A nighttime city skyline, likely New York City, featuring prominent skyscrapers like the Empire State Building. The scene is overlaid with numerous vertical lines of light in blue, purple, and white, each topped with a small glowing sphere, creating a digital or data-themed aesthetic.

华为云盘古大模型 AI工业化开发新模式

作坊式的AI开发无法满足行业高效高质量数字化转型需求

AI 走进**企业核心生产系统**，开始创造更大价值

“2025年，企业对AI的采用率 **86%**（当前**4%**）”

EI AI实践: **600+**项目, **30%** 进入生产系统, 但是

AI进入千行百业面临**诸多挑战**

- 1 **开发门槛高**, 需要专业人员持续支持
- 2 **场景碎片化**, 作坊式开发难以规模复制
- 3 **行业知识与AI技术**结合困难, 性能不达标
- 4 行业对普通AI模型**攻击、隐私、安全**担忧

投入人力多, 周期长, 性能很难做到极致



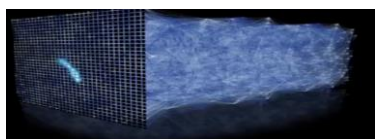
华为云EI的AI类产品设计理念:

- EI提供**基于行业的场景化解决方案**, 同时配套简化AI应用开发、落地的资产或工具。
 - 提供平台+生态类产品
 1. **开放**- 引入大量生态伙伴
 2. **AI Gallery** – 开放订阅, 吸引伙伴合作
- 关键资产 (**大模型**) + 工具 (**流水线**) + 简化AI APP

华为盘古大模型推动人工智能开发从“作坊式”到“工厂式”升级

工厂式开发

一次预训练+下游任务微调



盘古大模型 (L0)
「AI工业化开发」

行业预训练大模型
(L1)

泛化
复制

L2: 推理模型



预训练

微调&部署

迭代

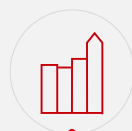
预训练大模型：百亿参数，十亿数据



参数数量
传统小模型vs.大模型

百万-千万参数

100亿参数，
第一梯队模型参数量：
Google 150亿，微软 30亿



小样本学习能力
(ImageNet 10%)
传统小模型vs. 大模型

ImageNet 10%
分类 精度**58%**

分类，业界领先 **81%**

微调部署：更高，更强，更便捷

- **更高精度**：20+电力视觉检测任务，相比传统小模型平均精度提升**5%~10%**
- **更强泛化能力**：适用**80%**以上分类、检测、分割、异常检测等AI任务
- **更便捷微调策略**：少量**1%-5%** 参数微调, 达到算法专家水准

优秀的泛化能力

充分挖掘数据内在联系，达到更高的泛化性能，对不同场景鲁棒性更强

高效样本筛选能力

海量无标注样本筛选，节省80%以上人力标注代价

小样本/零样本能力

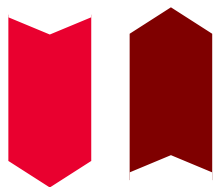
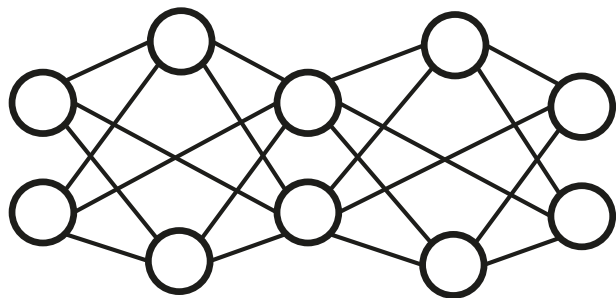
自动化数据增强，数据挖掘算法，实现零训练样本下缺陷样本识别，较传统方法发现率提升2~3倍

