



曙光人工智能管理平台

SothisAI V2.3

产品简介

平台逻辑架构

应用开发部署 (图像识别, 语音识别, NLP...)

深度学习软件框架 (Caffe, TensorFlow, MxNet...)

集群任务调度 (Slurm, Kubernetes, Mesos...)

资源分发部署 (容器化, 虚拟化...)



计算资源



存储资源



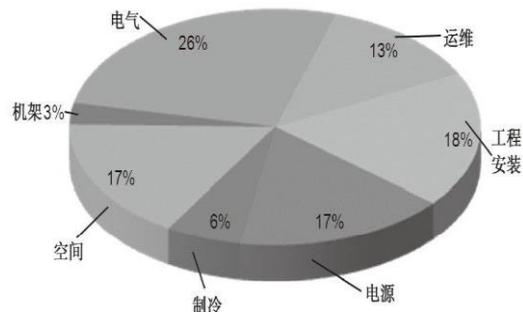
网络资源



人工智能研发痛点

■ 计算服务成本

- 集群购买及建设成本高
- 集群后期运行及维护成本高



■ 框架与部署技术门槛

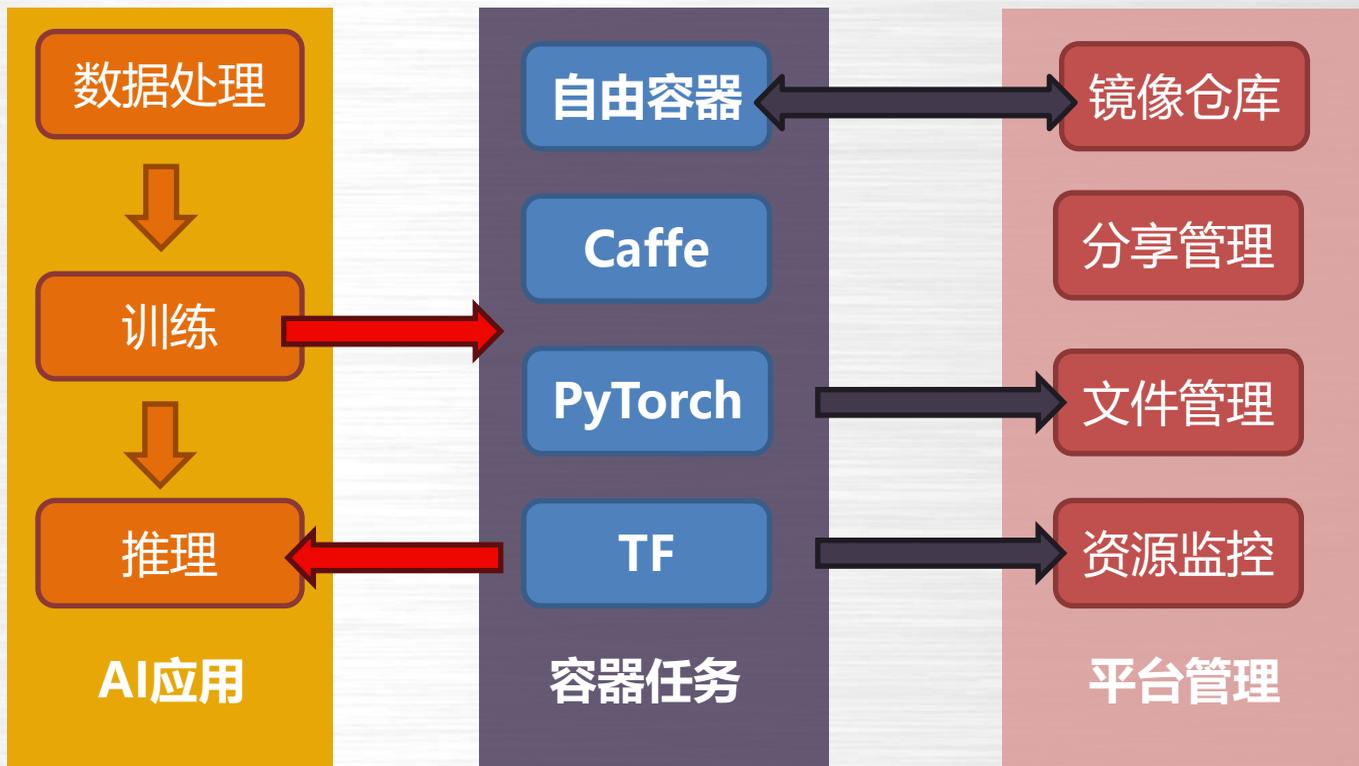
- 集群部署及运维需要相应技术基础
- 深度学习框架使用需要相应技术积累



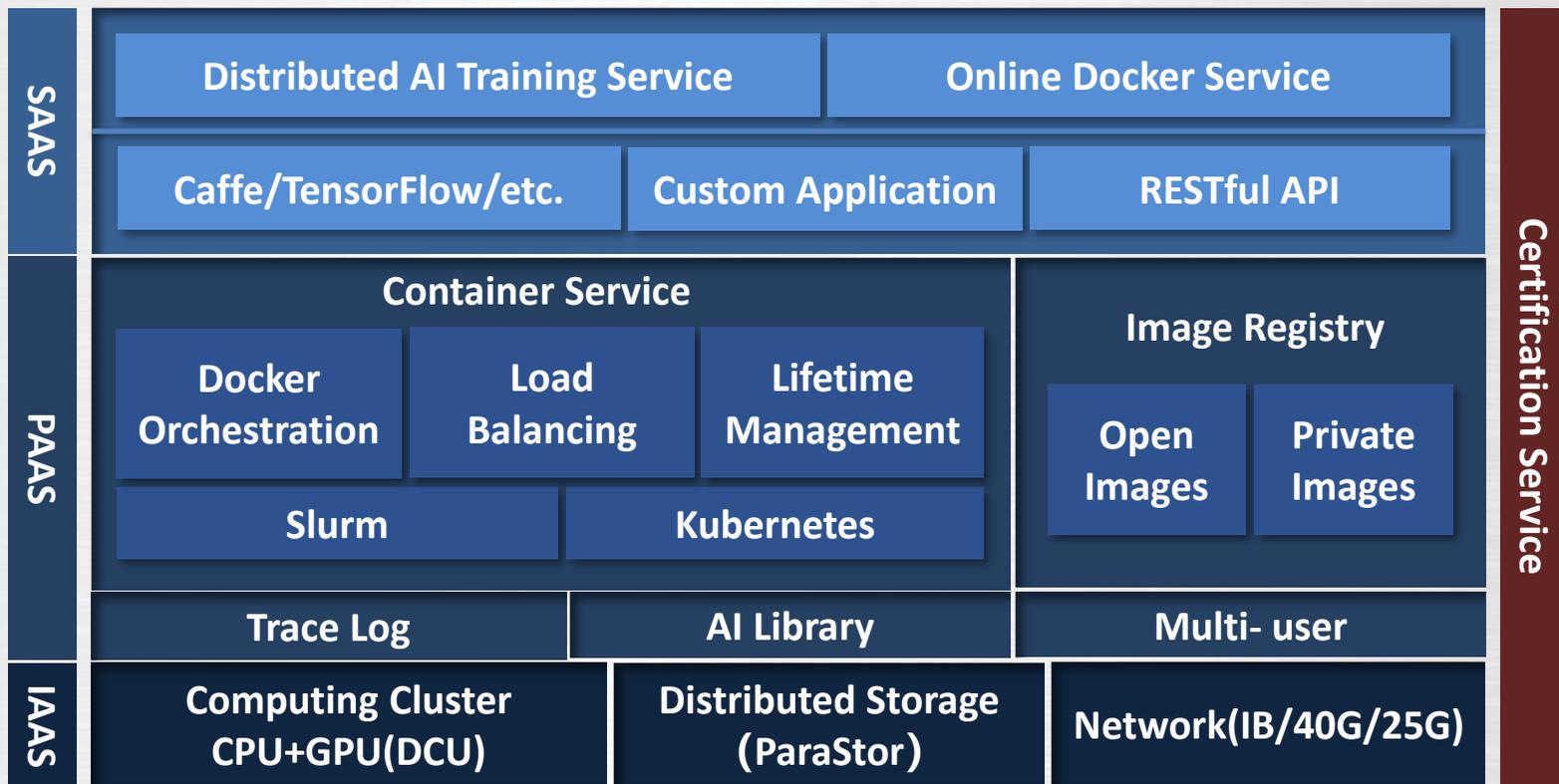
