DEEL

DEpendable & Explainable Learning

PAY ATTENTION TO YOUR LOSS:

UNDERSTANDING MISCONCEPTIONS ABOUT 1-LIPSCHITZ NEURAL NETWORKS

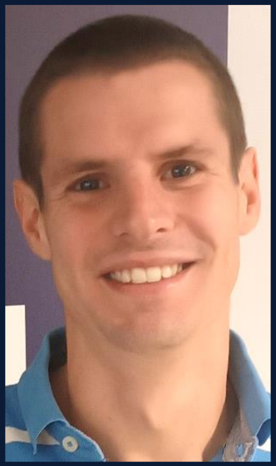
IVADO · IRT Saint Exupéry · CRIAQ · ANITI Université Fédérale Toulouse Midi-Pyrénées · Institut intelligence et données

DEEL

DEpendable & Explainable Learning

Louis
BETHUNE
louis.bethune@univ-toulouse.fr

Thibaut
BOISSIN
thibaut.boissin@irt-saintexupery.com



Corentin
FRIEDRICH
corentin.friedrich@irt-saintexupery.com

Alberto
GONZALEZ-SANZ
alberto.gonzalez_sanz@math.univ-toulouse.fr

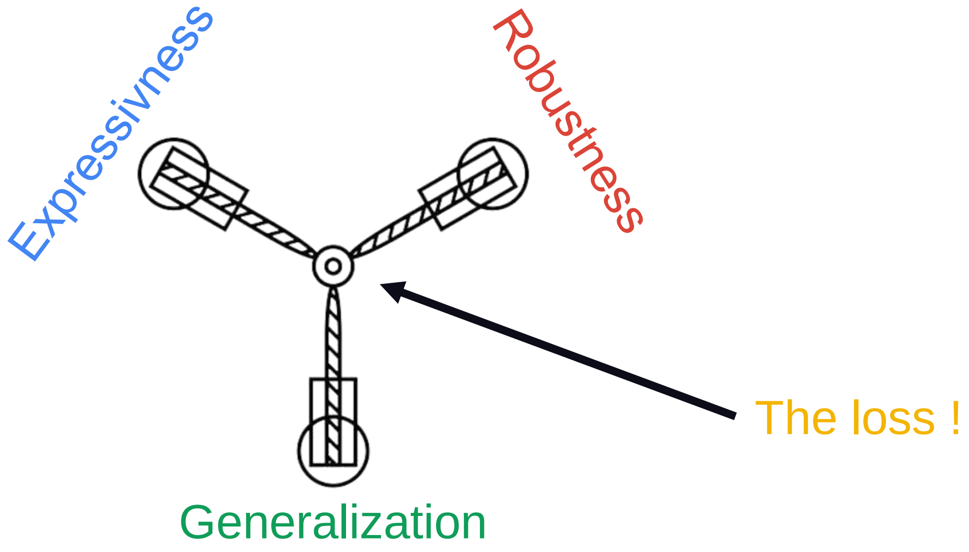
Franck
MAMALET
franck.mamalet@irt-saintexupery.com

Mathieu
SERRURIER
mathieu.serrurier@irit.fr

IVADO SAINT EXUPÉRY CRIAQ ANITI

Outline

- LipNet1 are **expressive**
- LipNet1 are **provably robust**
- LipNet1 have **generalization guarantees**



WHAT ARE 1-LIPSCHITZ NETWORKS ?

DEEL
DEpendable & Explainable Learning

Jax version incoming!