低门槛AI开发

提供自动化工作流，自动数据处理，自动化调参，自动化生成模型，减少对AI开发工程师专业依赖

盘古CV工具链

盘古大模型



知识积累

下游定制化场景



工具链核心目标

- 落地：提供将盘古预训练模型落地到定制化场景的能力
- 资产管理：对盘古团队各种模型（结构、行业）进行资产管理，并根据用户需求自动选型
- 积累：利用用户的使用记录，优化选型策略和调优策略

主要涉及的技术点和功能

- 超网训练+模型抽取：大模型抽取满足用户需求小模型的能力
- 模型蒸馏：大模型蒸馏超网训练
- 自动学习：盘古团队为每款模型积累了大量调优策略
- 知识积累：调优策略池根据用户使用，不断优化Ranking

盘古大模型三大独特优势

学术大模型—根技术做深 学术领先梯队

降低资源消耗

- 采用压缩学习+数据蒸馏来提速训练过程，降低训练成本50%以上

消除边际效应

- 在大模型规模保持领先的同时，保证性能随模型规模线性增长。

创新架构

- 基于CNN和Transformer联合架构, 兼具**局部特征和全局注意力**的优势

逻辑推理能力

- 构建大规模推理数据，探索显式推理与隐式推理的不同技术路线，引入知识图谱。

全栈AI开发流水线



独有先进的分层架构
自身工业知识沉淀优势，商用进展最快



盘古大模型的基石

能源

大金空调空气质量
检测准确率
提升15%

零售

米琪蛋糕无人门店
结账成功率
增加8%

金融

广发证券异常财务
检测精准
提升20%

工业

铁路故障检测
故障误报率
减少24.7%

医疗

尘肺检测
病例识别准确率
提升22%

环境

青海三江源野生动
物保护无需定制
识别100+类别动物

物流

浦发银行浦慧云仓
单个模型处理
100+仓库

科学计算

实时预测海浪高度
预测速度提升
提升10000倍



10+
业界挑战赛冠军

- nuScenes-3D检测竞赛冠军
- MS-COCO检测竞赛冠军
- WebVision分类竞赛冠军
- LVIS检测竞赛最佳创新奖
- ImageNet小模型分类领跑



40+
申请专利



IEEE ICME 2019 ACM CIKM 2018
最佳论文



70+
IEEE/ACM期刊论文

- BPM (1324次)
- CenterNet (1003次)
- PT-GAN (942次)
- P-DARTS (371次)
- PC-DARTS (361次)