SothisAI功能组成



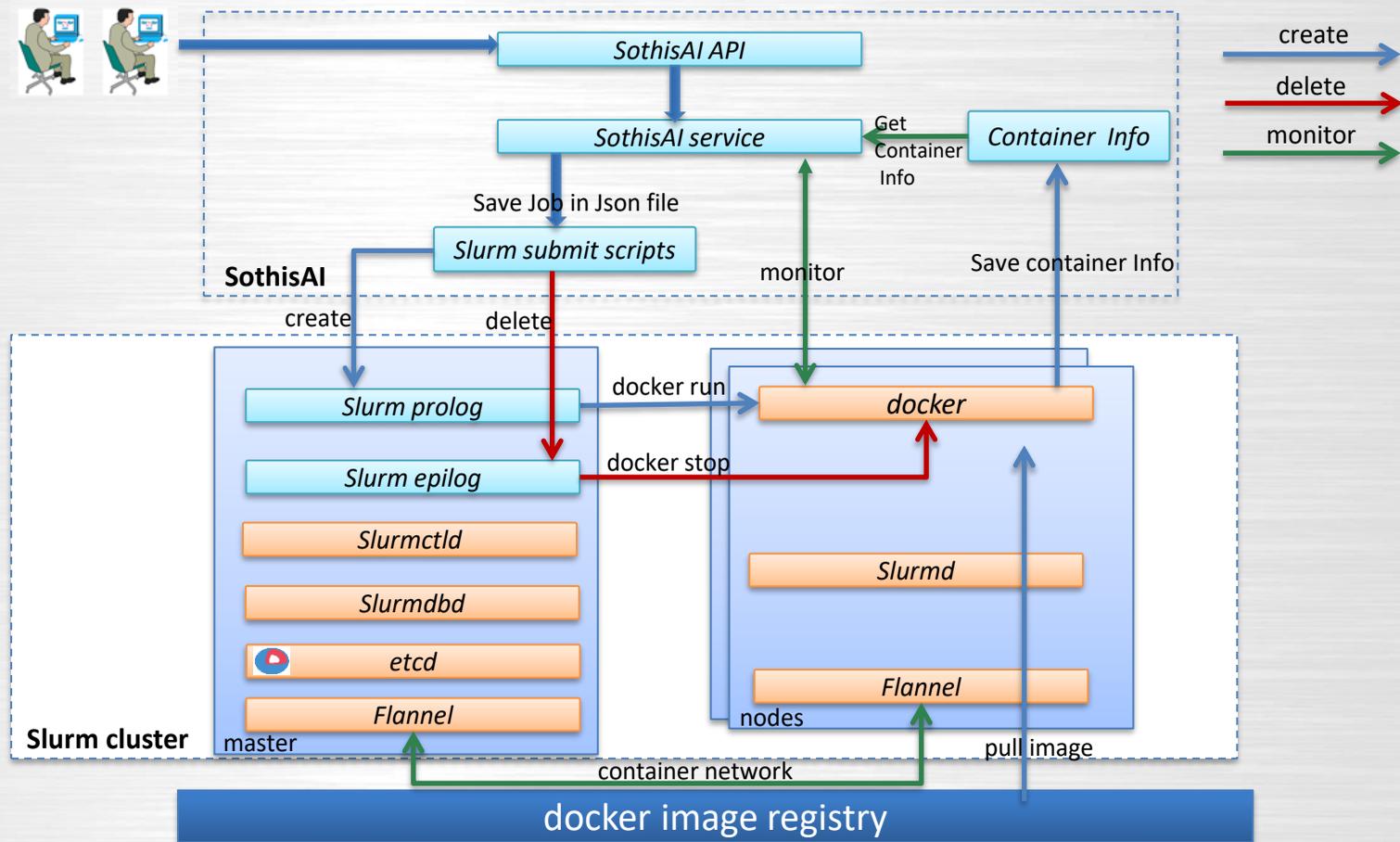
SothisAI主要功能结构



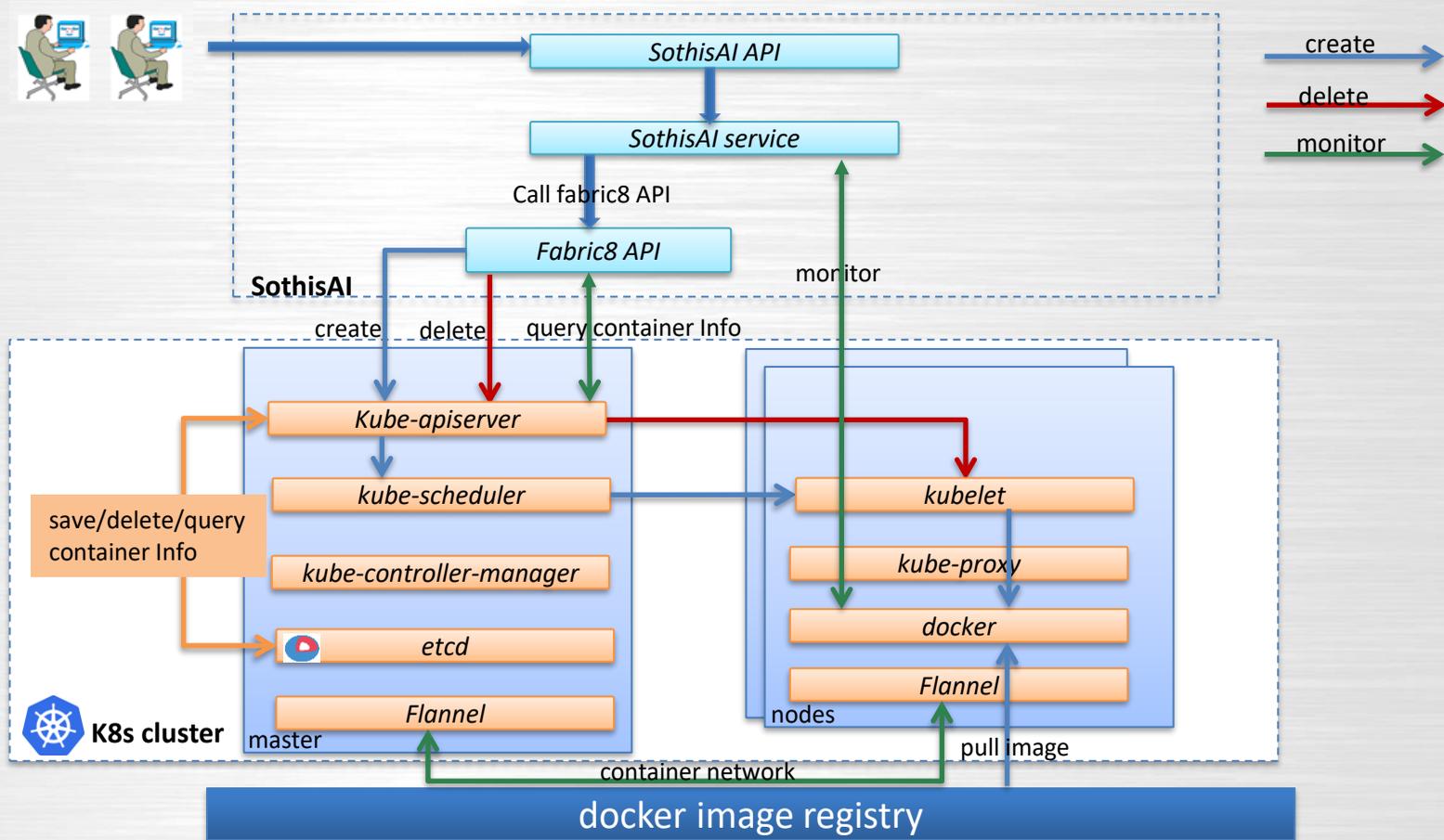
SothisAI架构



SothisAI Slurm Scheduler



SothisAI K8S Scheduler



SothisAI用户场景与核心优势

用户场景

AI转型期的**传统企业**，需要借助AI计算平台实现业务转型

中小型**AI创业企业**，聚焦算法构建，平台运维能力薄弱

学校和**科研院所**，有AI相关项目，或构建AI教学平台

核心优势

经过超大规模集群，上万加速器场景验证

全面支持DCU，开发环境同步集成

容器化环境部署，支持主流深度学习框架，具备丰富行业数据集

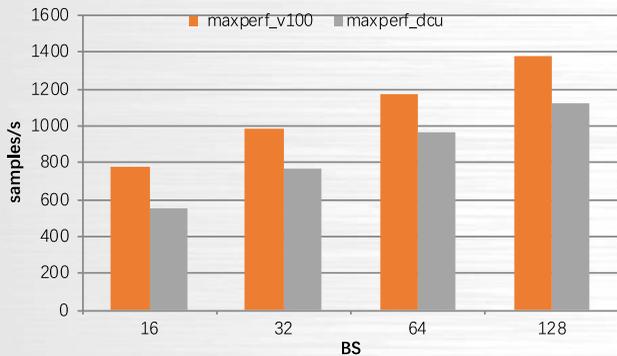