DEELLIP
LIPSCHITZ KERAS LAYERS

Tensorflow

Pytorch

Implemented in practice with:

- GroupSort (MinMAx) activation functions
- Differentiable projection of weights onto Stiefel manifold

Lipschitz constant $L(f)$:

$$\|f(x) - f(z)\|_2 \leq L(f)\|x - z\|_2$$

Conventional networks can be made 1-Lipschitz:

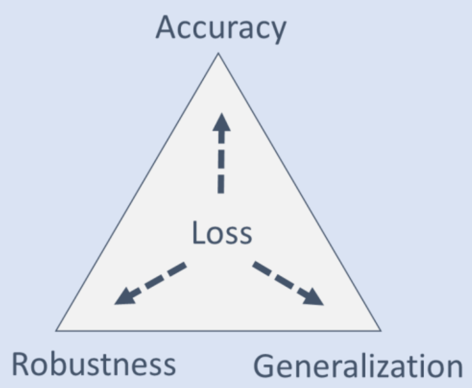
$$g^* = \arg \min_{g \in C(\mathcal{X}, \mathbb{R}^K)} \mathbb{E}_{x,y} \mathcal{L}(g(x), y) \qquad f^* = \frac{1}{L(g^*)} g^*$$

But $L(g^*)$ is often high, and finding it is NP-hard.

1-Lipschitz functions are approximated by constraining the weights of each layers. This is done in practice with **Deel-Lip** library:

$$f^* = \arg \min_{f \in \mathrm{Lip}_1(\mathcal{X}, \mathbb{R})} \mathbb{E}_{x,y} \mathcal{L}(f(x), y)$$

The choice of loss controls the tradeoff between accuracy, robustness and generalization.

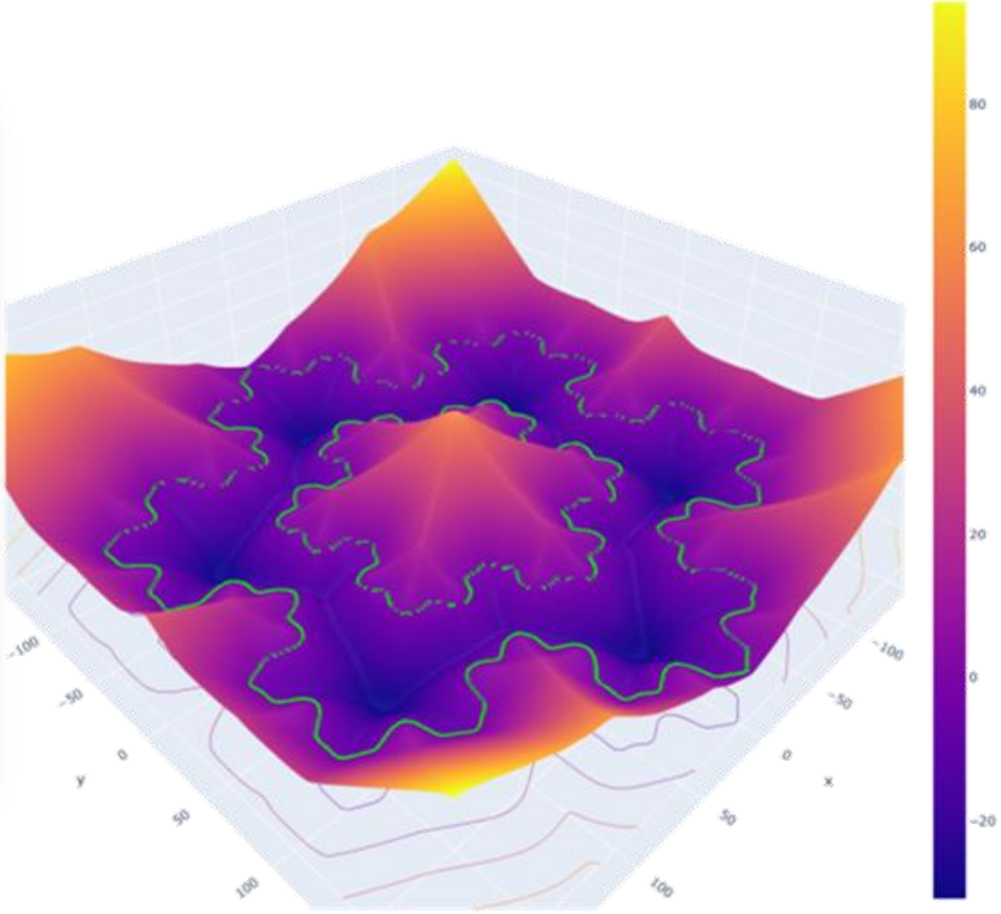


Why **1-Lipschitz networks** are **perceived as not expressive** ?