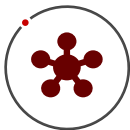


130+
CCF A类论文

- CVPR 2022: 11篇
- CVPR/ICCV 2021: 20篇
- CVPR/ECCV 2020: 48篇
- CVPR/ICCV 2019: 48篇

1

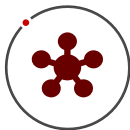
创新团队



- 国家级领军人物
- 3+华为天才少年
- 20+博士
- 30+工程师
- 50+高校实习生

1

共同目标

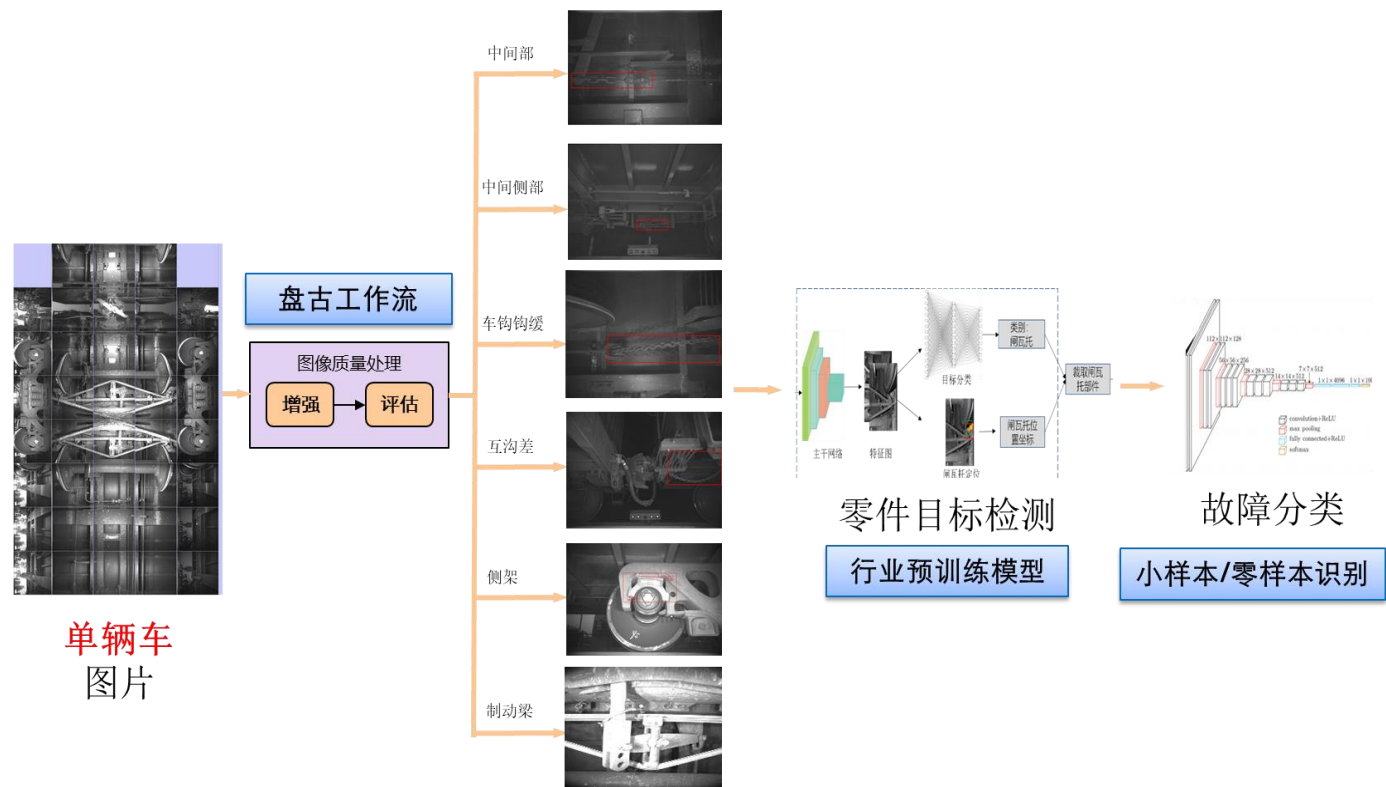


实现工业化AI开发新模式

降低了AI使用门槛， 践行普惠AI

盘古视觉大模型：铁路XXX故障识别

基于盘古行业预训练模型的智能故障识别



挑战:

- 部分图片拍摄质量差
- 故障类型多：上百种故障
- 样本分布不均衡：部分故障类型稀缺 (<10)
- 未知故障预测，要求高召回

平均召回率 **25%↑**

异常检测漏报 **80%↓**

现场实测情况:

郑州北车辆段5T检测车间集中作业分析的14条线路，在2021年9月19日至2021年10月20日期间，由5T检测车间动态检车员预报并经组长确认为提报故障的数据样本（故障图片）共计32007张。

	准确预测	漏报	故障总数	识别率
拦停	119	1	120	99.17%
重点	28280	506	28786	98.24%
一般	3084	17	3101	99.45%