数据集、训练、调参、推理，AI开发全流程覆盖

DCU深度学习框架支持

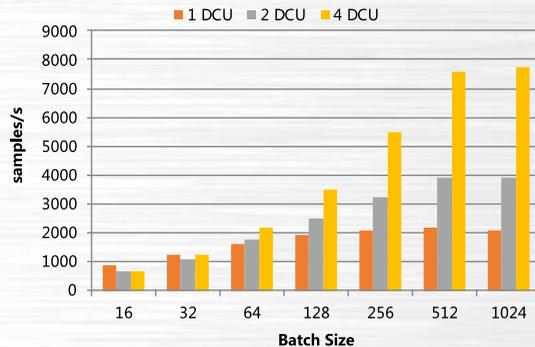
- 目前主流的深度学习框架Tensorflow、PyTorch、MXNet已可以在DCU上运行;
- 从测试结果来看, 部分TensorFlow测试DCU是V100性能的80%
- 基于Tensorflow完成了3000块DCU卡的大规模测试, 积累了性能与准确性调优方法

框架	单机单卡	单机多卡	多机多卡	备注	应用	网络	TensorFlow	PyTorch	MXNet
Tensorflow	√	√	√	horovod	Compute-Vision	ResNet50/VGG16/inception-v3	√	√	√
Pytorch	√	√	√	gloo	Object-Detection	Faster-RNN-ResNet50	√		
Mxnet	√	√	验证阶段		NLP	Bert/GPT-2	√	√	