EXPRESSIVENESS

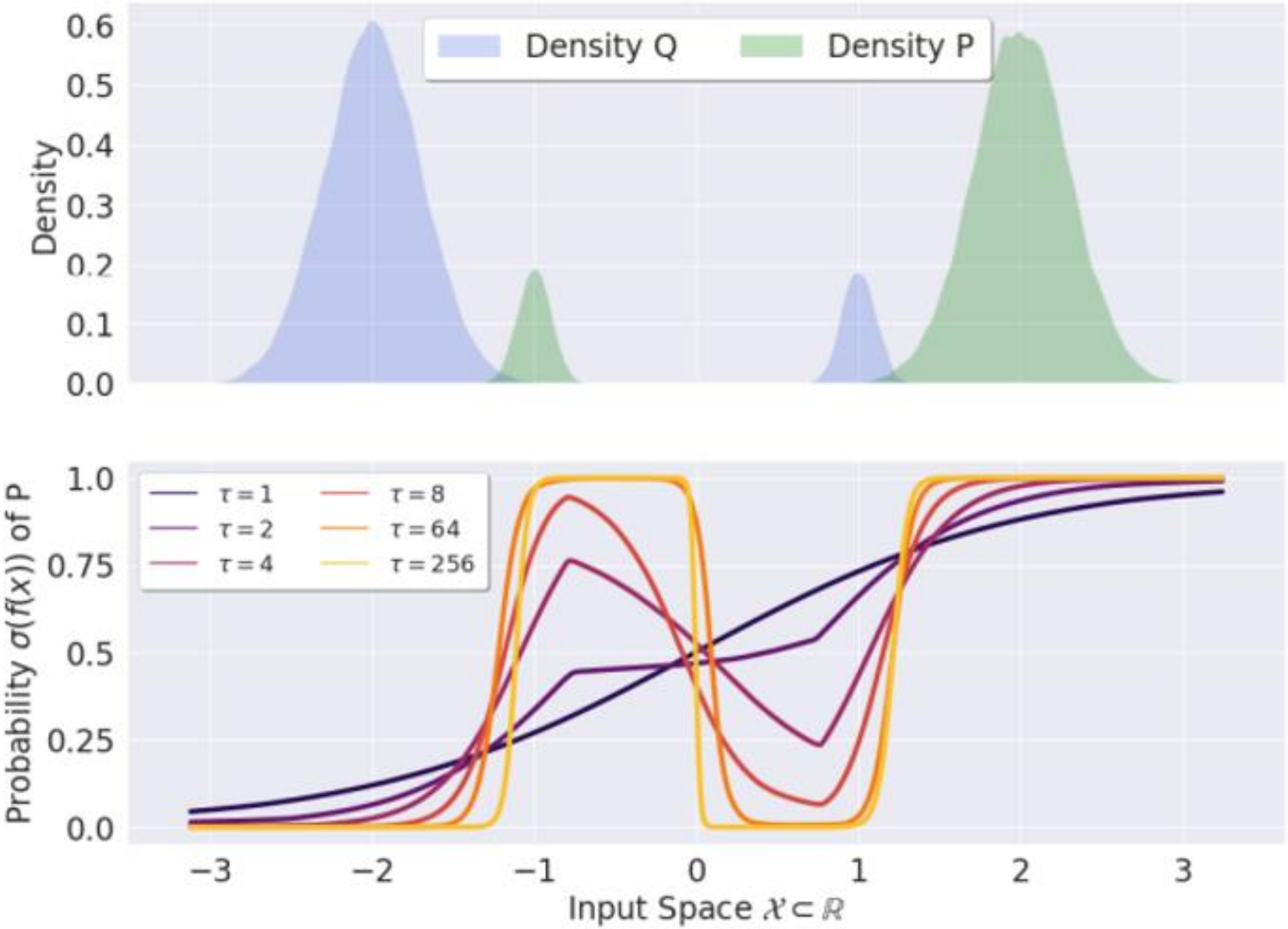
Proposition 1 (Lipschitz Binary classification). *For any binary classifier $c : \mathcal{X} \rightarrow \mathcal{Y}$ with closed pre-images ($c^{-1}(\{y\})$ is a closed set) there exists a 1-Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $sign(f(x)) = c(x)$ on $\mathcal{X}$ and such that $\|\nabla_x f\| = 1$ almost everywhere (w.r.t Lebesgue measure).*