盘古视觉大模型：XXX电力巡检



样本筛选效率



30倍 

识别精度



18.4% 

模型通用性



20+个  1个

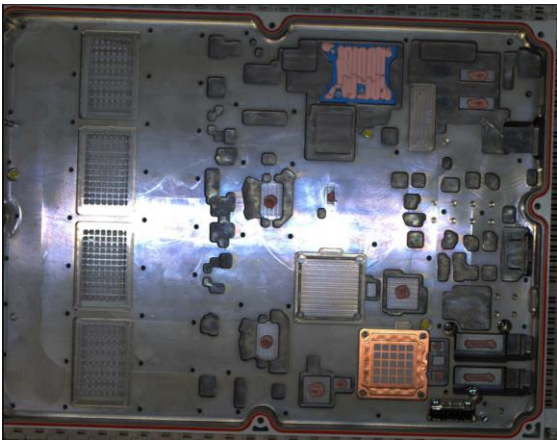
开发效率



10倍 

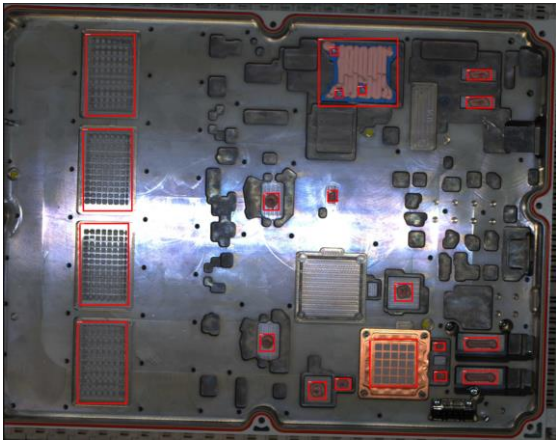
盘古视觉大模型：XXX制造部场景

基于盘古工作流的端到端自动化解决方案



盘古预训练超网

模型工作流
+
算法工作流



样本输入

- 全新类别
- 极少量样本

部件输出

- 满足超高的召回、精度要求
- 满足模型耗时、大小要求

自动化抽取任意耗时
需求模型<
自动化生成算法
解决方案



低成本+高效+可靠<

同样样本量14.5%↑ 1/2 样本量，优势明显

方法	训练集规模	Recall	Precision
Baseline	40	99%	~80%
盘古Flow	40	99.9%	99.5%
	30	99.5%	99.1%
	20	98.1%	97.8%
	10	94.7%	96.5%

