

段京辰

Info: 27 岁 男 中共党员

作品集&个人网站: www.djcom.top

Phone: 183-9826-0691

邮箱: jcduan1999@163.com



教育背景

2023.09 - 2026.06 (硕士研究生) 西南石油大学 (双一流) 计算机技术 绩点: 3.56

核心课程: 机器学习、深度学习、最优化理论与方法、高级算法分析、面向对象方法。

2016.09 - 2020.06 (本科) 四川农业大学 (211) 物联网工程 绩点: 3.57

核心课程: C/C++、java、数据库、计算机网络、传感器原理与应用、嵌入式系统开发与应用。

技能证书: 英语六级 (CET-6)、计算机网络高级工程师。

实习经历

西门子 (中国) 成都软件开发分公司 AI LAB 实习生 2024.08 - 2026.06

➤ 3D 视觉算法研发与工业软件落地 (算法+软件) 链接: djcom.top/portfolio/3d-industry-algorithm/

- 复现并改进 CVPR/SIGGRAPH 前沿点云与 Mesh 重建算法, 面向工业场景 3D 模型语义分割、分类及特征提取重构网络架构。对于点云模型设计识别算法: 吸收 GNN & PointNet++ 方案经验, 在 U-Net 框架内引入 FPS 采样与 KNN 分组构建层级化局部特征聚合机制, 设计 60% 比例随机掩码策略增强模型鲁棒性, 利用 Transformer 架构捕获全局依赖, 加入 3D 模型面-边拓扑关系的图神经网络 (GNN) 辅助位置编码。采用两阶段训练策略: 基于 ShapeNet、ABC Dataset 等进行大规模预训练, 结合特征权重下游迁移与内部小样本数据微调分割模型。该方法将 3D 模型子部件在复杂样本测试集上的识别 mIoU 由原始图论算法的 82.42% 提升至 94.75%。为后续提供 3D 样本 embedding 方案以支持 LLM 应用的相似度检索、RAG 召回、属性匹配。
- 完成 PyTorch 至 ONNX 的模型转换与 INT8 量化, 基于后台 C++ 开发 Siemens NX (CAD 软件) 插件 UI 与 API 逻辑, 打通 NX 交互界面~数据预处理~推理引擎调用~结果可视化的闭环。实现识别算法端到端部署与功能验证, 严格遵循需求分析、架构设计、单元测试及交付维护的软件工程规范。

➤ ER Assistant Web 服务质量保障体系构建 (测试) 链接: djcom.top/portfolio/er-web-test-system/

- 自动化测试与智能告警闭环: 搭建 API 自动化测试框架 pytest, 设计异常测试用例覆盖 ER Assistant Web 服务核心接口。
- 开发 Teams 自动告警: 通过 Python 调用 Microsoft Graph API, 编写自适应卡片 Adaptive Card 模板, 当测试失败时自动触发 Webhook 将报错信息推送至 Teams 频道, 并精准@责任开发人员。定期增量提取 Azure 服务器用户 ER 数据, 自动化清洗并更新 Elastic search 数据服务, 同步刷新前端页面。

中石化中原油田分公司工程技术研究院 测井解释部门实习生 2024.06 - 2024.08

➤ 勘探测井电磁信号的智能地层识别系统开发 (算法+软件) 链接: djcom.top/portfolio/logging-signal-ai-system/

- 针对油田公司研发的新型专利仪器进行相应的算法系统开发: 研发电磁测井信号处理模块, 对原始测井信号噪声干扰的问题, 设计深度条件引导的扩散去噪模型 (DCG-DDM), 平均提升信号信噪 42%。构建以 mWDN 为基础的多尺度时频形态融合特征提取框架 (MSTFF) 用于提取聚类与分类特征, 用 K-means++ 对预处理后的测井电信号进行无监督分层聚类, 识别套管接箍。对少量复杂数据增强, 用 CNN+SE 注意力结合 Transformer 编码捕获特征实现地质小层分类。该方案较人工解释效率提升数倍。
- 开发 QT 软件实现集成: 完成界面设计, 信号触发, 嵌入 ONNX 量化模型进行信号处理, 输出地质剖面可视化识别结果报告。
- 成果转化: 发表 CSCD 核心期刊论文《基于瞬变电磁的套管接箍信号智能识别方法》, 获《测井瞬变电磁信号分析软件》著作权。

