

个人简历

姓名：严雪



职位：机器视觉算法工程师

电话：13298303895 | 邮箱：yanxue6886@163.com

作品集：<https://lynnyanx.github.io/>

工作年限：3年+

个人优势 (Personal Highlights)

- 软硬全栈闭环能力：**拥有 3 年+ 机器视觉与工业物联网实战经验，具备“**算法研发(Python) + 工程落地(C#) + 边缘部署(C++)**”的完整技术闭环能力。能够独立主导从需求分析、模型选型、训练调优到桌面端软件交付的全流程。
- 多领域算法落地：**不仅精通视觉领域的 **SOTA 模型** (PatchCore 异常检测、YOLO 目标检测)，更具备 2.5D 视觉 (**光度立体/PMD**) 与时序分析 (LSTM/DTW) 的跨领域开发经验。擅长通过模型量化及 ONNX Runtime 技术，解决工业现场小样本、高耗时的落地痛点。
- 高性能混合计算：**精通 C# / C++ / Python 异构混合编程。针对工业视觉的高吞吐量需求，熟练运用 Cython 重构热点代码，结合 OpenMP 并行计算与 Halcon 算子加速，拥有将核心算法性能提升 3-8 倍的成功实战经验。
- 企业级架构设计：**深谙 Clean Architecture (**整洁架构**) 与 DDD (**领域驱动设计**) 思想。在复杂工业软件中，熟练运用 MVVM 模式、依赖注入 (DI/IoC) 及异步并发模型 (Async/Await)，致力于构建高内聚、低耦合、易于扩展的现代化工业软件系统。

技能清单 (Technical Skills)

- 核心语言：**Python (精通), C# (.NET 8/Core), C++, Cython, LabVIEW
- 机器视觉：**2.5D 视觉 (**PMD/Photometric Stereo**), Halcon 19.11, OpenCV, 3D 重构
- 核心算法：**无监督异常检测 (PatchCore), 目标检测 (YOLOv8/v11), 传统特征提取, 时序匹配 (DTW/LSTM)
- 架构与工程：**WPF, MVVM, Clean Architecture, DDD, 依赖注入 (DI), 混合编程 (P/Invoke, Pythonnet)
- AI 部署与 GenAI：**PyTorch, ONNX Runtime, vLLM (推理加速), Docker/WSL2, LLM 量化 (GPTQ/AWQ), RAG/Agent 架构
- 系统与运维：**Docker, Linux, MySQL, Redis, CI/CD, Inno Setup, 串口/Mesh/TCP 通信

工作经历 (Professional Experience)

创视自动化 | 视觉系统架构师 (核心骨干) 2025.06 – 至今

核心成就：主导下一代通用视觉平台研发，通过引入 2.5D 视觉技术解决高反光、微小形变等行业级检测难题，构建了研发(Python)与生产(C#)双轨并行的高效开发模式。

项目一：工业级 AI 视觉通用推理平台 (WPF + ONNX + MLOps)

- **痛点解决：**解决传统深度学习模型在工业现场部署难、环境配置繁琐、冷启动样本不足的问题。
- **项目描述：**基于.NET 8 (WPF) 开发的全生命周期 AI 视觉平台，集成了从“数据标注 -> 模型训练 -> ONNX 转换 -> 自动部署 -> 实时推理”的完整 MLOps 闭环。系统支持无监督异常检测 (PatchCore)、目标检测 (YOLO) 及图像分类 (ResNet) 三大核心任务。采用 MVVM + AvalonDock 架构，实现了类似 Visual Studio 的现代化停靠布局，支持多任务并行处理与实时监控。
- **核心技术栈：**
 - **框架：**.NET 8, WPF, MVVM (CommunityToolkit), MahApps.Metro
 - **UI组件：**AvalonDock (停靠布局), LiveCharts (实时图表)
 - **AI 引擎：**ONNX Runtime (推理), PyTorch (训练), Pythonnet/Process (互操作)
 - **数据/配置：**YamlDotNet, Newtonsoft.Json, SQLite/JSON (历史记录)
- **核心职责与技术亮点：**

1. MLOps 全流程自动化闭环 (The "No-Code" Pipeline)

- **训练流水线设计：**设计了 PythonProcessManager 服务，通过 Process 管道技术与标准流解析，实现了 C# 对 Python 训练脚本的全自动管控。
- **环境自适应：**实现了环境自动部署机制，无需用户手动配置 Python 路径，自动挂载 torch 与 onnx 依赖环境。
- **自动化部署：**构建了“一键训练”机制，训练结束后自动将 PyTorch 模型导出为 ONNX，并通过标准化配置自动更新模型注册表，实现了从“训练结束”到“上线推理”的零人工干预。

2. 高扩展性架构设计 (Architecture)

- **现代化 UI 架构：**引入 AvalonDock 实现可拖拽的停靠布局系统，支持检测视图、属性面板、统计仪表盘的自定义排列；基于 VS2022 Light 主题定制了统一的视觉风格。
- **插件化任务引擎：**基于工厂模式抽象了模型推理基类，统一了模型推理的调用接口，支持新算法的快速插件式扩展。

3. 核心算法与数据可视化

- **亚像素级缺陷分割：**针对 PatchCore 输出的热力图，引入 BFS (广度优先搜索) 洪水填充算法进行后处理，实现了异常区域的精确轮廓提取与 NG 溯源标记（红色虚线框 + 分数标签）。
- **实时统计仪表盘：**集成 LiveCharts 构建动态数据看板，实时计算良品率 (Yield Rate)、P95 延迟及各类缺陷分布 (OK/NG/Marginal)，支持按时间维度的历史回溯。