盘古大模型：煤矿场景



开发标准化、可视化，形成 workflow

掘进 workflow

数据标注 → 模型训练 → 模型评估 → 结果反馈 → 模型部署

主运 workflow

数据标注 → 模型训练 → 模型评估 → 结果反馈 → 模型部署

综采 workflow

数据标注 → 模型训练 → ... → 结果反馈 → 模型部署

人员安全 workflow

数据标注 → 模型训练 → ... → 结果反馈 → 模型部署

基于 workflow，场景化 AI 模型开发 (L2) 周期短，流程规范，单模型开发周由 **月级** → **天级**

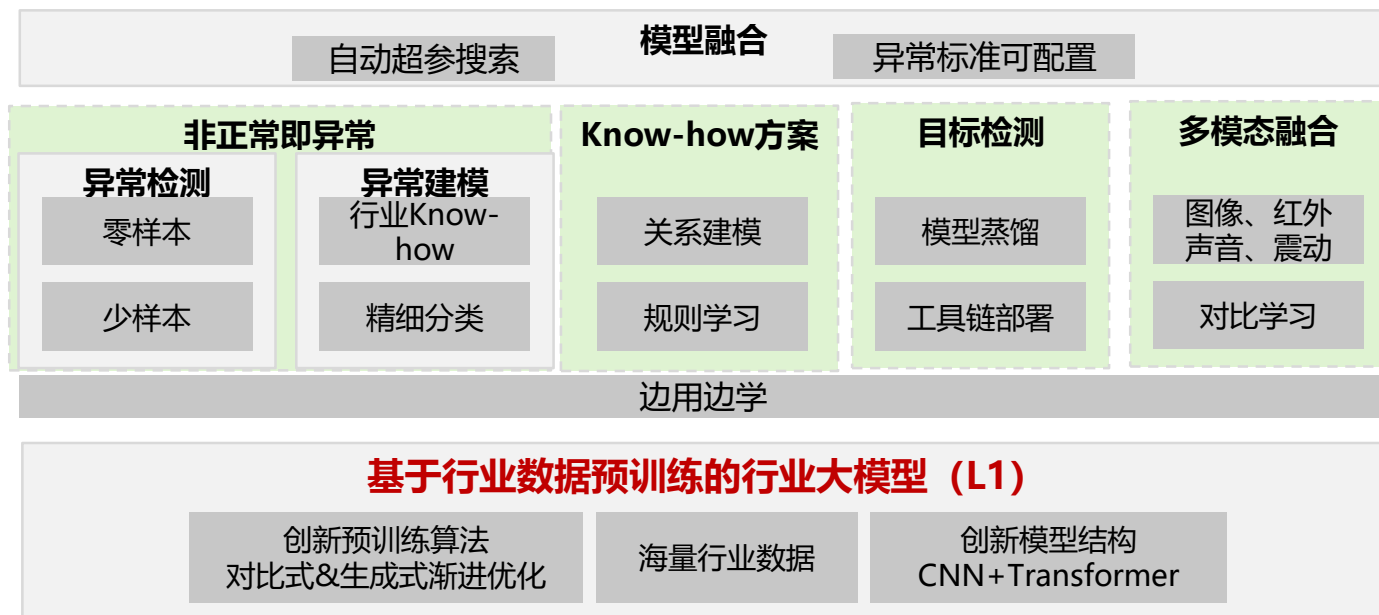
规范性强，统一的数据管理，统一的参数配置

交互性强，**可视化的开发**过程，直观易懂

AI 专家：不同场景基于需求对基础 AI 能力进行算法选择、模型参数调优等，开发出场景化的 AI workflow（支持柔性制造的流水线）

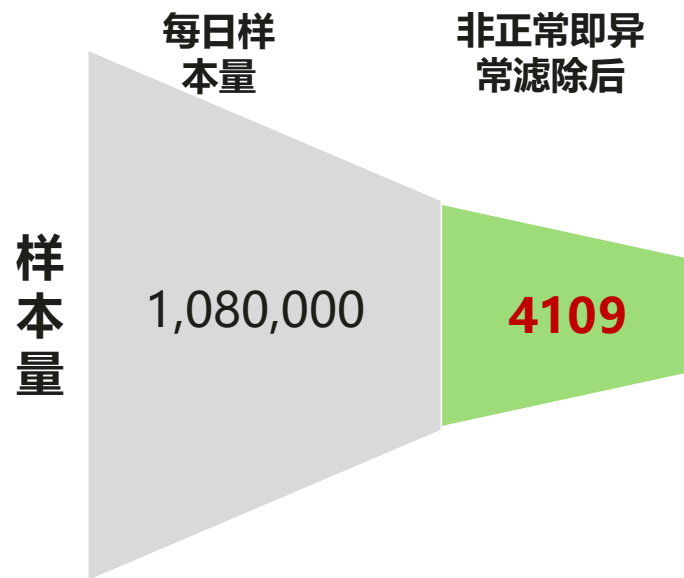
盘古大模型：基于行业数据的预训练，实现异常样本识别率大幅提升

现状：正常样本滤除率已达到99.6%



- 行业数据预训练作为底座，提供兼顾行业knowledge的基础网络模型
- 异常检测、异常建模、目标检测、行业know-how方案、多模态融合等根据场景进行多方案冗余设计，提供性能最佳的综合解决方案
- 目标检测：采用3亿量级盘古目标检测模型，进一步提升样本筛选效率

提升负样本筛选效率200倍以上



- 减少99.62%的负样本标注工作量
- 减少的负样本标注工作量 = 过滤掉的正常样本数/样本总量： $(1080000 - 4109) / 1080000 = 99.62\%$

