

多路径方法在神经网络验证中的 研究与应用

郑焯

导师：刘嘉祥

- 问题：神经网络验证
- 背景：界限传播方法
- 本文：多路径界限传播
- 本文：GPU 并行化
- 贡献和工作量

- 神经网络容易受到人为或非人为的**攻击**
- 安全攸关场景下神经网络的安全性需要得到可靠保证
- 基于**测试**的方法无法提供安全性保证（样本空间是无穷集）



自动驾驶系统事故



对抗路标

Image source: https://en.wikipedia.org/wiki/Smart_system; images from Google Image

- 验证给定的输入集合是否导致不安全输出
- 输入：无穷集（包含扰动阈值内的所有图片）
- 输出：安全或不安全（需要计算无穷集输入下的输出范围）

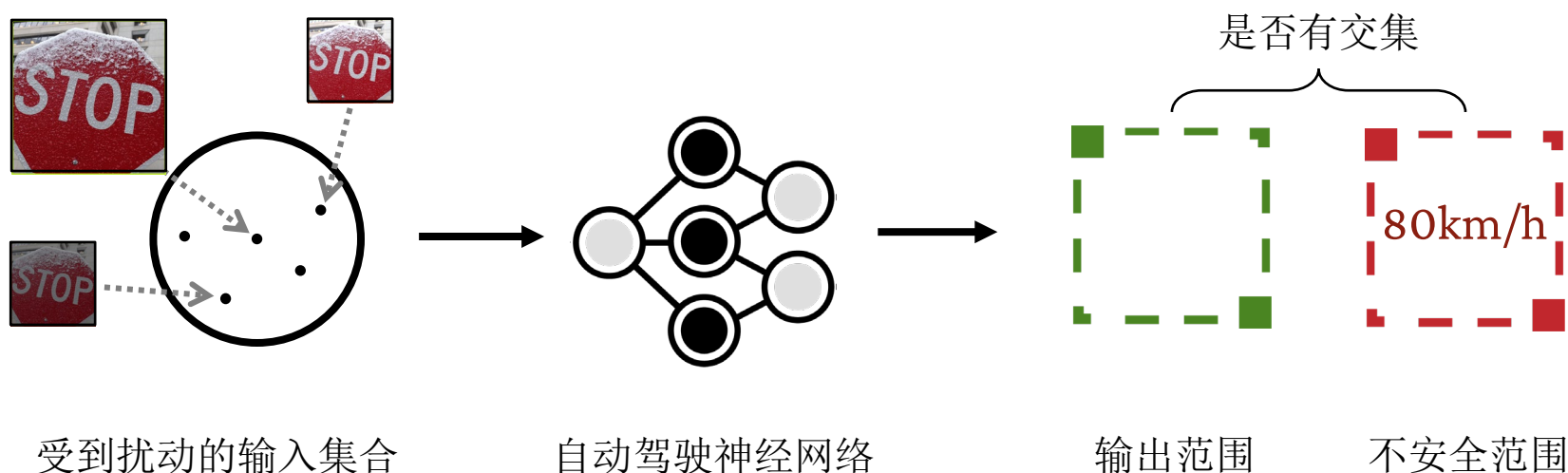


Image source: <https://www.businessinsider.com/why-are-stop-signs-red>

- 困难：非线性激活函数的复合（例如 ReLU）（NP-hard）
- 方法：约束求解、上近似+优化、界限传播
- 本文：将界限传播方法扩展（一般化）到多路径界限传播

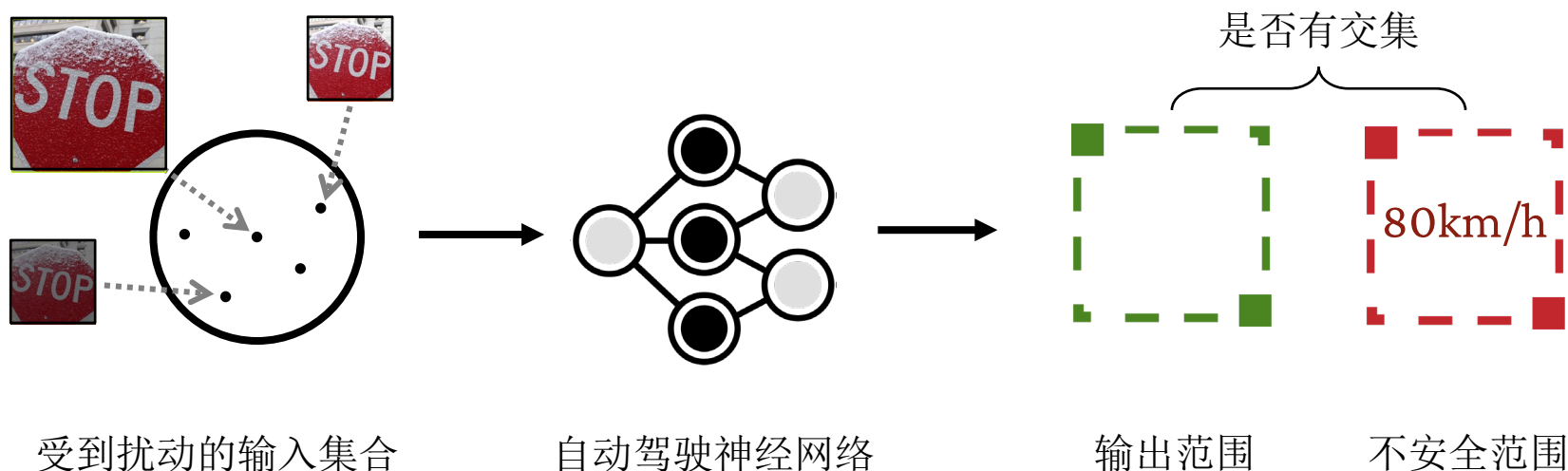
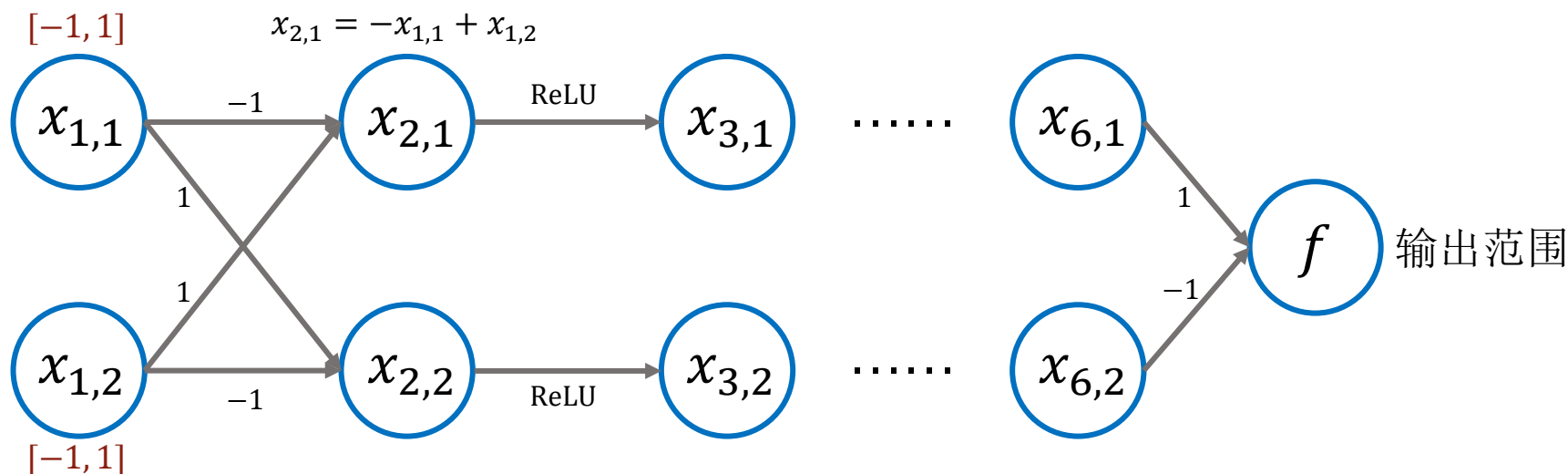


