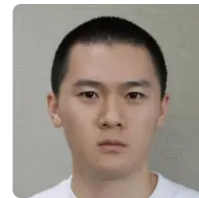


# 胡天



男 | 年龄: 32岁 | 18664779545 | hutian2013@qq.com

9年工作经验 | 求职意向: 深度学习 | 期望薪资: 25-30K | 期望城市: 深圳

## 个人优势

9年C/C++开发经验, 7年深度学习模型开发经验, 勤于思考和探索。熟练掌握 PyTorch、TensorFlow 等深度学习开发框架, 熟悉各类深度学习网络的结构和算子的实现原理, 并且在训练和推理模型的调试和调优方面有着实际的操作经验, 并在实际工作中应用来提高深度学习推理的效率和性能。熟练掌握深度学习算法的 NPU 和边缘端的部署。熟练掌握大模型的微调、训练和部署。熟悉 RAG大模型应用开发。

## 工作经历

广州傲程软件技术有限公司深圳分公司 算法工程师 2024.10-2025.06

- 负责智能核电计算中心相关算法的优化改造
- 负责智能核电知识中心 langchain 工单分析开发
- 与核电研究人员合作实现算法代码

趣动未来(深圳)科技有限公司 算法工程师 2024.04-2024.09

- MMPose人体检测, 人体关键点模型的训练。
- 人脸识别模型训练, 人脸特征入库和特征比对算法。
- 人体动作库构建以及人体动作识别算法开发。
- 模型算法在Android平台的RKNN部署。
- 调研利用ControlNet、IP-Adapter实现风格迁移的方法, 用于运动结束后生成分享图片。

深圳市视美泰技术股份有限公司 算法工程师 2019.03-2024.04

- 使用 PyTorch、TensorFlow、MxNet、PaddlePaddle 等框架开发人脸识别、商品识别、检测算法、分类回归等算法, 通过对卷积层、激活函数、损失函数等算子的优化提升模型的准确率和推理速度。
- 使用 cuDNN、TensorRT、OnnxRuntime、RKNN、NCNN 等深度学习推理框架进行模型部署, 部署平台包括 NVIDIA Jetson、rockchip RK3568、amlogic A311D、x86/ARM cpu等。
- 参与 Transformer 大模型算法开发, 成功使用 deepspeed、flash attention、vllm 等技术进行分布式模型加速训练和推理, 并对大模型进行预训练和 LoRA 微调, 对 prompt 的开发和优化使大模型应对不同类型的任务。
- 使用C/C++和 Python 语言, 开发深度学习模型应用程序。

北京思特奇信息技术股份有限公司 C/C++ 2017.02-2019.01

- 中国联通深港用户计费结算系统需求的开发和迭代。
- 新老系统割接过程中的数据迁移、容灾、回滚。
- 用户数据分析和报表生成脚本。

## 项目经历

负责项目中工单故障等级判断

- 1、大模型使用deepseek-32b
- 2、使用langchain框架对历史工单、领域知识、专有名词进行检索
- 3、构建prompt模板，对检索出的信息构建prompt
- 4、指定大模型输出固定格式的json信息用于下一流程，包括思考过程、判断理由、判断结果

时序数据异常状态检测

算法工程师

2025.01-2025.04

有以下几种类型：

- 1、从数据中心获取传感器实时数据，动态确定数值突变的起点和终点，与已知的异常类型做比对，确定此突变的类型。
- 2、用传感器历史数据做KNN分类训练，确定一段序列的异常类型。
- 3、用传统的规则判断对传感器数据进行异常分类。

滤网破损检测

算法工程师

2024.10-2025.03

核电领域某设备的滤网网面破损检测，使用yolov8训练检测模型，完成的工作包括数据标注清洗、训练、云化部署。

AEKE力量镜A1

算法工程师

2024.04-2024.09

- 1. 负责人体动作识别算法的开发，对MMPose进行再训练，提高人体关键点检测的精度。通过json配置人体各个部位的形态要求实现对动作的识别以及计数或者计时，比如监测交叉式运动的次数和时间，伸展运动的保持时间等。
- 2. 负责完成设备的人脸识别登录功能，对开源数据集进行清洗，用resnet50训练得到人脸特征模型，精度超越insightface。并完成人脸特征入库，人脸特征比对等算法逻辑。
- 3. 负责将模型和算法部署到Android RKNN NPU平台，包括模型到RKNN的编译，JNI编译以及Java接口。