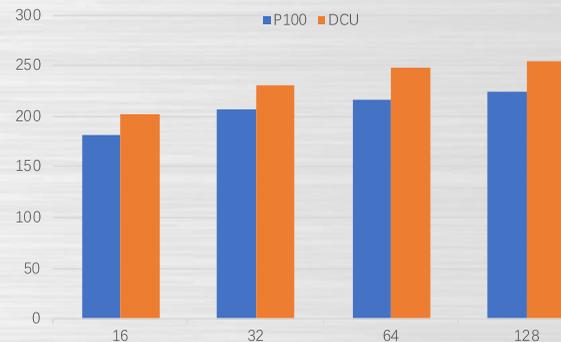
Tensorflow V100-DCU Resnet50 FP32性能对比测试



TensorFlow DCU Alexnet FP32单机多卡性能测试



Pytorch P100-DCU Resnet50 FP32性能对比测试



试用环境

www.hpccube.com



Login

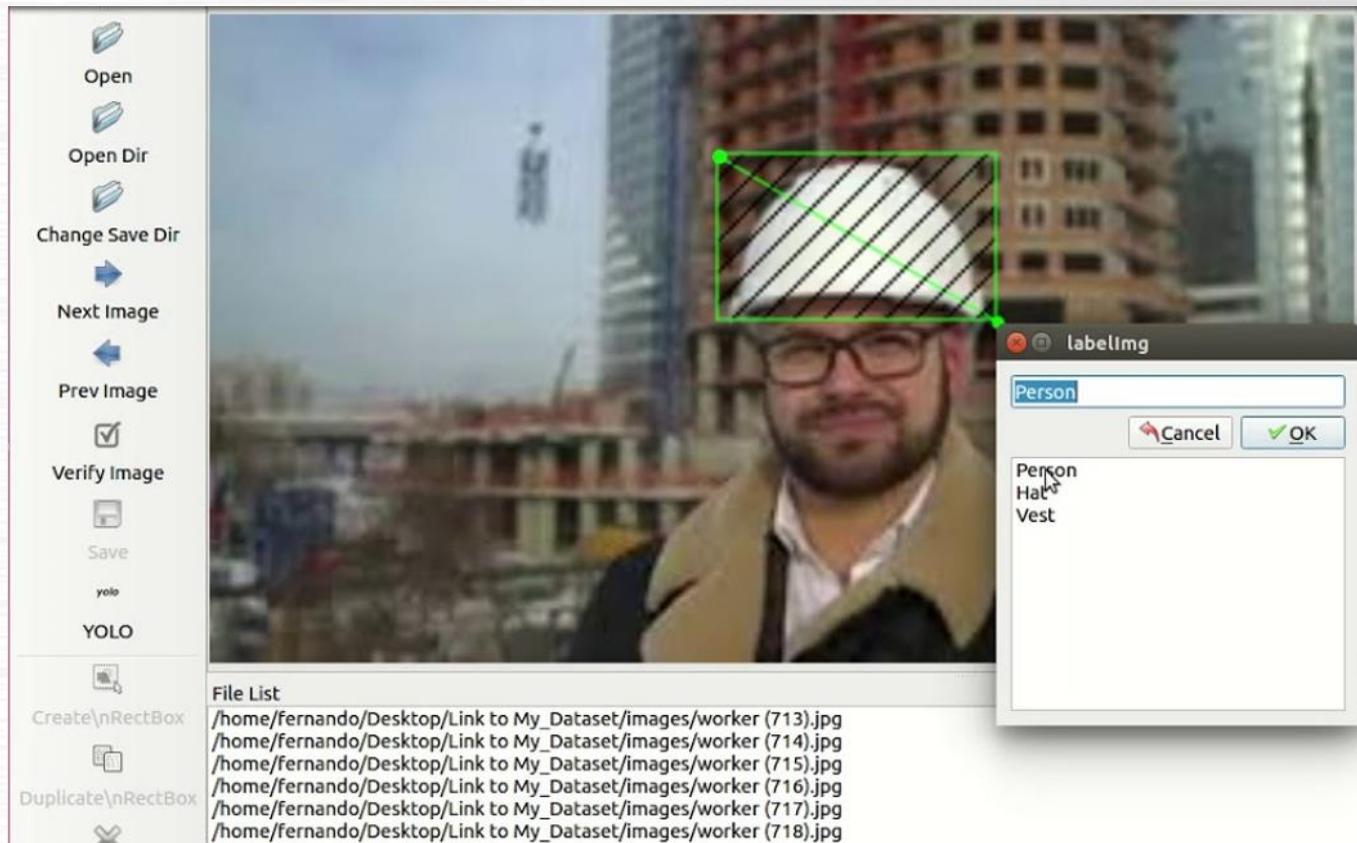
创世纪 算未来

企业级分布式深度学习平台 | 人工智能解决方案 | 一站式深度学习解决方案

产品详情

数据集标注

- 支持目标检测、图像分类等应用场景
- 支持VOC、YOLO等存储格式
- 可根据已训练模型自动标注，加快数据集处理效率



超参数自动调优

- 集成TPE超参数自动调优算法
- 支持多参数同步调优
- 支持分布式多机多卡参数调优



模型验证

- 支持图像分类、目标检测、语义分割等多种应用场景
- 支持单张和批量操作
- 在线图形化展示

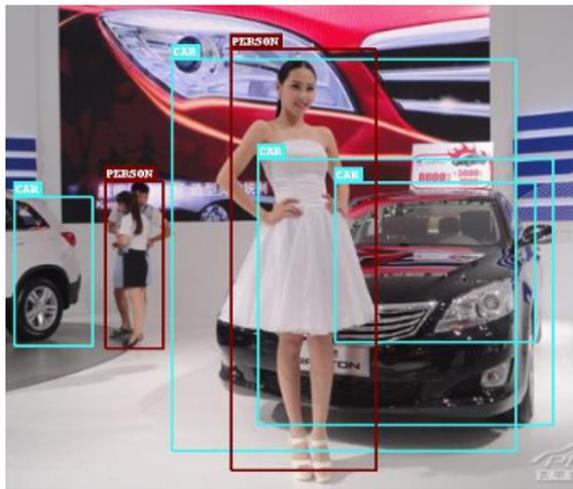
infer_ADE_val_00001519.jpg



ADE_val_00001519.jpg



car.jpg



412.jpg

Class Inferred dragon
Probability 0.94



