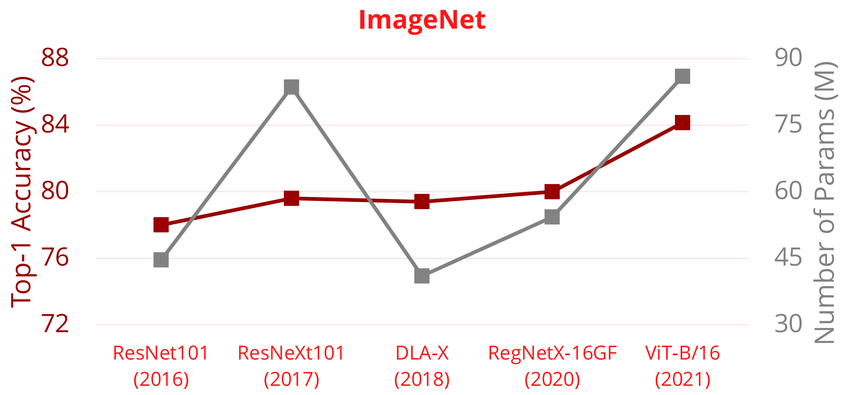
How To Accelerate Training Certifiably Robust Neural Networks

Pratik Vaishnavi[1], Kevin Eykholt[2], Amir Rahmati[1]
[1]Stony Brook University, [2]IBM Research

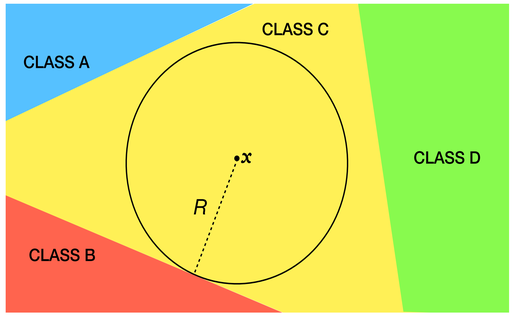
Stony Brook University
Computer Science

GitHub

NNs Are Continuously Evolving!



We Need Certified Robustness for Critical Applications



Decision boundary of a NN that is certifiably robust at x within neighborhood of radius R. This radius is called **certified radius**.

Training Certifiably Robust NNs Is Slow

Method	Training Slowdown Factor
SmoothAdv (2019)	46.20×
MACER (2020)	20.86×
SmoothMix (2021)	4.97×

- Preserving certified robustness across generations of NN architectures using existing methods result in high computational costs.

Certified Robustness Transfer (CRT)

Problem: Approximating certified radius R during training is slow.

Solution: (Knowledge Transfer) Indirectly maximize R by matching outputs with a certifiably robust teacher.

Benefit: Process of matching outputs adds **negligible overhead** to the training process.

CRT Training Objective:

$$\mathbb{E}_{\eta\sim\mathcal{N}(0,\sigma^2 I);\, x\sim\mathcal{D}}\left[z_\phi(x+\eta) - z_\theta(x+\eta)\right]$$

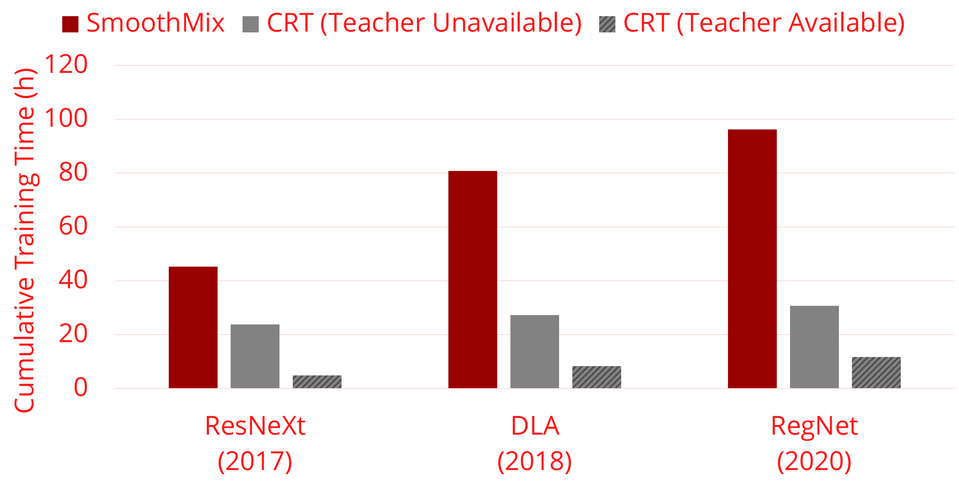


Figure 1: Cumulative time for training newly released NNs. **CRT significantly speeds up the process of training new certifiably robust NNs**, whether a pre-trained teacher is available or not.

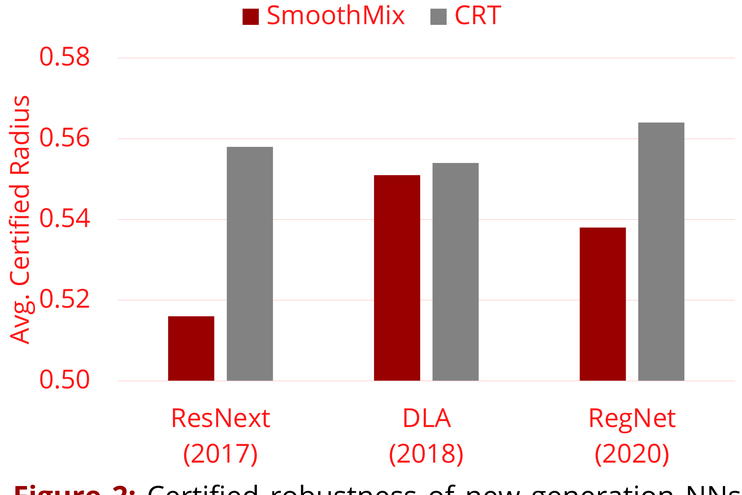


Figure 2: Certified robustness of new generation NNs. CRT trained NNs consistently exhibit **higher robustness** than their SmoothMix counterparts despite large generation gap with teacher (ResNet).

Method	Training Time (h)	ACR
SmoothMix	18.98	0.550
CRT (ResNet20 Teacher)	10.07	0.540

Table 1: When a **pre-trained teacher (ResNet) is not available**, we show that **CRT can also be used to accelerate the process of acquiring one** by using a smaller sized network as a proxy teacher. This speedup is achieved while **preserving certified robustness**.