项目经历

个人站点访客 AI 助手 RAG、LangGraph、Chroma、BM25 2026.07 - 2026.09

➤ 个人全栈项目 (LLM RAG 应用) 链接: djcom.top/services/#contact

- 重要成果: 搭建双层隔离知识库架构, 划分标准问答库 (JSON 结构化 QA) 与随笔日志库 (Markdown), 兼容结构化问答、个人随笔、PDF、Word、简历资源混合检索。设计多模型 Token 管控方案, 区分轻量任务与对话模型, 实现 LLM 实例懒加载单例复用, 降低重复连接开销, 独立追踪 Rewrite、检索、LLM 各阶段 Token 消耗并输出统计报告。

- **产品与架构**: 独立开发面向站点访客的智能问答系统, 支持访客问题解答、自助信息登记。基于 FastAPI 异步架构搭建后端服务, 实现混合检索、工具调用、结构化输出全链路。配套 Docker 容器化部署, 支持后续功能迭代与资源扩展。
- **RAG 与 Agent 链路**: 构建 Hybrid Search 混合检索管道, 实现向量检索+BM25 关键词双路召回。设置相似度阈值策略, 达标结果短路返回降低延迟, 未达标结果融合双路召回。设计分层 Prompt 体系 (Guard 安全边界、QueryRewrite、意图路由、Rerank), 基于 Pydantic 强制校验模型结构化 JSON 输出。落地 Tool Calling 与 Text-to-SQL 方案, 支持访客需求自动生成 SQL 写入 MySQL, 配套消息中断事务回滚机制保障数据一致性。
- **工程与交互**: 针对 PDF、Word 设计表格、图像感知的文档分块策略, 保留非文本元素上下文完整性。全链路配置规则兜底策略, 覆盖检索失效、LLM 调用异常场景, 支持自动重试与多模型降级。各模块异常隔离, 保障核心问答能力持续可用。

AI 个人智能健康管理平台

RAG、LangChain、AI Agent、FastAPI

2026.03 – 2026.04

➤ 个人全栈项目 (LLM RAG 应用)

链接: djcom.top/health

- **重要成果**: 实现 Tokon 消耗管控: 分用户等级 (VIP) 调用不同后台 LLM, 动态载入不同 prompt、会话上下文复用、标准化前置指令, 提升服务端 KV Cache 复用率, 减小 token 消耗。对 RAG 知识库按用户权限进行管理, 在向量分片元数据中增加权限标识字段, 于 Agent 编排层完成检索前置过滤, 实现知识库资源逻辑隔离, 阻止越权访问。完整的上下文处理策略, 会话在服务器持久保存, 短期记忆摘要压缩。
- **产品与架构**: 本人独自设计的一款面向个人用户的综合性健康管理 Web 应用, 支持运动记录、身体成分分析。嵌入独立的 AI 智能问诊、运动地点建议、饮食规划。实现了从数据采集、可视化分析到 AI 辅助决策的闭环, 基于 FastAPI+Vue 3 实现前后端分离, 并采用 Docker 容器化技术在阿里云 ECS 上完成了自动化部署与持久化运维。
- **RAG 与 Agent 链路**: 基于 LangChain 封装 qwen 系列模型与 Agent 工具调度, 接收用户需求参数, 利用 Chroma 向量知识库与 MySQL 个人健康数据实现双路召回, 融合检索结果构建 Prompt, 结合 Few-Shot 与 CoT 优化口语化健康 Query, 设计 Query Rewrite 到 Embedding 到噪声过滤到 Cross-Encoder Rerank 的 RAG 管道, 在本地测试数据上健康问答 Top-3 检索命中率由基线的 72.3%提升到 92.6%。实现 Rewrite、Embedding、Rerank、LLM 四阶段 Token 消耗独立追踪与终端输出报告。
- **工程与交互**: 集成高德 Web 与 JS API, 实现 LBS 定位, 周边健身、泳池、理疗场所检索及 Vue 3 实时地图跳转。后端采用 FastAPI 异步依赖注入体系, 配合 JWT+bcrypt 密码哈希完成安全认证, 提供健康数据 CRUD、聚合统计及大模型 SDK 封装接口。

专业技能

编程语言: C++、Python、Java、Matlab。

AI 算法框架: PyTorch 框架、LangChain 框架、ChromaDB、RAG 全链路开发、3D 视觉、大模型 Agent。

技术工具: 面向对象分析与设计, QT 开发, Git 版本控制, pytest 自动化测试, Docker 部署, AI 工具使用, Linux 系统基本操作及常用命令, MySQL 数据库, ONNX 量化。