Image source: <https://www.businessinsider.com/why-are-stop-signs-red>

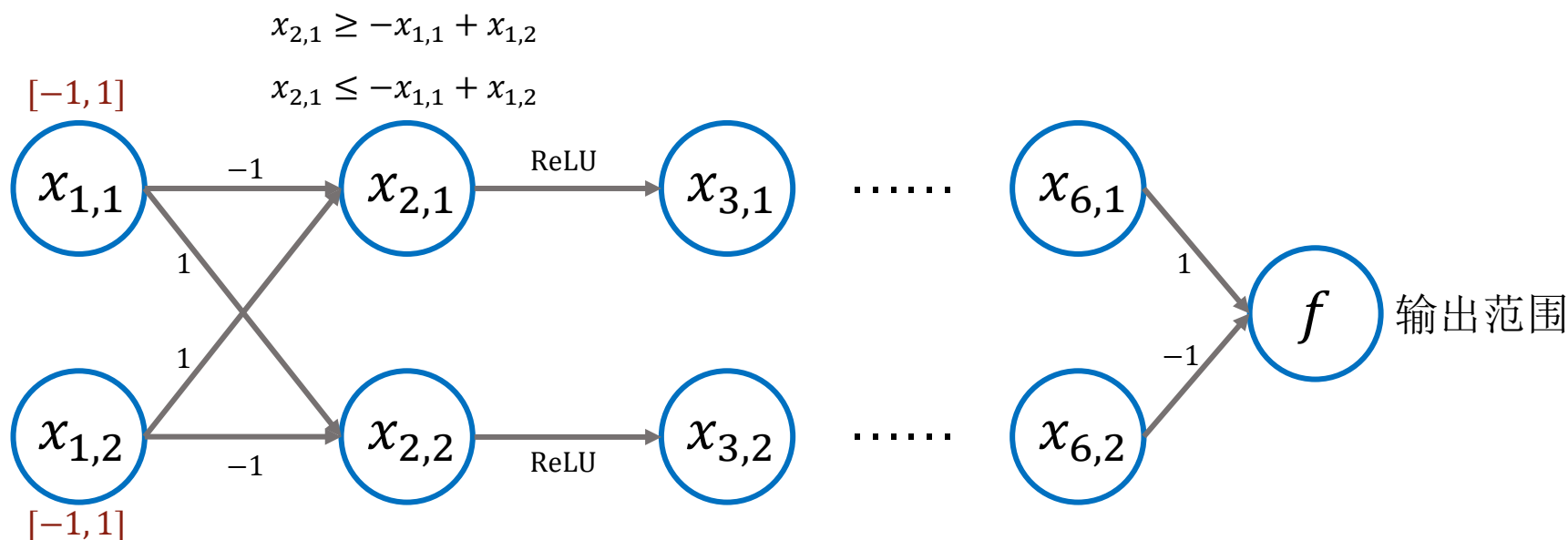
界限传播方法

- 沿神经网络传播界限函数
- 界限函数是关于输入变量的线性上下界



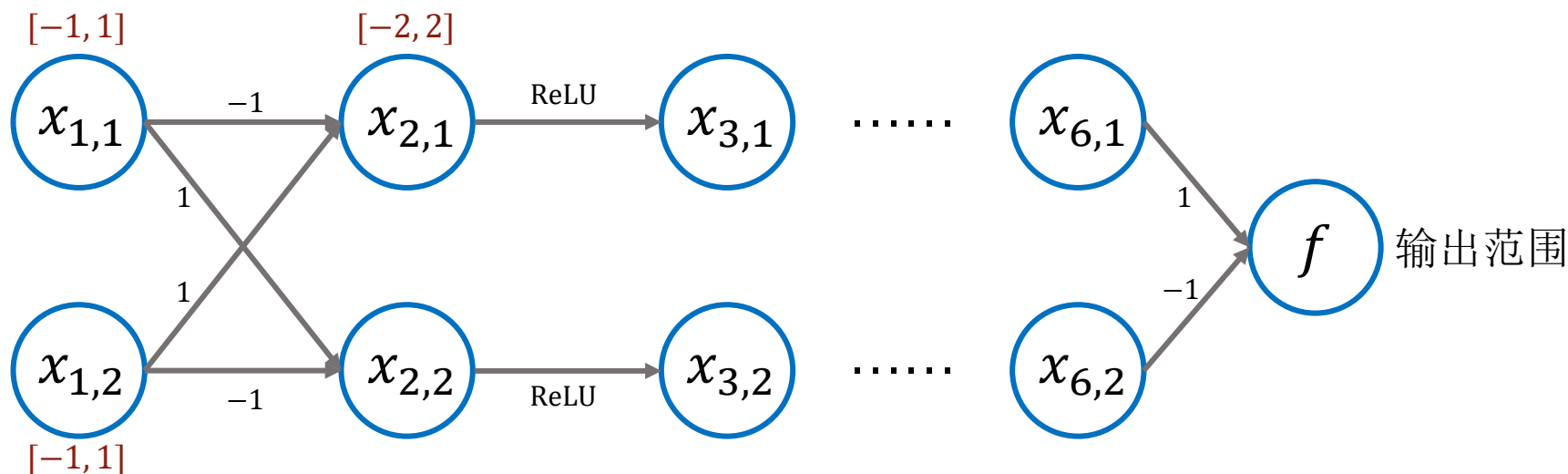
界限传播方法

- 沿神经网络传播界限函数
- 界限函数是关于输入变量的线性上下界



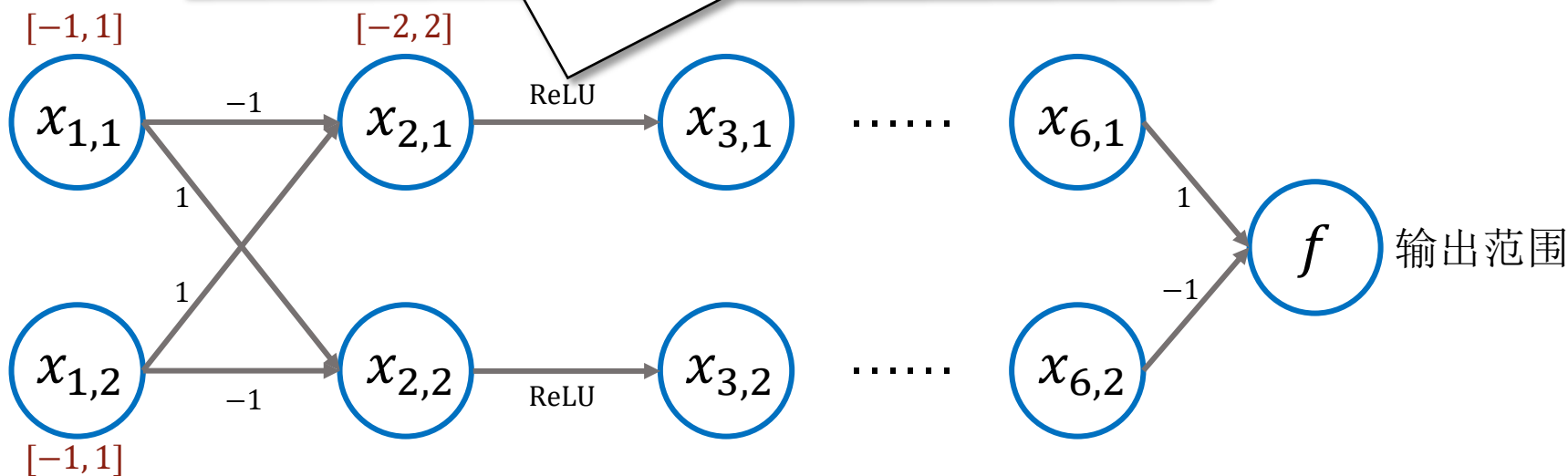
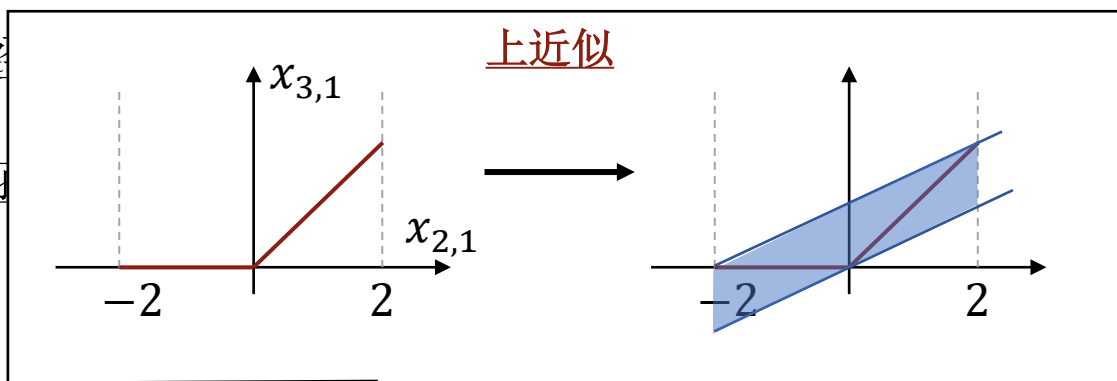
界限传播方法