413.jpg

Class Inferred dragon
Probability 0.98



416.jpg

Class Inferred dragon
Probability 0.97



417.jpg

Class Inferred dragon
Probability 0.95



500.jpg

Class Inferred elephant
Probability 0.57



501.jpg

Class Inferred elephant
Probability 0.69



504.jpg

Class Inferred elephant
Probability 0.56



505.jpg

Class Inferred elephant
Probability 0.58



服务发布

- 发布模型为web推理服务
- 面向第三方应用用户
- 容器化部署，资源按需定义

推理进度: 3 / 3 处理用时: 7 s

上传图片

下载结果



主流框架支持

- TensorFlow、PyTorch、Caffe等主流深度学习框架支持
- 图形化开发流程，全框架DCU适配，代码无感知迁移



环境自定义

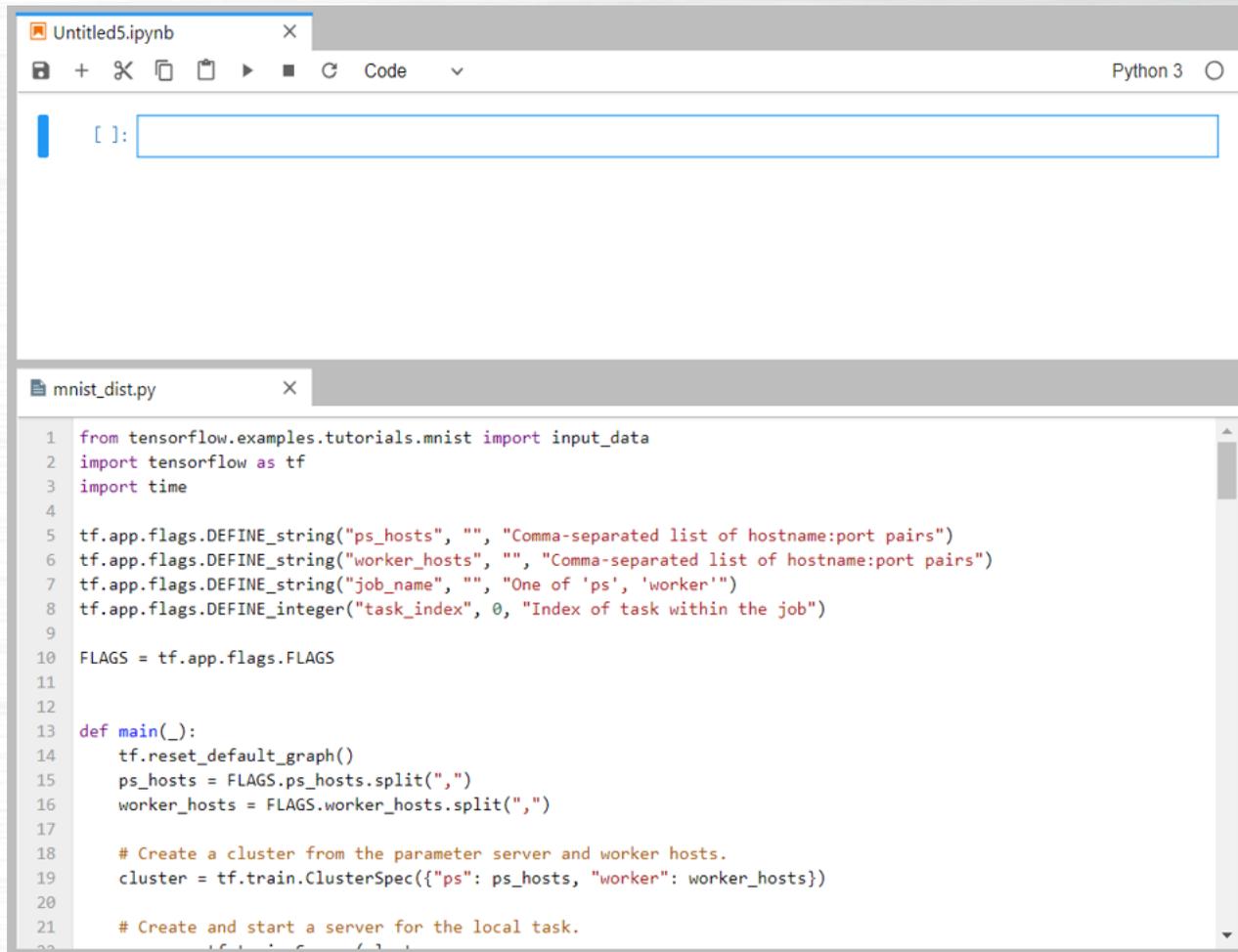
- 框架环境自定义，随时修改，按需固化
- 镜像库在线管理，随时订阅公有库资源，支持指定用户共享