注：
负样本标注工作量 = 负样本数量 * 单个负样本标注工作量

盘古大模型：良好的泛化性及高效性，提升算法通用性，实现AI快速应用

新场景高效适配：大模型的算法准确率提升15%左右

- 堆煤场景中，分别使用26，78，104，130个训练样本验证。
大模型全方位超过业界小模型，准确率普遍提升15%左右

通用性强：相似场景下，少量数据即可达到更高的准确率

- 在某矿上训练得到的皮带模型加入少量山东能源数据进行微调训练，相比业界小模型性能显著提升

模型	召回率	训练集数量	准确率	综合评定
小模型 (业界)	68%	26	65%	67%
	73%	78	66%	69%
	73%	104	66%	69%
大模型	68%	26	80% (+15%)	74%
	73%	78	79% (+13%)	76%
	73%	104	85% (+19%)	78%

模型	召回率	新增数量	准确率	综合评定
小模型 (业界) 原场景	79%	/	52%	63%
小模型 (业界) 新场景		28	74%	76%
大模型原 场景		/	78%	78.5%
大模型 新场景		28	85% (+11%)	82%

大模型的海量参数以及多种下游微调解决方案，支撑少样本数据的高效训练，以及对于同类场景下强适应性

依托AITISA产业联盟，联合4伙伴成立大模型产业推进组



2021：大模型产业推进组
华为、鹏城实验室、中国科技大学、上海交通大学、厦门大学
2021全球智博会产品金奖
首期大模型专题技术沙龙
盘古大模型产业发展联合专项工作组专题会议研讨

2022：大模型产业推进组
大模型技术沙龙
提升产业影响力
大模型白皮书
创新研究专题沙龙
高校老师罗马广场

2023：大模型产业推进组
大模型行业场景
借助联盟推进场景复制
金融（智能投研）
工业质检、农业
气象、金融

大模型



开场致辞

新一代人工智能产业技术创新战略联盟联合秘书长
张伟民



张伟民
新一代人工智能产业技术创新战略联盟联合秘书长

开场致辞

华为云人工智能领域首席科学家 田奇



田奇
华为云人工智能领域首席科学家

盘古大模型的研究进展和对产业生态构建的思考

华为云EI高级研究员，大模型推进组组长
谢凌曦博士



谢凌曦
华为云EI高级研究员，大模型推进组组长

大模型前沿进展与落地压缩

厦门大学信息学院教授
纪荣峰



纪荣峰
厦门大学信息学院教授

鹏城系列大模型基础研究及应用

鹏城实验室开源软件与平台研究所研究员
王晖



王晖
鹏城实验室开源软件与平台研究所研究员

推荐性国家标准项目建议书			
中文名称	信息技术 神经网络表示与模型压缩 第2部分：大规模预训练模型		
英文名称	Information Technology -- Neural Network Representation and Model Compression -- Part 2: Large Scale Pre-training Model		
制定/修订	<input checked="" type="checkbox"/> 制定 <input type="checkbox"/> 修订	被修订标准号	
采用国际标准	<input type="checkbox"/> 无 <input type="checkbox"/> ISO <input type="checkbox"/> IEC <input type="checkbox"/> ITU <input type="checkbox"/> ISO/IEC <input type="checkbox"/> ISO 确认的标准	采用程度	<input type="checkbox"/> 等同 <input type="checkbox"/> 修改 <input type="checkbox"/> 非等效
采标号		采标名称	
标准类别	<input type="checkbox"/> 安全 <input type="checkbox"/> 卫生 <input type="checkbox"/> 环保 <input checked="" type="checkbox"/> 基础 <input type="checkbox"/> 方法 <input type="checkbox"/> 管理 <input type="checkbox"/> 产品 <input type="checkbox"/> 其他		
ICS	35.040		
上报单位	国家标准化管理委员会		
技术归口单位 (或技术委员会)	全国信息技术标准化技术委员会		
主管部门	国家标准化管理委员会		



总结：盘古大模型解决传统AI开发的难题



谢谢！

