# 神经网络验证平台 PRODeep

张立军研究员 课题组
zhanglj@ios.ac.cn

随着人工智能的高速发展，神经网络模型被广泛应用于生产生活。然而模型缺乏解释、易受攻击等缺陷使得人工智能系统存在诸多安全隐患。为了实现人工智能技术在安全攸关领域的落地，发展可信赖人工智能受到学术界与工业界的高度关注，而其中，鲁棒性是神经网络模型最重要的安全性质之一。

鲁棒性：人工智能系统在任何情况下都能维持其水平的能力。

系统输入数据的变化　　系统行为符合设计期望

· 变化能够被具体刻画
如对抗条件下的鲁棒性（扰动形式、扰动幅度等特征可以被数学描述）

· 变化难以被具体刻画
如自然条件下的鲁棒性（雨雪雾霾、聚焦模糊、异物遮挡等现实因素造成的输入变化）

局部鲁棒性：具体样本邻域内神经网络行为的一致性

给定一个神经网络$f:\mathbb{R}^m \to \mathbb{R}^n$与输入样本$\hat{x} \in \mathbb{R}^m$。称神经网络$f$在$\hat{x}$的邻域$B(\hat{x}) \subseteq \mathbb{R}^m$内是局部鲁棒的，若对于任意$x \in B(\hat{x})$，有

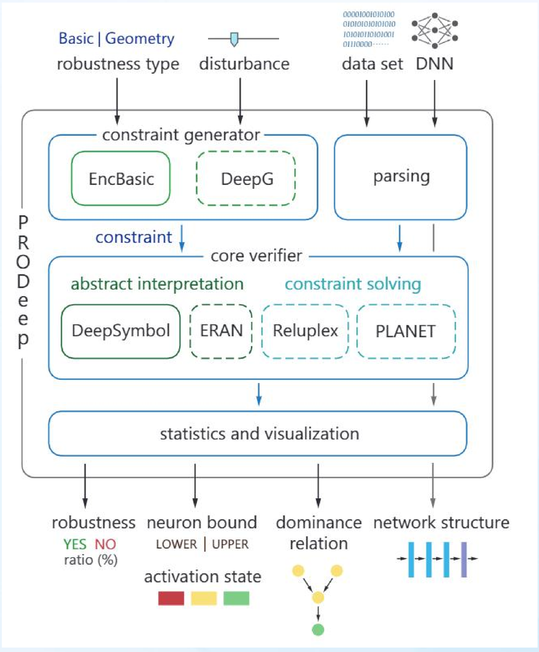
$$C_f(x) = C_f(\hat{x})。$$

局部鲁棒性是深度学习模型一种重要的鲁棒性刻画方式

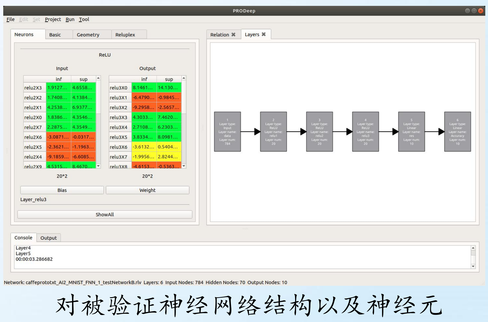
形式化验证是保障计算机系统与其性质具有一致性的重要技术。面向人工智能模型，课题组基于形式化方法研发了神经网络验证平台 PRODeep[1]，实现对深度学习模型的鲁棒性验证。PRODeep 是一个专注于神经网络鲁棒性的验证平台，其中内嵌了多个主流神经网络验证器：

· 基于 SMT 方法　　Reluplex 和 PLANET，
· 基于抽象解释方法　　DeepSymbol [2] 和 ERAN

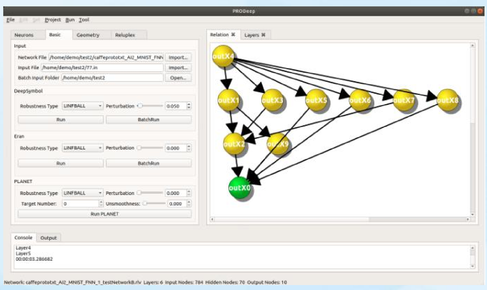
PRODeep采用C++作为底层实现算法语言，并基于Qt进行上层开发，具有较高的可移植性、可拓展性和可维护性，是目前国内领先的神经网络验证工具。 同时，PRODeep提供了用户友好的交互界面，大大缩短了配置时间和用户的学习成本。




PRODeep设计架构：约束生成、模型解析、验证器、可视化输出四大模块



对被验证神经网络结构以及神经元激活状态的可视化



基于哈斯图对输出层节点取值区间偏序关系的可视化

[1] PRODeep: a platform for robustness verification of deep neural networks. ESEC/SIGSOFT FSE 2020: 1630-1634
[2] Analyzing Deep Neural Networks with Symbolic Propagation: Towards Higher Precision and Faster Verification. SAS 2019: 296-319

# 工具/系统/平台/解决方案名称

主要完成
人

联系方式（**请务必写上主要联系人、手机号、邮箱**）

## 具体内容介绍

1. 可从系统简介、功能指标、创新点、所取得的标志性技术进步、可应用领域（应用情况）等方面进行介绍；
2. 请多些图例，**尽量让描述通俗易懂**，更具知识性与趣味性。

**海报的实际制作尺寸为80*180cm，**
**正文部分的大小颜色搭配请自行设计**