- 沿神经网络传播界限函数
- 界限函数是关于输入变量的线性上下界



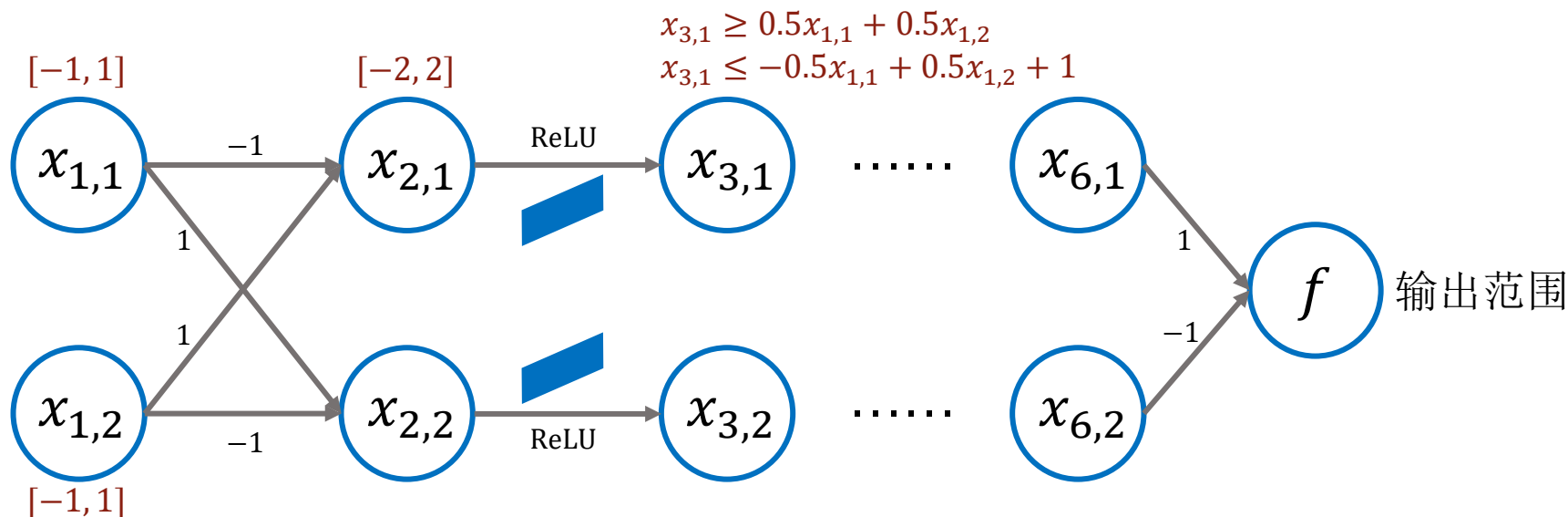
界限传播方法

- 沿神经网络
- 界限函数



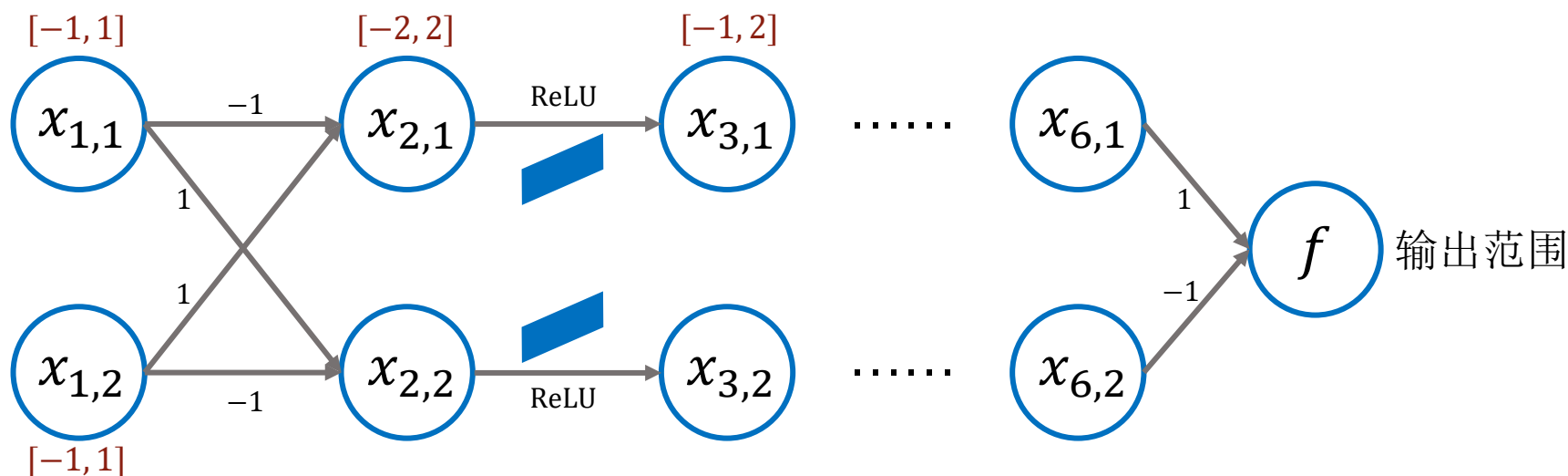
界限传播方法

- 沿神经网络传播界限函数
- 界限函数是关于输入变量的线性上下界