支持技能: 熟练办公软件, 能高效进行团队协助, 有良好的规范化代码编写习惯, 能独立撰写技术手册、操作手册及项目交付文档。

校园经历

计算机学院文体与宣传委员

西南石油大学

2023.09 - 2024.09

负责专业文体活动的宣达, 包括运动会、文艺汇演、主题班会等、制定班级宣传方案, 涵盖活动预告、成果展示、公众号展示等。

学院文体部部长与小班体委

四川农业大学

2016.09 - 2020.06

负责小班体育活动的开展与参与, 以身作则, 多次带领同学获得如校运会, 趣味运动会的奖项。负责学院文娱活动如: 歌手大赛, 新生才艺大赛, 朗诵比赛的活动策划与后台舞台背景视频特效制作, 协作舞蹈团, 乐团的器材的使用与日常训练。

数学建模协会培训部副部长

四川农业大学

2017.09 - 2018.06

负责培训教授协会内部成员基本的数学建模软件使用、论文写作、算法模型的使用, 以及如何写程序代码, 指导协会成员参与数学建模竞赛并获奖。

个人评价

学习适应能力突出, 执行力强, 工作严谨务实。拥有扎实的算法基础、软件开发功底, 熟悉深度学习、大模型理论, 具备完整 LLM 应用开发、模型开发、部署测试及 RAG 工程落地实战经验。我先后在**中石化**与**西门子的**技术部门长期实习, 积累了工业场景下软件开发、AI 研发与项目落地经验、抗压能力强, 具备良好的团队协作精神与沟通能力, 能快速融入团队并承担核心模块的开发工作, 坚持以技术落地为导向, 运用技术手段解决真实业务痛点。

Jingchen Duan

Age: 27, Male
Phone: (86)183-9826-0691

Portfolio Website: www.djcom.top
Email: jcduan1999@163.com



Education Experience

Sept. 2023 - Jun. 2026 Southwest Petroleum University Master of Computer Technology GPA: 3.56
Core Courses: Machine Learning, Deep Learning, Advanced Algorithm Analysis, Object-Oriented Methodology

Sept. 2016 - Jun. 2020 Sichuan Agricultural University Bachelor of IoT Engineering GPA: 3.57
Core Courses: C/C++, Java, Database, Computer Networks, Sensor Principles & Applications

Certifications: CET-6, Senior Engineer of Computer Networks

Internship Experience

Siemens (China) Co., Ltd. Chengdu Software Development Branch | AI Lab Intern Aug. 2024 – Jun. 2026

➤ 3D Vision Algorithm R&D & Industrial Software Implementation (Algorithm + Software)

- Reproduced and optimized state-of-the-art **point cloud** and **Mesh** reconstruction algorithms published on **CVPR** and **SIGGRAPH**; **reconstructed network** architectures for semantic segmentation, classification and feature extraction of industrial **3D models**.
- Designed an identification algorithm for point cloud models: integrated technical merits of GNN and PointNet++, built hierarchical local feature aggregation modules via **FPS** sampling and **KNN** grouping under the **U-Net** backbone. Adopted a 60% random masking strategy to boost model robustness, leveraged Transformer to capture global dependencies, and embedded GNN with 3D model vertex-edge topological relations for auxiliary position encoding.
- Implemented a two-stage training pipeline: conducted large-scale **pre-training** on ShapeNet, ABC Dataset, then migrated feature weights and **fine-tuned** segmentation models with internal small-sample industrial data. The optimized approach lifted the **mIoU** of 3D model sub-component recognition on complex test datasets from 82.42% (original graph-based algorithm) to **94.75%**. Delivered a 3D sample **embedding** solution to support LLM downstream applications including similarity retrieval, RAG recall and attribute matching.
- Completed end-to-end model engineering: converted **PyTorch models** to **ONNX** format and applied **INT8 quantization**; developed UI and API logic for Siemens NX (CAD software) plugins with C++. Built a closed workflow covering NX interface interaction, data preprocessing, inference engine invocation and result visualization, delivering fully verified recognition functions in compliance with standardized software engineering processes (requirement analysis, architecture design, unit testing, delivery & maintenance).

➤ Construction of ER Assistant Web Service Quality Assurance System (Testing)

- Built an API **automated testing** framework based on **pytest**, designed comprehensive abnormal test cases to cover core interfaces of the ER Assistant Web service.
- Developed automatic alert workflow for Microsoft Teams: invoked Microsoft Graph API and **Webhook** via Python and customized Adaptive Card templates to push error logs to designated Teams channels with targeted @mentions of responsible developers upon test failures.
- Regularly extracted incremental ER user data from Azure servers, automated data cleaning and synchronized updates to **Elasticsearch** to refresh front-end display data.

