张可为

(+86) 155-2770-6812 | 个人主页 | xiwen.kwzhang@gmail.com | zkwsdsg | github.com/xiwen1

基本信息

武汉大学 计算机学院 软件工程 预计毕业时间：2026 年 6 月

GPA: 3.91 / 4.0 | GPA 排名：11 / 222 | 综合排名：1 / 222

- **专业能力**：计算机图形学基础（98）、概率论与数理统计（93）、线性代数（95）、离散数学（97）、面向对象程序设计（94）、计算机组成与设计（90）、数据库系统（93）、软件需求与建模（97）、系统级程序设计（93）、科技写作（95）、数字图像处理（94）、商务智能（94）、中级项目实践（95）。
- **编程能力**：我对编程充满热忱并经验丰富，熟练使用包括 Python、Rust、Go、C# 以及 JavaScript 在内的多种编程语言以及相关开发框架及深度学习框架，了解 CUDA、Triton 等底层深度学习开发工具和 GPU 编程，掌握 Accelerator、DeepSpeed 等多卡并行训练方法，累计编程时长超过 1800 小时。[编程时间统计]
- **工程经验**：我的工程经验丰富，擅长构建功能划分清晰、工程结构严谨的深度学习代码，精通 web 全栈开发（前端：React.js、Next.js，后端：Axum、Gin、Django 等）以及各类常见桌面端应用开发技术，在 Github 拥有众多开源项目[Github 主页]，累计获得 31 个 star。
- **英语能力**：全国大学英语六级考试 586 分，独立完成并投稿多篇英语学术论文。[Google Scholar]
- **在校荣誉**：在本科的前五个学期中，我曾获得国家奖学金、雷军计算机奖学金（￥10000）、甲等奖学金、新生奖学金，在 2023 年获得第十八届中国“挑战杯”竞赛全国一等奖，多次在蓝桥杯、“互联网 +”以及美国大学生数学建模大赛等学科竞赛中斩获奖项并收获奖金，累计获得奖金达 47000 元。

研究经历 [计算机视觉安全]

重新审视对抗性补丁：设计相机无关的攻击以对抗行人检测 已收录于 NeurIPS 2024 | **共同第一作者**

研究方向：对抗攻击、对抗训练 | [项目主页]

- **研究背景**：现有基于补丁的行人检测攻击方法在物理空间部署时面临显著挑战：1）对抗补丁需经打印-拍摄流程，其攻击效果受相机 ISP（Image Signal Processing）参数影响显著，实验表明不同设备间攻击成功率波动高达 47.3%；2）尽管物理条件鲁棒性已被广泛研究，但相机 ISP 对攻击效果的调制作用尚未被系统探索；3）我们首次将相机 ISP 建模纳入对抗补丁优化框架，旨在实现相机无关的物理攻击。
- **研究方法**：1）提出**可微分 ISP 代理网络**，通过参数化建模模拟不同相机 ISP 对图像色彩空间的非线性映射，有效桥接物理-数字域鸿沟；2）设计**对抗性优化框架**，将 ISP 代理网络作为防御模块与补丁攻击模块进行对抗训练：攻击模块优化扰动以最大化检测失败率，防御模块优化 ISP 参数以最小化攻击效果，实现相机无关的鲁棒攻击；3）构建**多设备评估基准**，涵盖 12 种主流相机型号，支持端到端物理攻击性能评估。
- 实验结果表明，1) 我们的方法成功在多种不同的相机下隐藏了人体，在多种 ISP 的场景下超越 SOTA 方法约 15% 的攻击成功率；2）我们的对抗性优化框架被证明是极为有效的，消融实验表明，相比于不适用对抗性优化框架的情况，我们的方法在物理空间提升了 31.5% 的攻击成功率。