The screenshot displays a management interface for environment images, divided into '公有库' (Public Library) and '私有库' (Private Library) tabs. The interface includes action buttons for '选择所有' (Select All), '订阅' (Subscribe), and '取消订阅' (Unsubscribe), along with a '选择类型' (Select Type) dropdown. A grid of environment cards is shown, each with a unique icon, name, user, and creation timestamp. The first card is highlighted with a blue border and a checkmark. A tooltip is visible over the second card in the middle row.

Environment Name	User	Creation Time	Category
v1:v2.0	nsccsz_leifl11	2019-12-01 16:02:25	Other
doubi:v1.2	liu_18021504833	2019-12-01 16:01:26	Other
mongo:v4.1	admin	2019-11-30 01:42:47	Other
rocm_tf_ssh:rocm3.3_tf2.1_v1.0	admin	2020-04-22 11:23:07	Jupyter
rocm_pytorch:v1.3_rocm2.9_py3.6	admin	2020-04-07 16:35:30	Jupyter
pytorch:rocm2.9_py3.6	admin	2020-04-03 20:04:13	Pytorch
rocm_jupyter:dcu	admin	2020-02-22 15:52:59	Jupyter
rocm_mxnet:v1.4_py3.6	admin	2020-02-22 15:52:59	Jupyter
rocm_pytorch:v1.3_py3.6	admin	2020-02-22 15:52:59	Jupyter

多种接入方式

- 支持SSH、Jupyter等多种接入模式
- 集成Jupyter开发工具，可随时在线接入
- 页面SSH支持shell风格快捷键操作



The screenshot displays a Jupyter Notebook environment. The top window is titled 'Untitled5.ipynb' and contains a single code cell with a cursor. The bottom window is titled 'mnist_dist.py' and contains the following Python code:

```
1 from tensorflow.examples.tutorials.mnist import input_data
2 import tensorflow as tf
3 import time
4
5 tf.app.flags.DEFINE_string("ps_hosts", "", "Comma-separated list of hostname:port pairs")
6 tf.app.flags.DEFINE_string("worker_hosts", "", "Comma-separated list of hostname:port pairs")
7 tf.app.flags.DEFINE_string("job_name", "", "One of 'ps', 'worker'")
8 tf.app.flags.DEFINE_integer("task_index", 0, "Index of task within the job")
9
10 FLAGS = tf.app.flags.FLAGS
11
12
13 def main(_):
14     tf.reset_default_graph()
15     ps_hosts = FLAGS.ps_hosts.split(",")
16     worker_hosts = FLAGS.worker_hosts.split(",")
17
18     # Create a cluster from the parameter server and worker hosts.
19     cluster = tf.train.ClusterSpec({"ps": ps_hosts, "worker": worker_hosts})
20
21     # Create and start a server for the local task.
```