Sinopec Zhongyuan Oilfield Company | Logging Interpretation Department Jun. 2024 – Aug. 2024

➤ Intelligent Stratum Identification System for Exploratory Logging Electromagnetic Signals (Algorithm + Software)

- Developed algorithm modules: proposed a Deep Condition-Guided Diffusion Denoising Model (**DCG-DDM**) to suppress noise in raw logging signals, improving the average signal-to-noise ratio by **42%**.
- Constructed a Multi-Scale Time-Frequency Morphology Fusion feature extraction framework (**MSTFF**) based on

mWDN; adopted **K-means++** to perform unsupervised hierarchical **clustering** on preprocessed electromagnetic logging signals for casing collar detection.

- Applied data augmentation on limited complex samples; combined CNN with SE attention mechanism and **Transformer** encoder to extract discriminative features for geological sublayer **classification**, drastically boosting interpretation efficiency compared with manual work.
- Developed an integrated **QT** desktop application: implemented graphical interface design and signal trigger logic, embedded **ONNX quantized models** for real-time signal processing, and generated visualized geological cross-section identification reports.
- Achieved academic and intellectual property output: published a **CSCD** core journal paper *Intelligent Identification Method of Casing Collar Signals Based on Transient Electromagnetism*, and obtained **software copyright** for *Transient Electromagnetic Logging Signal Analysis Software*.

Project Experience

➤ AI Personal Intelligent Health Management Platform | Tech Stack: RAG + LangChain + AI Agent

➤ Independent Full-Stack Personal LLM Application Project (LLM)

- **Product & Architecture Design:** Independently built a full-featured Web application for personal health management, supporting exercise tracking, body composition analysis, **AI medical consultation**, fitness location recommendation and personalized diet planning. Implemented a closed loop from raw data collection, visual analytics to AI-assisted decision-making. Adopted front-end and back-end separation architecture with **FastAPI + Vue 3**, containerized the project via **Docker** and completed automated deployment & persistent operation on Alibaba Cloud ECS.
- **RAG & Agent Pipeline Development:** Wrapped **qwen-max** LLM invocation and Agent tool scheduling logic with **LangChain**. Designed dual-path retrieval combining **Chroma** vector knowledge base and **MySQL** personal health datasets, fused multi-source retrieval results to construct optimized prompts. Integrated **Few-Shot** and **CoT Prompt engineering** to refine colloquial health queries; built a complete RAG pipeline including Query Rewrite, Embedding, Noise Filtering and Cross-Encoder Rerank. The Top-3 retrieval hit rate of health Q&A increased from 72.3% (baseline) to **92.6%** on local test data. Implemented independent token consumption tracking and terminal reporting for four stages: **Rewrite**, **Embedding**, **Rerank** and **LLM** inference.
- **Engineering & Interactive Functions:** Integrated **Amap Web & JS API** to realize LBS positioning, supporting retrieval of surrounding gyms, swimming pools and physiotherapy venues with real-time Vue 3 map navigation. Built asynchronous dependency injection system on FastAPI; adopted **JWT + bcrypt** password hashing for secure authentication, and encapsulated standardized APIs for health data CRUD, statistical aggregation and LLM SDK calls.

Technical Proficiencies

Programming Languages: C++, Python, Java, Matlab

AI & Algorithm: PyTorch, LangChain, ChromaDB, RAG Pipeline, 3D Computer Vision, LLM Agent

Tools & Middleware: QT, Git, pytest, Docker, Linux & Shell Commands, MySQL, ONNX

Soft Supporting Skills: Office Suite, teamwork and communication capabilities, standardized coding habits

Self-Evaluation

Quick learner with outstanding adaptability and rigorous, results-driven execution. Possess solid foundational knowledge of algorithms and software development, proficient in deep learning and large language model theories, with end-to-end practical experience covering LLM application development, model training, deployment testing and industrial RAG engineering implementation.

Completed long-term internships in technical departments of **Sinopec** and **Siemens**, accumulating abundant industry experience in AI R&D, software engineering and technical landing. Resilient under high work pressure with strong teamwork and cross-department communication skills. Oriented toward practical business value, adept at leveraging AI and software technologies to resolve real industrial pain points.