate
----  ---------------  ------  --------------  -----------  -------------------
 323  test-ruwei         1487  LINE_DETECTION  TENANT       2019-10-14 08:51:42
 187  dms-train-test1   10059  DETECTION       GROUP        2019-07-05 09:53:53
ubuntu@qianzhan-se-02:~$ igear dataset info 323
  id  name        taskType        secondaryTaskType    datasetFormat    datasetStatus
----  ----------  --------------  -------------------  ---------------  ---------------
 323  test-ruwei  LINE_DETECTION  LANE_LINE_SPLINE     TXT              COMPLETED
ubuntu@qianzhan-se-02:~$ igear task ls
id    name    task_type    status    priority    create_date
----  ------  -----------  --------  ----------  -------------
ubuntu@qianzhan-se-02:~$ igear task
Usage: igear task [OPTIONS] COMMAND [ARGS]...

  Subcommand for training task operations.

Options:
  --help  Show this message and exit.

Commands:
  info   Echo training task detail info by id.
  logs   Download training task logs by id.
  ls     List all training task.
  train  Deploy training task with yaml.
```


外部接入系统
智驾车端
采集系统
零部件产线
摄像机
研发台架
数据采集仪
燃料电池车TBox
大数据业务系统
NAS 阵列
IoT平台
HDFS
Spark
iGear on Kubernetes
数据接入模块
S3 adapter
NFS adapter
GlusterFS
adpater
HDFS adapter
ClickHouse/Kafka
adapter
业务逻辑模块
Scheduler
Service
Logging Service
TrainingTask
Service
Message Service
ACL Service
Material Service
Model Service
LabelingTask
Service
Docker Images
Auth Service
用户交互界面
iGear SDK
JupyterLab
模型开发IDE
Web 管理UI
iGear Web
标注工具
数据存储
Docker Harbor
MySQL
MongoDB
ClickHouse
GlusterFS
Ceph S3

目前使用情况

存储数PB级数据

每辆车每天产生1T数据，10辆采集车在一个月内即达到300T，目前已积累数PB级不同路况数据

标注团队分布在全国各地

原始数据的筛选和标注是异常繁重的工作，在尽可能自动化处理的情况下，工作量仍然大得惊人，这些都将成为智能的源动力

管理大型GPU服务器集群

每台服务器4张~8张NVIDIA Tesla GPU卡，数百张GPU卡加上匹配上百台PC服务器组成异构的工作集群

同时运行上千个训练任务

通过带优先级的队列可同时运行上百个训练任务，包含不同的训练模型、不同的资源配置

支撑近百名算法工程师的开发工作

自有开发团队成员加上合作算法开发人员均通过Web化的开发环境进行调试

产品不断迭代更新

精益和持续迭代的思路贯穿始终，不断支持人工智能的开发集成

几点建议

1. 数据是核心，构建数据收集、标注处理的能力是关键
2. AI业务对技术平台提出各种定制化的要求，需要建立持续迭代开发的能力
3. 加快推进商业价值驱动的场景化探索和落地：场景→数据→训练→落地

iGear团队的一些技术文章

1. ［凡益堂］AI的一般套路
2. ［凡益堂］如何在Kubernetes实现GPU调度及共享
3. ［凡益堂］ iGear平台DevOps实践
4. ［凡益堂］ iGear平台权限管理架构解读
5. …