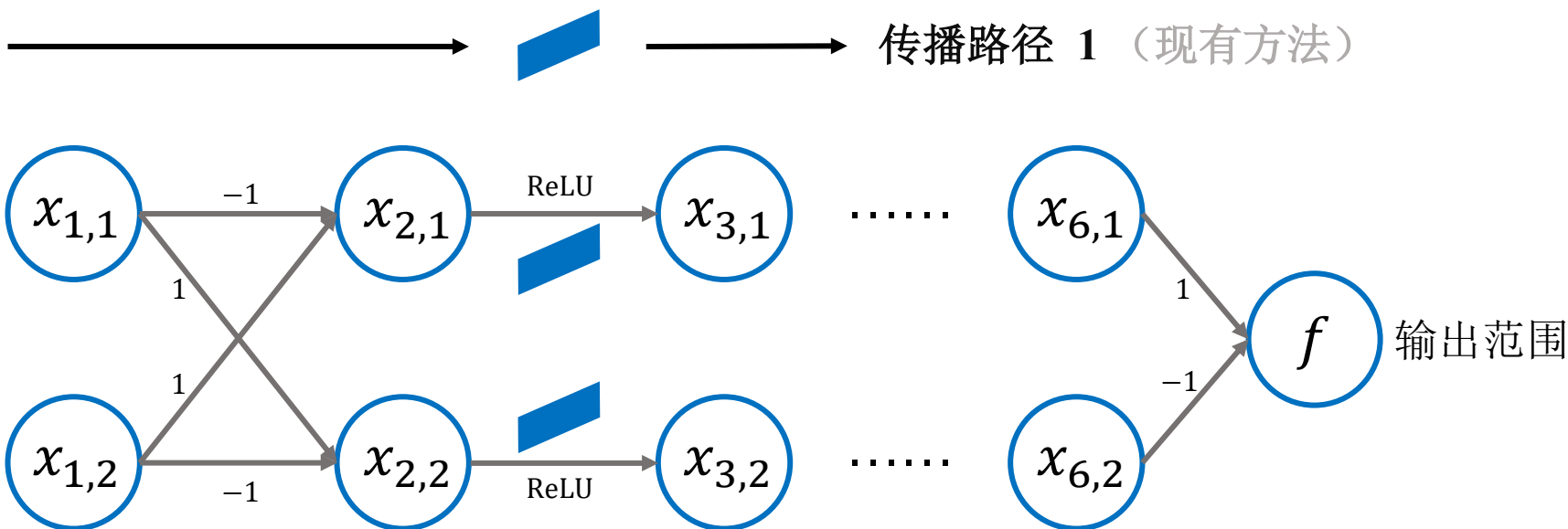
界限传播方法

- 沿神经网络传播界限函数
- 界限函数是关于输入变量的线性上下界

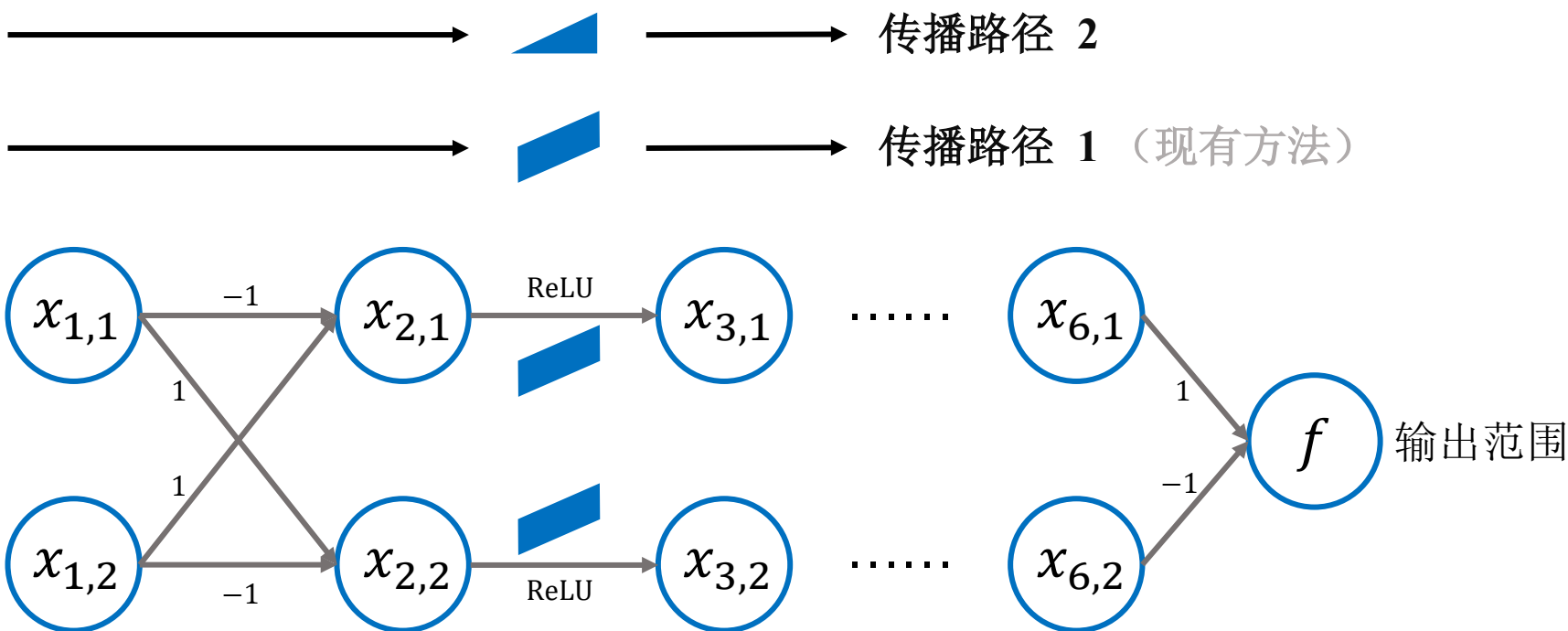


本文：界限传播路径

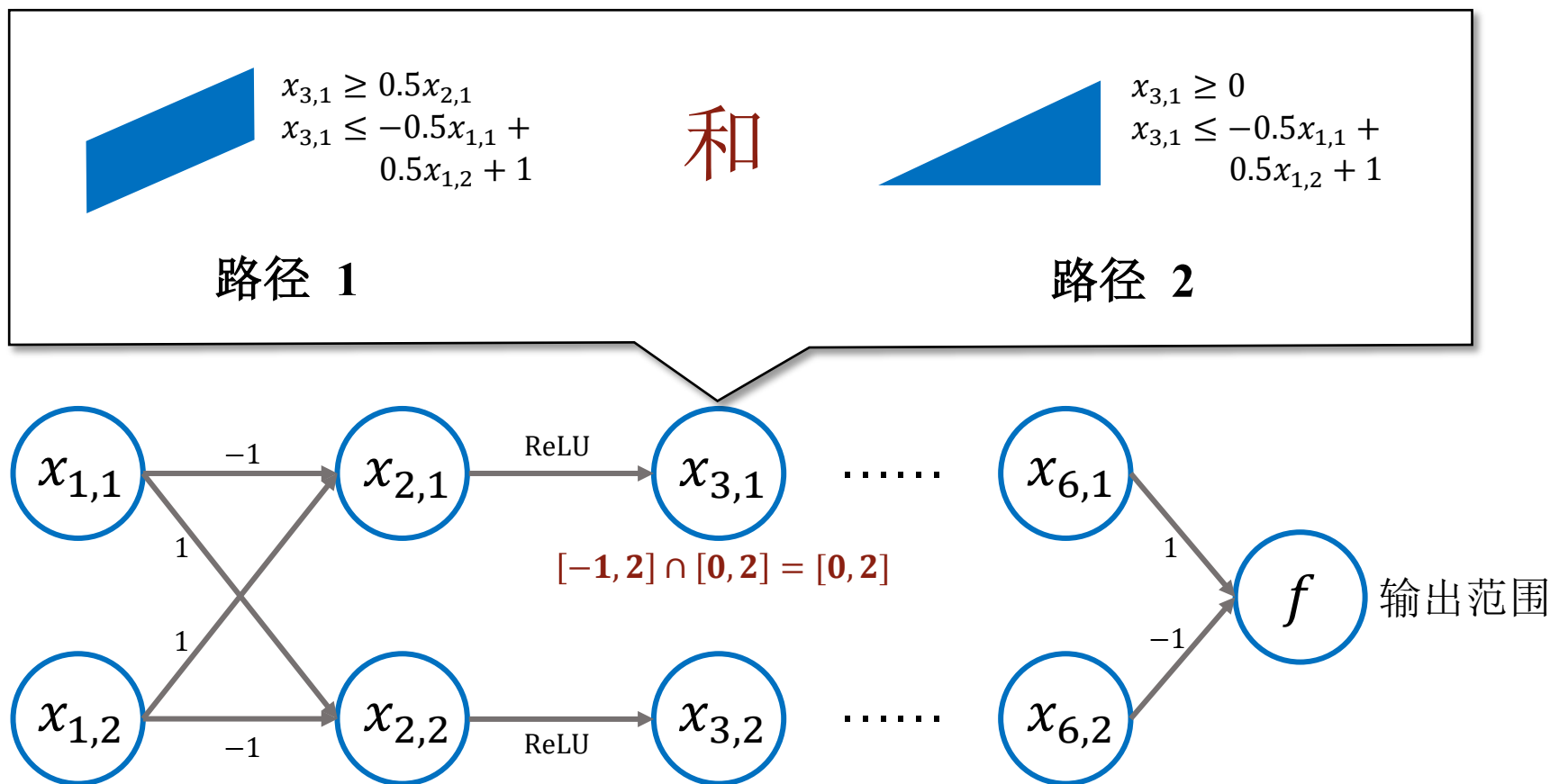
- 传播过程确定了一条“传播路径”