4. 健壮性与数据治理

- **配置热备份：**开发了配置快照系统，支持配置文件的 ZIP 一键导出/导入，并在导入前自动执行完整性校验与自动备份，防止现场误操作导致配置丢失。
- **数据集向导：**设计了可视化的数据集管理向导，支持 MVtec/YOLO/ImageFolder 三种标准格式的自动结构化创建与校验，降低了数据准备门槛。

项目二：高性能大语言模型私有化部署架构 (Local LLM Private Cloud Architecture)

2025.10 – 至今

项目背景：针对数据隐私敏感与离线开发场景，基于消费级硬件 (RTX 50-Series) 构建高吞吐、低延迟的本地大模型推理集群，实现代码辅助与知识库的私有化落地。

• 容器化异构计算架构 (WSL2 + Docker + vLLM)：

- **架构设计：**设计了基于 Windows 宿主 + WSL2 子系统 的混合部署架构。利用 Docker 容器化技术隔离推理环境，解决了 CUDA 版本冲突与依赖问题。

- 硬件直通 (Passthrough): 配置 NVIDIA Container Toolkit 实现 GPU 算力在虚拟化层的零损耗透传，成功在 WSL2 环境下调用 RTX 5070 Ti 全部算力。
- 网络拓扑: 通过 Windows 端口转发 (Port Forwarding) 与 Docker 网桥配置，实现了从宿主机 localhost 到局域网 (LAN) 任意终端的无缝访问，构建了家庭/办公室局域网内的 AI 服务中心。

- 推理性能极致优化 (Inference Optimization):

- 显存管理: 针对 16GB 显存限制，采用 GPTQ-Marlin 4-bit 量化 技术部署 Qwen2.5-14B 模型。在保持模型精度的前提下，将显存占用从 28GB 压缩至 9GB，通过 vLLM PagedAttention 机制最大化 KV Cache 利用率，实现了长上下文 (8k+) 的流畅推理。
- 高并发引擎: 摒弃低效的 HuggingFace Transformers 推理，采用 vLLM 推理引擎，利用 Continuous Batching 技术，将推理吞吐量 (Tokens/s) 提升 10 倍以上。

- 全栈应用集成 (Full-Stack GenAI):

- API 标准化: 封装 OpenAI Compatible API 接口，实现了后端模型与前端应用的解耦，为后续接入 LangChain/AutoGen 智能体框架预留了标准接口。
- 交互层交付: 部署 Open WebUI 可视化前端，实现了类似 ChatGPT 的对话体验，支持多模型切换、参数热调及 RAG (检索增强生成) 知识库扩展。

项目三：2.5D 多模式精密检测系统 (PMD 相位偏折 + Photometric Stereo)

- 商业价值与痛点解决:

- 低成本替代 3D: 针对昂贵的 3D 线扫方案，研发基于 2.5D 视觉 (PMD + PS) 的低成本替代方案。有效解决了微小形变与高反光表面的检测难题，硬件成本大幅降低。
- 多模式融合:
 - 集成 相位偏折 算法，通过面阵光源以及线扫时序频闪光源，重建6张形貌图，提高表面的检测能力。
 - 集成 光度立体 算法，通过 24 分区光源控制，合成法线图/高度图/曲率图等 6 通道数据，大幅提升了对表面微观纹理的检测能力。

- 高维度架构设计 (Clean Architecture + DDD):

- 硬件抽象层 (HAL): 通过接口设计，隔离了海康 SDK 与控制器私有协议，增强了不同品牌相机与不同光源控制器的扩展能力。
- 策略引擎: 基于策略统一了不同的采集流程，实现了采集逻辑的“热插拔”。

- 极致性能调优 (Performance):

- 异构计算: 设计 Python (研发) 与 C++ (生产) 双后端。Python 端通过 Cython + OpenMP 并行化重写解包算法，将处理耗时从 4s 压缩至 40ms，整体性能提升 5-10 倍。
- 混合编程: 通过 P/Invoke 封装 C++ 动态库，配合内存指针实现零拷贝传输，进一步提高性能。

东莞创机电业制品有限公司 (TTI) | 机器视觉开发工程师 2022.07 – 2025.06

核心职责：负责工业物联网 (IIoT) 平台搭建、标准化测试软件开发及 AI 算法在测试领域的落地应用。

1. 算法工程化落地 (AI & Algorithm)

- 感温线混淆识别 (fastDTW 算法): 针对产线感温线型号混用痛点，独立开发基于 fastDTW (动态时间规整) 的时序匹配算法，嵌入 LabThinker 平台。从源头杜绝了数据采集错误，大幅减少了后续的人力 Review 成本。
- 电机故障诊断 (LSTM): 负责从数据清洗、特征工程到模型训练的全流程，基于 LSTM 网络实现了对设备异常电流波形的自动分类与预测。

- **图像去模糊 (GAN)**: (研究型项目) 基于 DeblurGAN-v2 引入边缘损失函数，训练端到端图像去模糊模型，有效提升了运动模糊场景下的图像识别率。

2. 工业物联网与标准化平台 (IIoT & Platform)

- **LabThinker 标准化测试平台**: 参与开发覆盖中国、越南多地工厂的标准化测试架构。集成数据库、TCP/UDP 通讯、电池管理模块，实现了测试流程的规范化与可复用化。
- **Climate Chamber 物联网系统**: 设计并开发连接云端与硬件的通信控制系统。采用 Mesh 组网 与串口多线程并发技术，实现了对 100+ 节点设备的实时数据采集与 Dashboard 可视化监控 (Flask + MySQL + Docker)。

教育经历

- 河南工业大学 | 电子信息工程 | 学士 | 2018.09 – 2022.07