南京邮电大学

毕业设计(论文)任务书

|  |  |
| --- | --- |
| 题 目 | 基于真实数据分布的后门攻击软件设计与实现 |
| 专 业 | 信息安全 |
| 学生姓名 |  |
| 班级学号 |  |
| 指导教师签字 |  |
| 指导教师职称 | 教授 |
| 指导单位 | 计算机学院、软件学院、  网络空间安全学院 |
| 单位负责人签字 |  |
| 学院领导签字 |  |
| 日 期 |  |

|  |  |  |
| --- | --- | --- |
| 题　　目 | | 基于真实数据分布的后门攻击软件设计与实现 |
| 课题类型 | | 工程设计√ 理论研究□　 其他□ |
| 设计内容与技术要求  、  成  果形式 | 一、设计内容  深度神经网络的训练依赖海量的数据，这些数据可能具有多个来源，并且攻击者无法完全访问单个数据源。这种情况下，传统的后门攻击受限于良性数据和恶意数据之间的特征纠缠，不能满足真实的攻击场景。该课题旨在提出一个基于真实数据分布的后门攻击算法，并设计实现基于该算法的后门攻击软件系统。具体任务及要求如下：  （1）熟悉Python编程语言和PyTorch深度学习框架；  （2）熟悉机器学习基本概念，学习后门攻击的基本原理；  （3）设计基于真实数据分布的后门攻击方法；  （4）实现后门攻击算法，对计算机视觉领域主流任务（例如图像分类）具有不同分布特征的真实数据进行后门攻击，影响主流图像识别模型（例如ResNet、MobileNet）的识别结果；  （5）在实现后门攻击的基础上，编写代码完善系统界面。系统界面能够选择真实数据的分布特征，并可视化提出的后门攻击方法与传统后门攻击方法的结果对比。  二、成果形式   1. 软件； 2. 毕业设计论文、软件说明书及其系统源码。 | |
| 设计进度 | 2025年1月11日前：完成开题报告  2025年4月xx日前：完成中期报告  2025年5月xx日前：完成毕业论文 | |
| 参考资料 | （建议5篇以上）   1. Gu T , Dolan-Gavitt B , Garg S .BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain[J].arXiv, 2017.DOI:10.48550/arXiv.1708.06733. 2. Gao Y, Doan B G, Zhang Z, et al. Backdoor attacks and countermeasures on deep learning: A comprehensive review[J]. arXiv preprint arXiv:2007.10760, 2020. 3. Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PMLR, 2021: 8748-8763. 4. Wu B, Liu L, Zhu Z, et al. Adversarial machine learning: A systematic survey of backdoor attack, weight attack and adversarial example[J]. CoRR, 2023. 5. Li Z, Sun H, Xia P, et al. Efficient backdoor attacks for deep neural networks in real-world scenarios[J]. arXiv preprint arXiv:2306.08386, 2023. | |