本文：界限传播路径

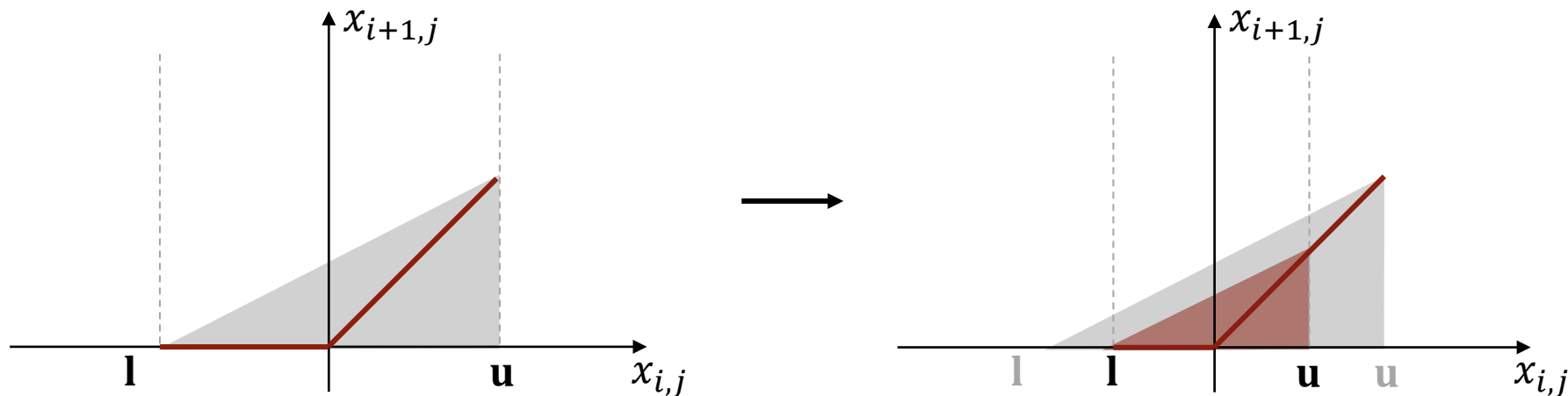


本文：双路径界限传播



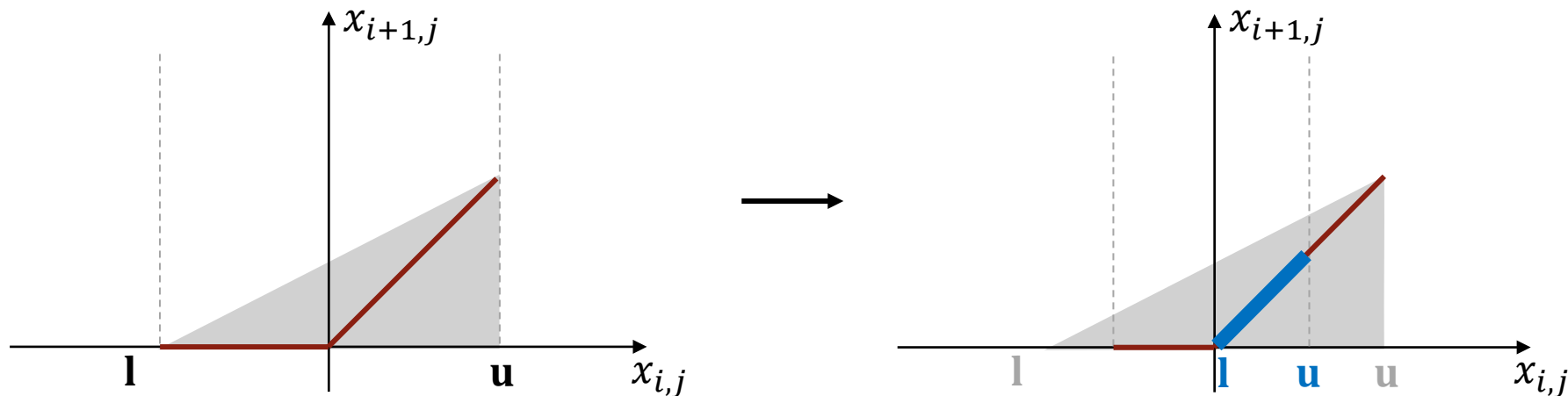
优势：精度积累

- 每个节点的上下界均更精确（相比已有方法）
- 带来更精确的 ReLU 上近似



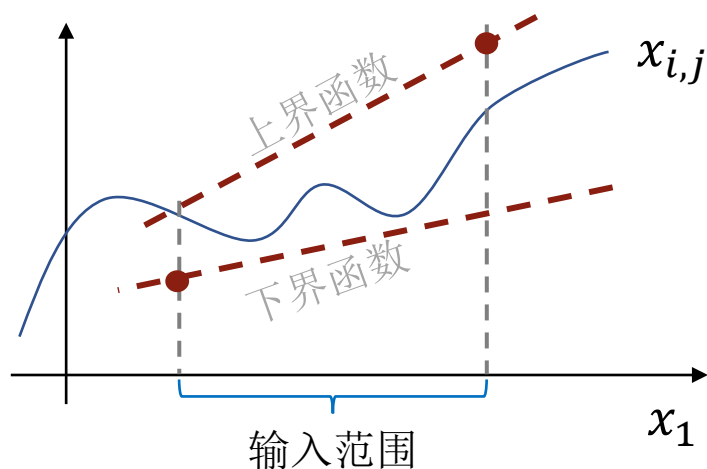
优势：精度积累

- 每个节点的上下界均更精确（相比已有方法）
- 带来更精确的 ReLU 上近似

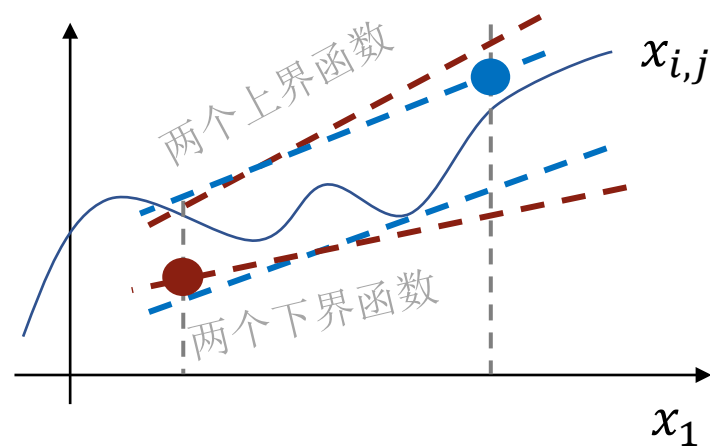


- 一维输入 (x_1) 示例 (此时每个节点 $x_{i,j}$ 是关于 x_1 的一维曲线)

传统界限传播

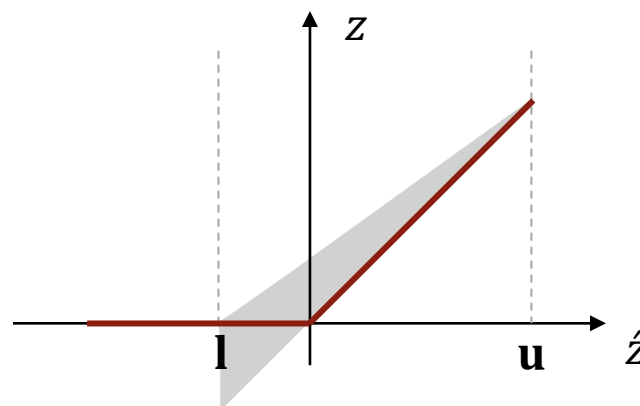