LLM大模型应用开发

算法工程师

2023.02-2024.04

调研语言大模型的模型结构，训练方法，实现大模型的训练，LoRA 微调以及部署，结合公司业务探索大模型的应用。

项目职责：

- 1. vicuna、chatGLM2、qwen等 transformer 模型全参训练和微调。
- 2. 利用 prompt 提示词、qwen7b开源模型、qwen7b LoRA 微调模型、langchain 等技术实现了基于特定知识库的智能助理数字人应用程序。

基于边缘、端侧的算法模型部署

算法工程师

2020.03-2024.04

将深度学习算法模型部署到不同的平台，包括 NCNN 、瑞芯微 RKNN、TensorRT、amlogic 等平台，并封装成统一的接口，根据各个平台对不同算子的支持程度，优化原始网络结构和对算子的优化方法，减少算子在不同平台的精度损失。

项目职责：

- 1. 不同推理平台模型移植和算子优化，模型转换脚本的开发。
- 2. 不同目标平台下的C++推理接口代码封装。

AI电子称

算法工程师

2021.08-2022.12

内容：

AI电子秤解决方案可准确快速识别蔬菜、水果、干货、零食等，将商品放置于秤盘上，即可自动对物品进行识别，摄像头捕捉到图像后，就能迅速识别出物品，配合系统完成自动称重。

项目职责：

- 1. 数据的采集、清洗和模型训练。

- 2. AI 电子称应用程序开发，包括实现特征提取，特征入库，特征搜索，识别策略等功能。
- 3. 模型的 NCNN 和边缘端的 NPU 移植。

业绩：

专利：

CN202211413255.1 物品称重自动识别方法、装置、计算机设备及介质

人脸识别算法模型研发      算法工程师      2019.03-2021.08

内容：

人脸识别相关算法研发，包括人脸检测、人脸追踪、活体检测、人脸识别、人脸特征库管理等。

项目职责：

- 1. 模型精度和速度的持续优化，人脸特征库搜索算法。
- 2. 各功能模块的 API 封装。

业绩：

专利：

CN202310137061.1 一种人脸数据库快速搜索方法及系统

CN202011003125.1 基于人脸识别更新人脸特征的方法、装置和计算机设备

CN202010797552.5 人脸偏转角度的确定方法、装置、计算机设备及介质

分布式内存数据库系统      C++工程师      2017.05-2018.10

- 负责内存数据库生产站点和容灾站点之间数据复制和数据持久化功能的实现和完善以及内存数据库批量操作工具脚本编写。
- 采用 TCP 协议对生产内存库站点的增删改操作复制备份到容灾站点。内存库内容的每次变动都产生一条 log 信息，将 log 信息放入 FIFO-queue 供后续处理。
- 实现内存库的 Python API，并用 Python API 实现内存库的批量操作工具。

综合采集预处理      C++工程师      2018.03-2018.06

- 负责新增处理香港和记运营商网元请求以及落单规则。
- 解耦和重构多种请求相关代码，新增预处理配置文件和相关规则的回调函数处理，传入下一级。

新一代BSS云化系统割接      C++工程师      2017.08-2018.01

- 负责香港联通40万用户数据的迁移，计费相关程序的云化接口改造以及云化系统告警平台的构建
- 40万香港联通用户数据的近100张表的数亿条记录数据从 Oracle 迁移到 DMDB 分布式内存库
- 将 Oracle 接口改成 DMDB 分布式内存库接口，文件I/O改成 HDFS 接口，新增 Zookeeper 程序调度
- 告警平台：Python 编写主框架，监控项目可配置，配置项为监控项目、执行脚本路径，主框执行脚本并接收固定格式的返回信息构建告警信息，若有异常则调用短信平台接口发送告警信息到相关人员。监控项目比如进程状态、CPU 使用率、内存使用量、每日处理数据量，错单量等。为系统割接后的正常运行提供了重要保障

## 教育经历

东华理工大学      本科      电子科学与技术      2012-2016