摩尔纹后门攻击 (MBA)：物理世界中行人检测器的新型触发器 已收录于 ACMMM 2023 | **第三作者**

研究方向：后门攻击、对抗攻击

- **研究背景**：1）现有针对行人检测系统后门攻击存在研究空白：现有工作主要集中在交通标志识别等有限场景，缺乏对动态人体目标的针对性攻击方法；2）传统空间域触发器（如斑块图案）存在显著可见性缺陷，在物理部署中易被防御系统检测；3）自然现象启发的触发机制研究不足，未充分利用成像系统固有特性构建隐蔽攻击通道；
- **研究方法**：摩尔纹效应是由相机传感器引起的常见现象，出现在密集条纹的数字图像中，如纱窗、LED 显示器和带条纹的衣物，特征明显且常见。本研究提出将摩尔纹作为行人检测器的后门攻击触发器，基于摩尔纹的形成原理，在数字空间中进行了建模并模拟多种可能的现实空间变换，以此在行人数据集的图像上生成摩尔纹中毒样本。通过向训练数据集中按照少量比例注入中毒样本，便可以在多种目标检测器模型上产生十分明显的后门攻击效果。
- 我们的方法为在人体衣物上产生摩尔纹提供了一种十分有效的建模方法，为攻击者提供了充足的自由和灵活性。
- 实验结果表明我们的方法仅需要对于 10% 的样本下毒就可以在目标模型上达到 77.3% 的攻击成功率。
- 针对横跨多个后门攻击触发器的隐蔽性用户调研表明，仅有 8.3% 的用户能够发现我们方法的触发器，相比于其他方法下降了 73.4%。

姿态确实重要: 为有效的隐藏人体攻击设计的关键点引导的对抗性补丁 已提交至 CVPR 2025 | **第一作者**

研究方向：对抗攻击、关键点检测

- **研究背景**: 1）现有行人检测对抗攻击存在**评估指标误导性问题**：主流方法通过降低平均精度（AP）制造攻击成功假象，但实验表明其本质是诱发检测器输出多个重叠误检框（检测框数量增加约 4 倍），实际人体区域仍被持续识别；2）特征层面分析表明单补丁方案存在显著覆盖盲区，未扰动区域在 ResNet-50 第四卷积层仍保有 86% 原始激活强度；3）现有方法未考虑物理部署中人体姿态变化导致的补丁形变，数字仿真与物理实现间存在系统性偏差。
- **研究方法**: 1）我们提出了一种关键点引导的多补丁攻击方法，通过集成 OpenPose 模块检测人体关键点数据，并建立 17 个解剖学锚点与 8 个补丁部署位置的动态映射关系，以此部署对抗补丁，从而保证数字空间中补丁的位置与物理空间的对齐；2）我们还提出了**与姿态对齐的真实变换**，可以根据关键点数据中的姿态信息，融合**透视变换核**与**遮挡概率矩阵**，实现数字空间中的补丁变换与物理空间高度对齐，在训练阶段预补偿物理形变导致的攻击性能衰减；
- 我们的方法在解决了多框检测问题的同时，也在**隐藏攻击成功率**上以大幅领先先前的方法 63%，大量物理实验表明，我们的方法是目前唯一能在各种使用官方权重的检测器下使人体几乎完全隐身的方法。

多样化操作下的对抗性水印: 基准测试与超越 已提交至 TIP | **共同第一作者**

研究方向：水印保护、对抗攻击

- **研究背景**: 1）揭示现有水印保护方法在复合攻击场景下的脆弱性：传统方案在经历大多数单个变换（如旋转、裁剪和压缩）后对水印去除网络的防御成功率普遍下降超过 30%，面对多个变换则性能退化更加明显；2）指出数字版权保护领域的评估体系缺陷：在水印保护的场景下 SSIM/PSNR 等指标与人类视觉感知相关性很低，无法有效量化水印防御性能；3）对抗攻击者可通过级联图像处理操作构建攻击流水线，现有防御方案缺乏系统性应对策略。
- **研究方法**: 1）引入了**变换集成模块** (TEM)，构建包含 9 种可微分图像处理算子的参数化空间，通过蒙特卡洛采样生成复合攻击序列，提高对抗水印应对不同变化序列的鲁棒性；2）设计注意力引导的扰动分配方案：基于类激活映射（CAM）定位水印敏感区域，在保证水印攻击性基本不变的前提下将有效防御区域缩小至水印周边的环切范围，显著提升水印隐蔽性；3）我们还提出了 **水印透明度一致性模拟** (ACE)，推导水印透明度参数与图像像素矩阵的耦合关系方程，通过拟合经过水印去除网络后的水印透明度，准确评估对抗水印的保护效果。
- 在 CLWD 测试集上达到 95.3% 的水印保留率（较 SOTA 提升 23.7%）。针对多个开源模型的攻击测试显示，本方案在白盒/黑盒场景下分别保持平均 91.4% 和 84.6% 的防御成功率，针对三种商业 API 的攻击测试同样达到了平均 65.7% 的防御成功率。在面对多种图像变换和压缩操作下稳定性高达 98%。