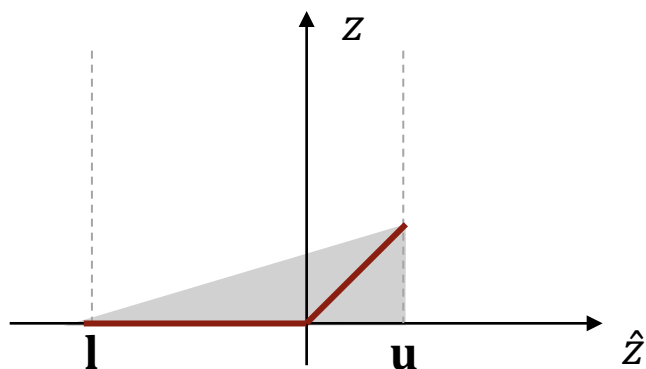


双路径界限传播



- 更好的上下界 \rightarrow 更准确的输出范围 \rightarrow 更高的验证精度

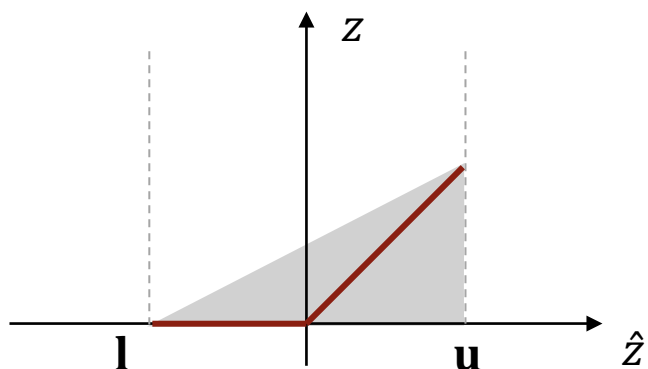
- DeepPoly / CROWN 路径 (baseline, 1-path)



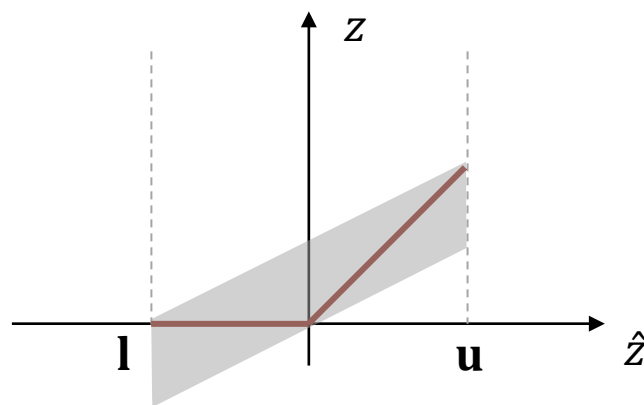
$$z \leq \frac{u}{u - l} (\hat{z} - l)$$

$$z \geq \alpha \hat{z} \quad \alpha = 0 \text{ if } u \leq |l| \text{ else } 1$$

- DeepPoly / CROWN 路径 (baseline, 1-path)
- 零下界 (2-path), 平行下界 (3-path), 以及 ...