平台资源共享

- 支持用户间容器镜像、数据集、代码等多种内容的分享和订阅
- 分享权限控制，支持共享给指定用户
- 海量行业数据集集成

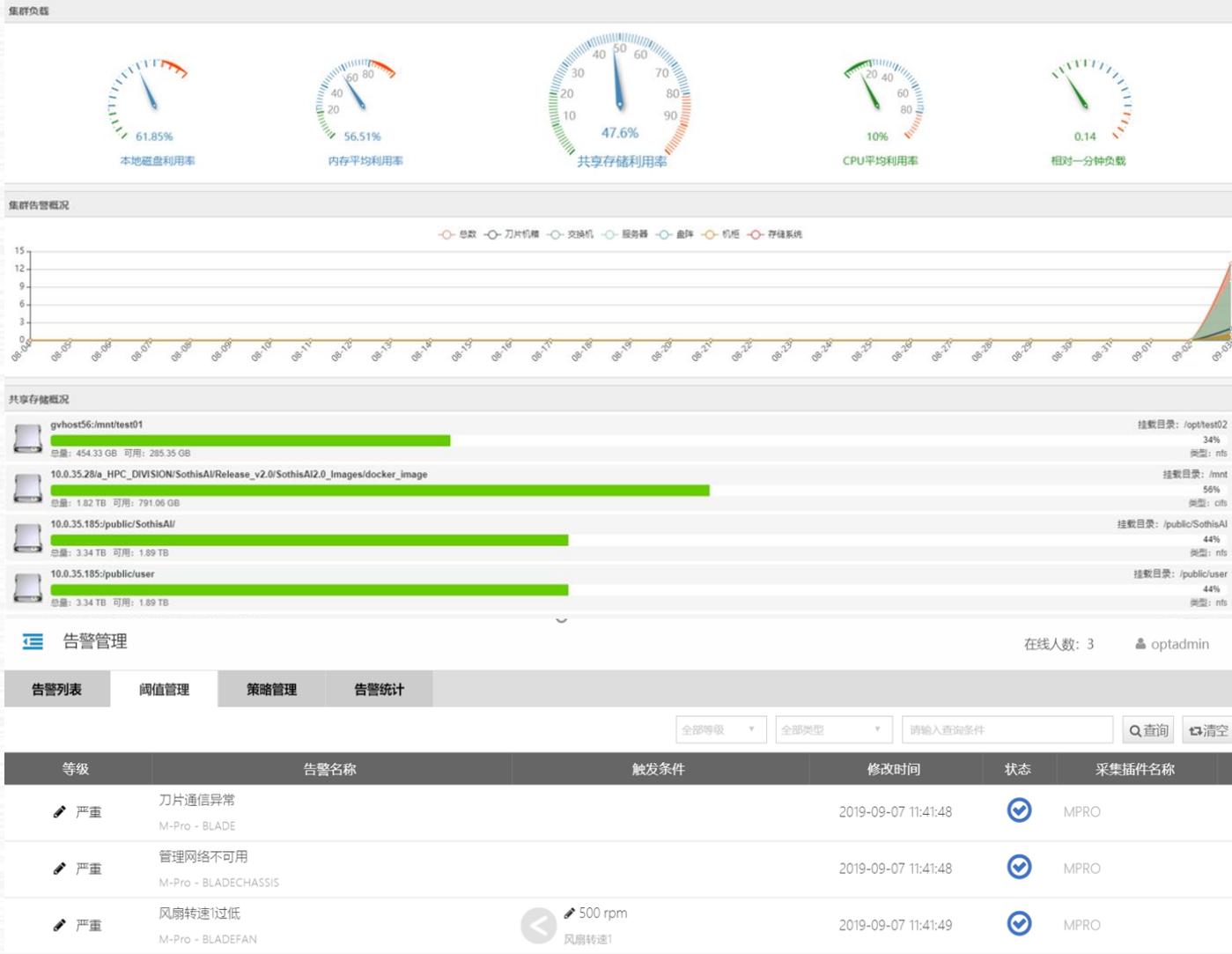
数据 代码 其他

+ 添加共享 取消共享 选择共享者

<p>LITS 共享者: chacky 日期: 20-05-19 13:27</p> <p>描述信息 Liver segmentation LiTS</p>	<p>Youku_VESR 共享者: wangxl 日期: 20-05-06 14:34</p> <p>描述信息 优酷视频处理数据集 本数据集包括视频数据、评测程序和数据说明...</p>	<p>Product_Desc 共享者: wangxl 日期: 20-05-06 14:31</p> <p>描述信息 淘宝商品描述数据集 阿里平台有很多达人、商家等撰写的文案。这...</p>	<p>Euler_Graph 共享者: wangxl 日期: 20-05-06 14:27</p> <p>描述信息 Euler图学习开源数据集 本数据集根据2018年10月23...</p>
<p>pytorch_inference_datase t_scripts 共享者: zhangyu 日期: 19-11-30 00:20</p> <p>描述信息 pytorch推理所需数据集及脚本</p>	<p>pytorch_train_dataset_sc ripts 共享者: zhangyu 日期: 19-11-30 00:20</p> <p>描述信息 pytorch训练所需数据集及脚本</p>	<p>tf_semantic_segmentation 共享者: zhangyu 日期: 19-11-30 00:20</p> <p>描述信息 TensorFlow语义分割所需模型及标签</p>	<p>CIFAR 共享者: admin 日期: 19-11-30 00:20</p> <p>描述信息 CIFAR-10和CIFAR-100是带有标签的数据集，都出...</p>

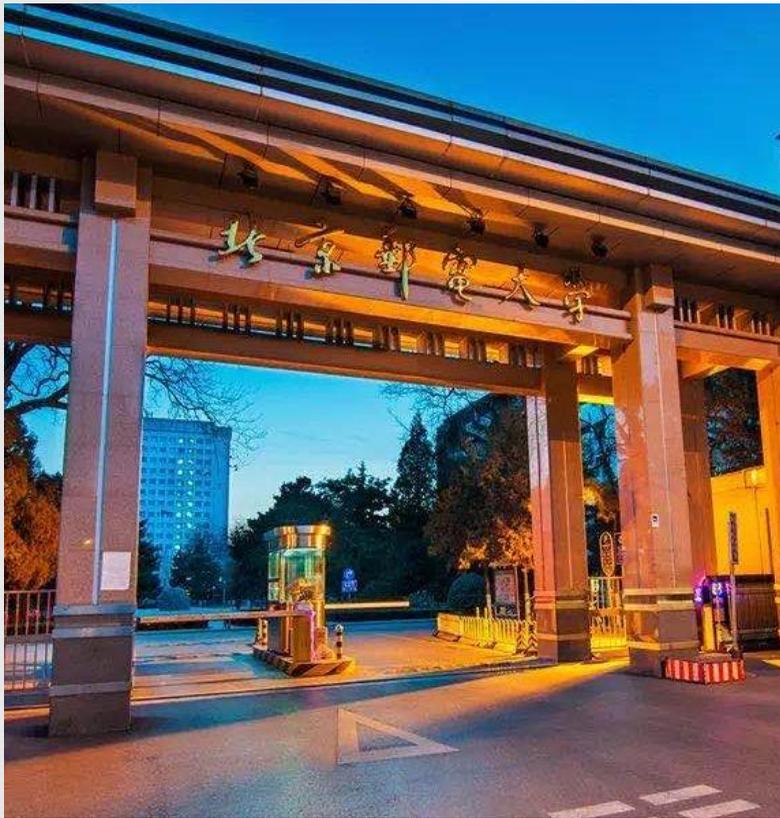
集群资源管理

- 提供集群资源监控、集群资产管理、用户管理、日志审计、告警管理等多种集群管理功能



典型案例

北京邮电大学校园安防系统



基本情况

- 2019年，北京邮电大学着手落地科研用校园智能安防系统项目，通过构建人工智能科研基础平台推动校园安防体系建设。
- 由于高等学校建筑物多、人员密集复杂、环境功能主体繁杂，通过构建人工智能科研基础平台推动校园安防体系建设，形成以校园安防可视化综合平台为基础的，实时监控视频存储、出入口人脸抓拍识别、校园异常事件检测报警等功能；