隐身衣: 面对各种各种失真场景鲁棒的对抗纹理 计划投稿于 ACMMM 2025 | **预计共同第一作者**

研究方向：物理对抗攻击、目标分割

- **研究背景**: 1）现有物理对抗攻击方法在以**动态模糊为主图像失真**场景中存在显著性能衰减；2）图像修复算法对对抗纹理的二次干扰尚未被系统化研究，传统方法未建立从图像采集到后处理的全流程攻击模型；3）先前工作中获取对抗纹理的方法存在很严重的局限性：主流工作 AdvTexture 依赖周期性补丁拼接策略，导致数字仿真与物理部署间的域差异，从而潜在地时攻击性能受限；AdvCaT 采用 3D 建模合成训练数据集的方案，存在生成样本多样性不足与现实场景适配性差的双重缺陷，同时在数据集的获取上也相当不便；
- **研究方法**: 1）分析并证明了在常见的图像失真和图像修复中，图片中的低频部分通常能被完整保存，通过谱分解技术分离图像高低频分量，设计频域敏感损失函数（$\mathcal{L}_{LF} = \|\mathcal{F}^{-1}(H(u,v) \cdot \mathcal{F}(x))\|_2$）强化低频对抗模式；2）构建**端到端失真管道**，集成运动模糊等 12 种失真算子进行联合对抗训练；3）开发**自适应部署系统**，基于 FastSAM 实现实时衣物分割与透视变换补偿，实现纹理部署时数字空间与物理空间的对齐，通过几何感知网络动态调节纹理密度。
- 构建包含 12 种复杂失真类型的多模态测试集，在数字仿真环境中达到 82.4% 的平均攻击成功率，物理部署测试显示该方法在 3-15 米有效攻击距离内保持 78.2% 的稳定性能。对图像修复网络 Resformer 对攻击成功率达到 79%。

研究经历 [可信生成大模型]

偏见发现:LVLM 基于反事实解释的自动化模型诊断 计划投稿于 NeurIPS 2025 | **预计第一作者**

研究方向：LLM、模型诊断、因果推理

- **研究背景：**当前视觉模型诊断面临双重困境：现有方法依赖人工预设偏见类型（如 UMO），且仅作相关性分析而缺乏系统性针对诊断结果的反事实验证机制，诊断缺乏可信度且人工验证困难。同时，当前的模型诊断工作中缺乏对于该任务的基准测试，难以比较各个模型诊断方法之间的性能高下。
- **研究方法：** 1）提出**两级反事实解释框架**：通过 LVLM 构建语义-视觉联合推理模块，首先解析模型的错误模式空间（Error Pattern Space），生成可验证的偏见假设（如" 模型将长发错误关联为女性特征"），再通过对于错误样本进行干预生成反事实样本进行因果验证；2）设计**动态因果干预协议**：集成 FlowEdit 等工具链，在保持图像语义完整性的前提下，对目标属性进行精准干预；3）建立**自适应测试基准**：开发参数化偏见注入引擎，支持通过贝叶斯网络动态生成包含显式/隐式偏见的诊断数据集，并定义来量化评估指标。

- 1）首创基于 LVLM 的模型诊断范式，极大提高了偏见诊断结果的可信度与可解释性；2）构建首个开放域动态诊断测试平台，支持对多种任务模型进行脆弱性分析；3）首次提出了模型诊断领域的量化评估方法，结果显示我们的方法在模型诊断的准确性上性能最优。