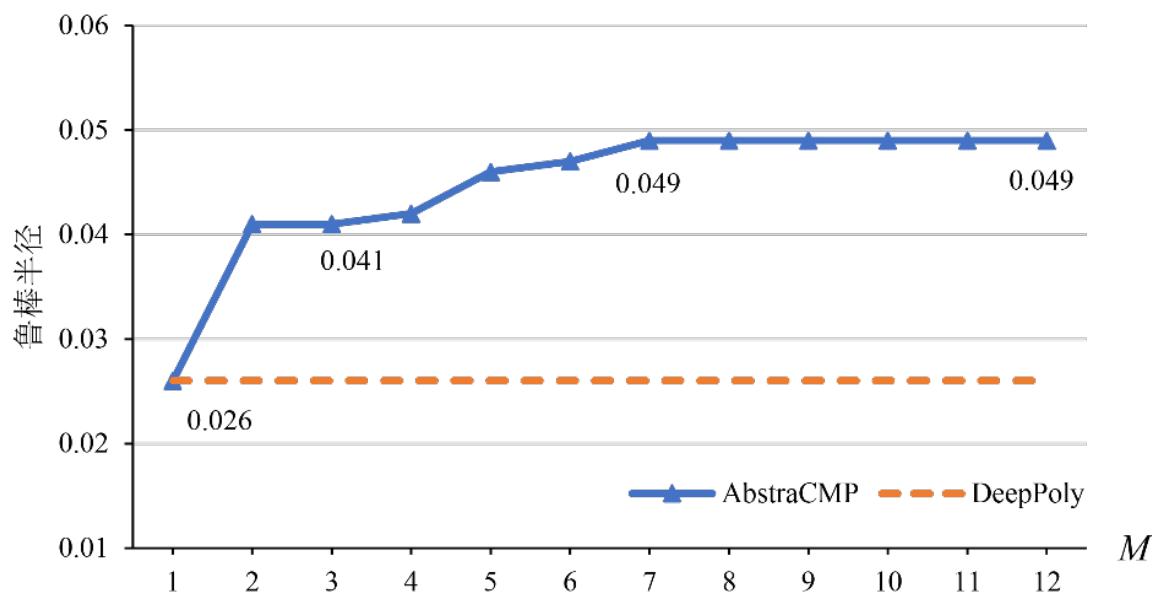


$$0 \leq z \leq \frac{u}{u-l} (\hat{z} - l)$$



$$\frac{u}{u-l} \hat{z} \leq z \leq \frac{u}{u-l} (\hat{z} - l)$$

- DeepPoly / CROWN 路径 (baseline, 1-path)
- 零下界 (2-path), 平行下界 (3-path), 以及 ...
- 理论上路径数目越多越好, 但提升并非可持续的


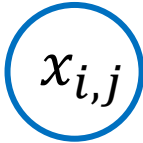


- 验证安全的数目对比，越大表示精度越高（黑体标出）

工具		模型和扰动大小 δ			
		MNIST FFNN			
		0.0014	0.0018	0.0022	0.0026
FBP	→ MpBP	73	62	51	40
	LiRPA	69	59	48	33
FBBP	→ MpBP	86	78	69	58
	LiRPA	83	77	66	56
		CIFAR-10 CNN		Tiny ImgNet CNN	
		0.0010	0.0014	0.0010	0.0014
BBP	→ MpBP	61	38	27	22
	LiRPA	56	36	25	19
	GPUPoly	56	36	-	-

- 提出界限传播路径的概念
- 将界限传播方法扩展到多路径界限传播
 - 现有方法是单条传播路径的特例
 - 包括反向界限传播、前向、前向+反向等
 - 形式化说明了上述方法，理论上证明其可靠性和精度优势
- 实验上说明其验证精度提升（相对于 SOTA）

- 每条路径之间独立，可并行化
- 使得多路径界限传播的时间代价降低到与单条路径相当
- 在 PyTorch 上实现为 MpBP 工具

PyTorch	MpBP
	
点值	界限函数 数值上下界

- 界限函数表示为高维张量

(batch 大小, 路径数目, 本层节点数, 输入层节点数)

- 界限函数表示为高维张量

(batch 大小, 路径数目, 本层节点数, 输入层节点数)