解决方案

- 该系统由曙光中标并独立构建，基于曙光天阔系列X745-G30 GPU服务器，搭配曙光人工智能管理平台SothisAI，实现了整体系统的高效性、稳定性和易用性。

客户收益

- SothisAI极大的降低了研发人员切入深度学习领域的经济成本和技术成本，提供一套容器化的集群调度以及深度学习私有云服务方案。基于容器技术，动态弹性扩展，实现深度学习开发和应用环境秒级部署。
- 实现“AI全流程覆盖”，即对深度学习中的数据预处理、模型在线编写、模型训练、超参调优、模型验证和模型发布等流程的完全覆盖。
- 基于曙光先进计算的解决方案，有效的加速了目标监控，人脸识别，车牌检索等校园安防项目的研发进度，为高校科研、教学工作探索提供了新思路、新途径、新手段。

山东大学齐鲁软件学院AI研发平台



基本情况

- 山大齐鲁软件学院建院以来，为我国经济结构战略性调整和软件产业发展急需的人才培养做出了重要贡献。学院拥有软件工程一级学科博士学位授权点和软件工程博士后流动站。目前设立人机交互与虚拟现实、软件与数据工程、可视媒体智能处理、信息安全、人工智能等五大学科群。

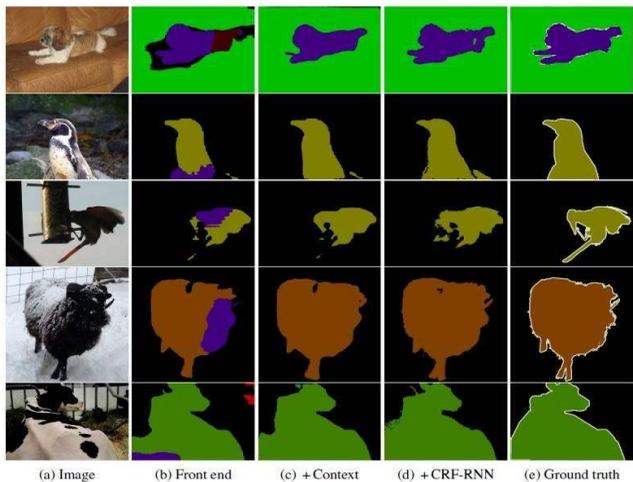
解决方案

- 中科曙光于2019年为齐鲁软件学院构建人工智能研发平台，面向学院内部的科学研究及教学工作。
- 平台基于曙光天阔系列GPU服务器及容器化人工智能管理平台SothisAI构建。

客户收益

- 该平台面向不同应用行业提供了丰富的典型数据集支持。
- 集成深度学习领域主流的Caffe、TensorFlow、PyTorch等框架，提供GUI、SSH、Jupyter等多种接入方式，加速不同开发背景的学生和老师的上手速度。
- 推动交叉学科研究与深度融合、催生新兴学科建设与现代科学技术在高校的普及，助力基础科研工作者追踪国际科技前沿，为高校综合科研实力和教学水平提升发挥了重要作用。

中科院动物所人工智能开发平台



基本情况

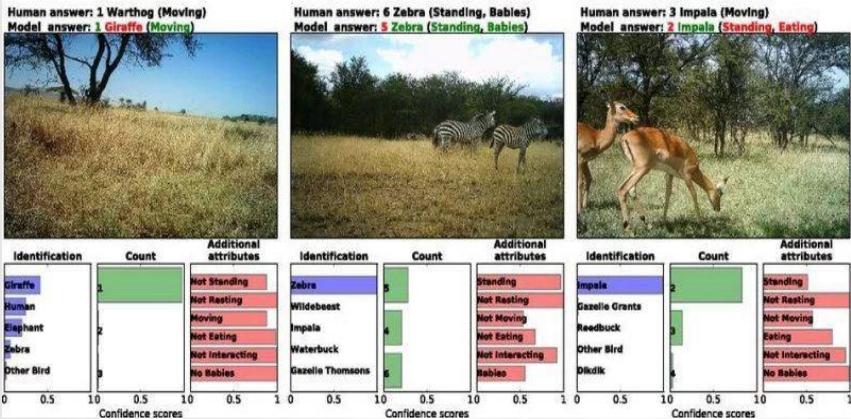
- 中科院动物研究所以野生动物和模式动物为研究对象，开展现代动物学研究，服务于人口健康、农业和生物多样性保护等国家重大需求。在细胞编程与重编程的机制、生殖与发育调控、生物灾害爆发机制与控制、物种濒危机制与保护等领域发挥引领作用；在动物分类与进化、农业虫鼠害防控和濒危动物保护中发挥不可替代的作用。
- 基于技术融合思想，需要构建人工智能研发平台，将图像分类、目标检测等深度神经网络应用思路与现有研究方向结合。

解决方案

- 采用搭载V100 GPU的高密度服务器，搭配曙光人工智能管理平台SothisAI的解决方案。
- 通过进一步合作，协助用户构建生物领域AI算法的在线服务功能，并探讨数据和算法进驻先进计算平台的可行性。

客户收益

- 基于容器的环境快速部署解决了用户开发框架多样、计算环境复杂的问题，将有限的硬件资源高效便捷的分发到开发人员手中。
- 图形化的模型训练及监控界面极大的降低了深度学习框架的学习成本，降低了跨行业AI应用开发的技术门槛。动物所有效的加速了传统学科与人工智能技术的融合进度，促进了基础科学研究的发展。



案例与合作



携手成就梦想



谢谢!