主观相机: 无需训练的场景草图引导生成与空间奖励优化 计划投稿于 ICCV 2025 | **预计第三作者**

研究方向：图像生成、扩散模型

- **研究背景**: 当前文生图模型在细粒度空间控制上面临多重挑战：1）纯文本输入难以精确描述复杂场景的几何布局；2）现有草图控制方法依赖高质量轮廓输入（如 ControlNet 草图控制训练中需专业级草图），限制了非专业用户的使用；3）同时，多目标场景生成存在语义-几何错位问题，仅凭口头语言描述和单一草图输入很难达到输出结构与用户预期的视觉一致性。本研究致力于构建无需训练的双模态控制框架，通过草图-文本联合优化突破上述限制。
- **研究方法**: 1）提出**分层感知优化框架**，通过 CLIP/BLIP 语义对齐损失、PickScore 生成质量损失和美学评估损失构建多模态奖励函数，采用交替优化策略在潜在空间迭代优化生成结果；2）设计**增量式布局生成算法**，基于 FreeControl 的无训练控制范式，通过用户交互式输入物体轮廓与位置坐标，采用分阶段注意力重加权技术实现从空白画布到完整场景的渐进式构建；3）创新**涂鸦真实化模块**，通过可逆 DDIM 过程建立涂鸦域与真实图像域的映射关系，利用预训练扩散模型在低迭代步数下完成风格迁移，有效解决用户草图粗糙导致的图像失真问题。
- 我们的方法特别关注到了非专业用户的图像创作需求，首次为精细化重构人类主观印象图片提供了可行且高效的解决方案，尤其在多目标场景的创作中，我们的方法在用户调研中主观满意度远超其他方法。

增强型基于扩散路径融合的图像生成技术 《数字图像处理》Excellent Project（课程得分：94）

研究方向：图像生成、扩散模型 | [项目链接]

- **研究背景**: 当前图像生成技术面临双重挑战：1）MultiDiffusion虽能实现任意尺度图像的布局控制，但在细节生成上存在显著不足（如局部纹理失真、特征模糊等）；2）基于草图的 ControlNet 等方法虽能精确控制轮廓，但在多物体场景中难以兼顾全局一致性与局部细节（实验表明超过 3 个物体时特征匹配准确率下降约 50%）。为此，我们提出融合两类方法的优势，实现从全局布局到局部细节的精准控制。
- **研究方法**: 1）融合布局引导与细节控制：将 ControlNet 注入 MultiDiffusion 的管道中，实现了对每个扩散通道单独匹配控制草图，实现了对于每个目标的情却控制。2）开发区域化控制协议：支持用户为不同画幅区域独立设置草图引导与文本提示，实现” 全景规划-局部雕琢” 的创作模式。
- **技术实现：** 1）采用 React.js 开发交互式前端，支持草图绘制、区域框选、参数设置等操作；2）基于 FastAPI 构建后端服务，实现多通道扩散路径的并行计算；3）设计模块化架构，支持 ControlNet 等多种控制方法的即插即用。

荣誉奖项 [精选]

一等奖 | 第 18 届“挑战杯”全国大学生课外学术科技作品竞赛 (总决赛). [报道链接] 2023 年 9 月
Honorable Mention | COMAP 2024 MCM/ICM. 2024 年 4 月
二等奖 | 第 15 届蓝桥杯 Python 程序设计竞赛 (湖北赛区). 2024 年 6 月
二等奖 | 第 14 届蓝桥杯 C/C++ 程序设计竞赛 (湖北赛区). 2023 年 6 月

学生活动与课外工作

- 担任武汉大学微软俱乐部技术部部长，负责技术沙龙以及校内比赛的组织与主持，参与社团招新与社员培训。
- 曾担任武汉大学团委青年志愿者协会干事，负责志愿者管理以及信息整理，多次参与组织志愿者活动。
- 曾担任武汉大学计算机学院团委理论研究会干事，承担院内青年大学习数据统计与每周公众号的推文撰写。
- 曾主持武汉大学计算机学院新生购机讲座主持人，指导 2023 级新生选购个人电脑并持续答疑。