

基本信息

姓名: 王添毅

电话: 13514685743

个人主页: <https://hill-wu-1998.github.io/>


专业: 控制科学与工程

微信: hillwu1998

毕业时间: 2026.7

电子邮箱: zjuwty@outlook.com

Github: [Hill-Wu-1998](#)



教育背景

浙江大学 (985 工程)

控制科学与工程 (博士研究生)

2023.9-至今

• 浙江大学网络传感与控制研究组安全组, 指导老师[王聪研究员](#), 研究方向 AI 模型及系统攻击与防御

浙江大学 (985 工程)

电子信息 (硕士研究生)

2021.9-2023.7

• 浙江大学网络传感与控制研究组安全组, 指导老师[程鹏教授](#), 研究方向协议模糊测试

郑州大学 (211 工程)

自动化 (本科)

2016.9-2020.7

• 毕业设计为[基于 STM32 的同步机械臂](#) (16K+阅读, 360+收藏), 曾获多次获得三好学生, 校级奖学金等荣誉

科研经历

2024.1-2024.8

CNN-based 目标检测模型鲁棒性优化技术 (项目地址: [Underload Code](#))

CVPR 2025 一作

◆ 在延迟攻击下, 发现被攻击样本中的对象区域存在天然鲁棒性, 并依此设计注意力区域加入对抗训练算法中

◆ 使用 Nvidia Nsight 相关工具进行性能分析, 发现针对不同设备在延迟攻击下存在性能瓶颈迁移现象

◆ 提出硬件自适应对抗训练算法维持延迟攻击下的推理实时性 (13 FPS 到 43 FPS), 同时维持模型精确度

◆ 论文: [Learning Robust and Hardware-Adaptive Object Detectors against Latency Attacks for Edge Devices](#)

2024.8-2025.3

Transformer-based 检测模型鲁棒性分析 (复现项目地址: [Attn Fool Code](#))

ICCV 在投一作

◆ 针对 DETR 系列模型存在的独特攻击面设计一种新的对抗攻击方法, 基于攻击中发现的特殊现象调整攻击算法

◆ 复现针对 Attention 机制进行对抗攻击的工作 ([Attention Fool 论文](#)), 并结合我们的攻击发现一些有意思的现象

2023.9-2024.8

面向分布式端侧系统的异构模型融合优化技术

AAAI 2025 三作

◆ 通过对抗攻击方法探测异构模型决策边界, 利用 PGD 攻击对最近边界点的动态性进行建模

◆ 优化损失函数使得蒸馏过程关注接近决策边界的样本, 支持卷积与 Transformer 模型混合分布式架构, 将异构模型的分布式训练精度提升 0.5%-3.5%。

◆ 论文: [Fed-DFA: Federated Distillation for Heterogeneous Model Fusion through the Adversarial Lens](#)

2024.3-2024.12

面向端侧异构设备的多任务混合模型并行推理加速技

ICDCS 2025 三作

◆ 提出了一种多任务混合模型并行推理机制, 该机制采用动态规划和负载均衡双层优化策略, 旨在减少处理器间内存带宽竞争开销, 实现模型的流水线并行推理。

◆ 在包含 ARM CPU、OpenCL GPU、华为 DaVinci NPU 等异构多核处理器的并行任务环境中进行测试, 对于麒麟 990、高通骁龙等多架构 SoC, 该机制可将推理速度提升 2-8 倍, 显著提升了多任务在端侧的混合推理效率。

◆ 论文: Hetero²Pipe: Pipelining Multi-DNN Inference on Heterogeneous Mobile Processors under Co-Execution Slowdown

其他相关成果

论文: [KDD 2025 在投] Learning Pairwise Federated Distillation Online via Bandits with Hidden Context for Heterogeneous Model Fusion

CCF A 类会议 二作

论文: [IEEE TSE 录用] Better Pay Attention Whilst Fuzzing

CCF A 类期刊 六作

论文: [INFOCOM 2024 录用] Reverse Engineering Industrial Protocols Driven By Control Fields

CCF A 类会议 五作

漏洞: CNVD 证书 17 项, 其中高危 6 项, 中危 11 项