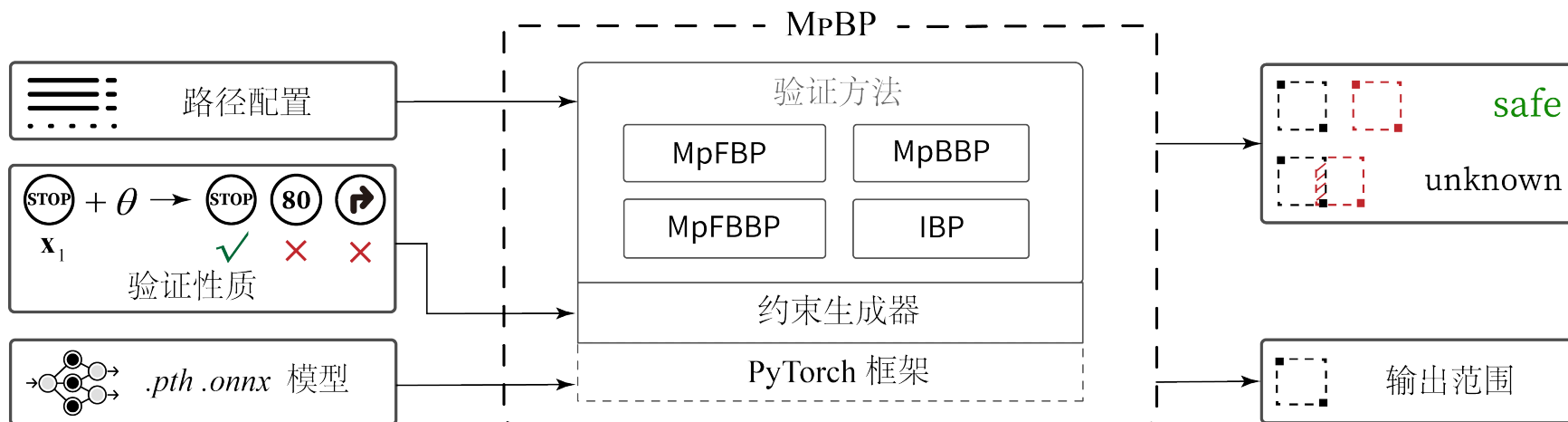
Tensor 1

- 界限函数逐层传播

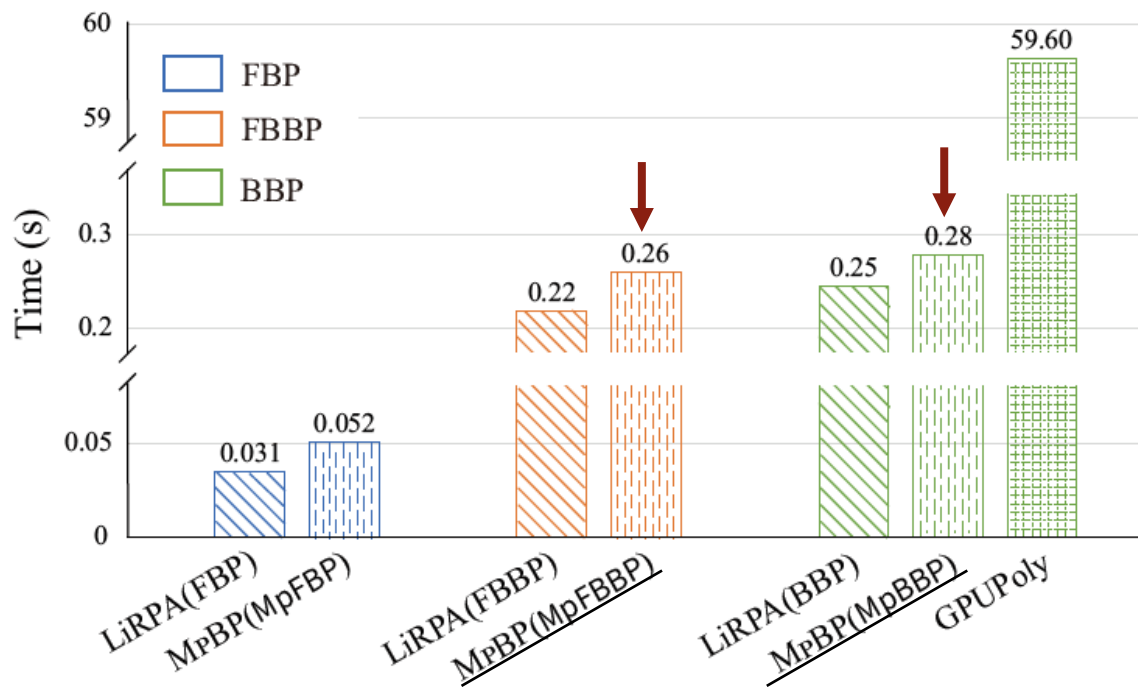
Tensor 1 × Tensor 2 × Tensor 3

- 乘法: 沿用 PyTorch 的高效张量乘法 (包括 CUDA 并行)

- 基于 GPU 并行的多路径界限传播工具
 - 高效验证、支持 CNN 网络结构、类 PyTorch 用法
 - 完善“训练-验证”流程



时间消耗对比



精度更高

&

时间相当

- 提出界限传播路径的概念，将界限传播方法扩展到多路径界限传播
- 并行化多条路径，开发了领先于 SOTA 的工程实现
- 两个开源工具（ AbstraCMP 和 MpBP ） *
- 两篇已发表论文
 - 郑焯，施晓牧，刘嘉祥. 基于多路径回溯的神经网络验证方法. 软件学报（CCF-T1）
 - Ye Zheng, Jiayang Liu, and Xiaomu Shi. MpBP: Verifying Robustness of ... FSE（CCF-A）

多路径方法在神经网络验证中的研究与应用

神经网络验证

深圳大学
SHENZHEN UNIVERSITY

- 验证给定的输入集合是否导致不安全输出
- 输入：无穷集（包含扰动阈值内的所有图片）
- 输出：安全或不安全（需要计算无穷集输入下的输出范围）

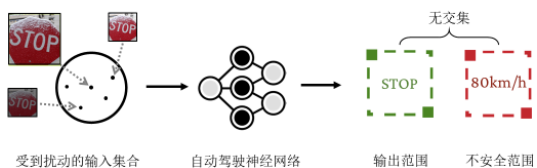


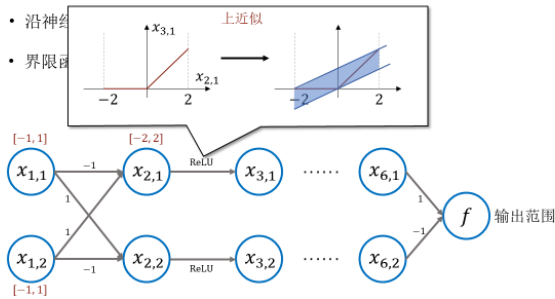
Image source: <https://www.businessinsider.com/why-are-stop-signs-red>

郑烽 多路径方法在神经网络验证中的研究与应用

4

界限传播方法

深圳大学
SHENZHEN UNIVERSITY

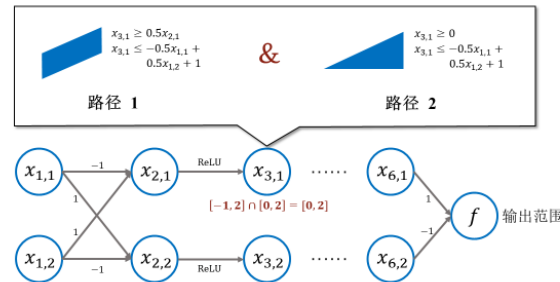


郑烽 多路径方法在神经网络验证中的研究与应用

8

本文：双路径界限传播

深圳大学
SHENZHEN UNIVERSITY



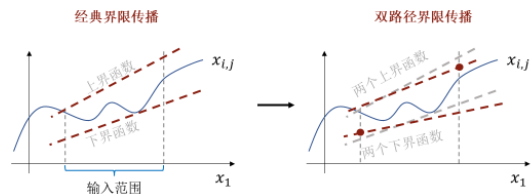
郑烽 多路径方法在神经网络验证中的研究与应用

13

本文：双路径界限传播

深圳大学
SHENZHEN UNIVERSITY

- 一维输入 (x_1) 示例（此时每个节点 $x_{i,j}$ 是关于 x_1 的一维曲线）



- 双路径界限传播为每个节点 $x_{i,j}$ 计算两组上下界函数，可得到更紧的上下界，提高验证精度

郑烽 多路径方法在神经网络验证中的研究与应用

15

GPU 并行化

深圳大学
SHENZHEN UNIVERSITY

- 界限函数表示为高维张量
(batch 大小, 路径数目, 本层节点数, 输入层节点数)

Tensor 1

- 界限函数逐层传播

Tensor 1 × Tensor 2 × Tensor 3 ……

- 乘法：沿用 PyTorch 的高效张量乘法（包括 CUDA 并行）

郑烽 多路径方法在神经网络验证中的研究与应用

25

工作量

深圳大学
SHENZHEN UNIVERSITY

- 提出界限传播路径的概念，将界限传播方法扩展到多路径界限传播
- 并行化多条路径，开发了领先于 SOTA 的工程实现
- 两个开源工具（AbstraCMP 和 MpBP）*
- 两篇已发表论文
 - 郑烽, 施晓牧, 刘嘉祥. 基于多路径回溯的神经网络验证方法. 软件学报 (CCF-T1)
 - Ye Zheng, Jiaxiang Liu, and Xiaomu Shi. MpBP: verifying robustness of ... FSE (CCF-A)

https://github.com/formes20

郑烽 多路径方法在神经网络验证中的研究与应用

28

Thank you!