In our experiment, 1-Lipschitz Network reached **99.9% accuracy** on Cifar-100 with **random labels.**



Fractal decision boundary in 2D with Von Koch snowflake.

EXPRESSIVENESS



Sigmoid Cross-entropy temperature tuning in 1D classification.

Default to 1. in most frameworks!

Why **1-Lipschitz Networks** are perceived as **more robust** ?

ROBUSTNESS

- Robustness certificates

$$\|f(x) - f(x+\delta)\| \leq L\|x - (x+\delta)\| \quad \forall x, \delta \tag{D.3}$$

$$\iff \|\delta\| \geq \frac{\|f(x) - f(x+\delta)\|}{L} \tag{D.4}$$

Positive even when x is misclassified

- MCR: mean certifiable robustness

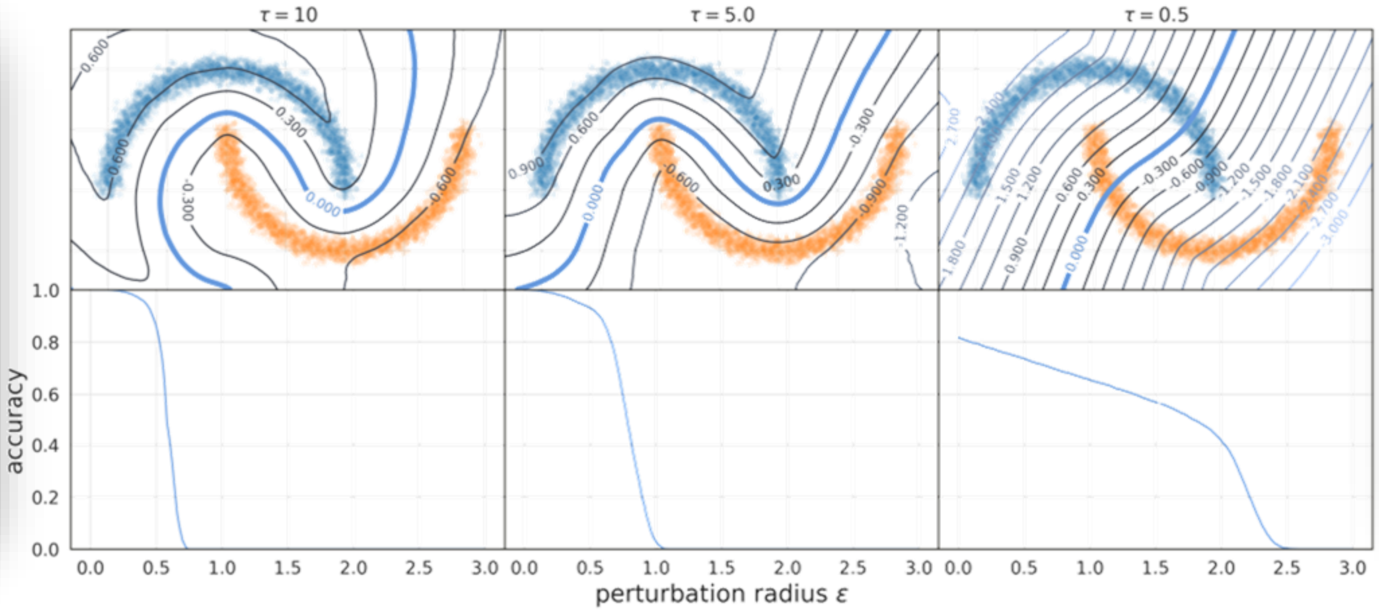
Negative when x is misclassified

$$\begin{aligned} \mathrm{MCR}_{XY}(f) &= \mathbb{E}_{x,y}[\mathbb{1}\{yf(x) > 0\}|f(x)|] \\ &+ \mathbb{E}_{x,y}[-\mathbb{1}\{yf(x) < 0\}|f(x)|] \\ &= \mathbb{E}_{x,y} yf(x) \end{aligned}$$

ROBUSTNESS

Accuracy ⟷ Robustness

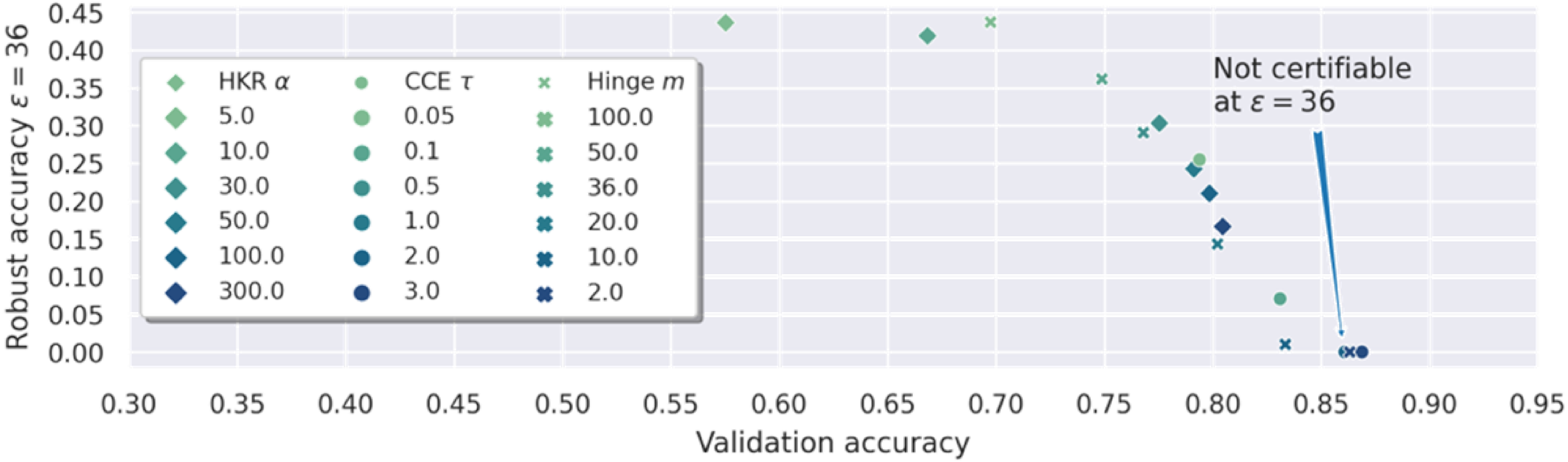
Highest accuracy and robustness certificates.
Corresponds to 1-nearest neighbor classification.



Classifiers with highest Mean Certifiable Robustness (MCR):

- Are WGAN discriminators,
- Have usually low accuracy,
- Correspond to low cross-entropy temperature.

CONTROL OF THE TRADEOFF

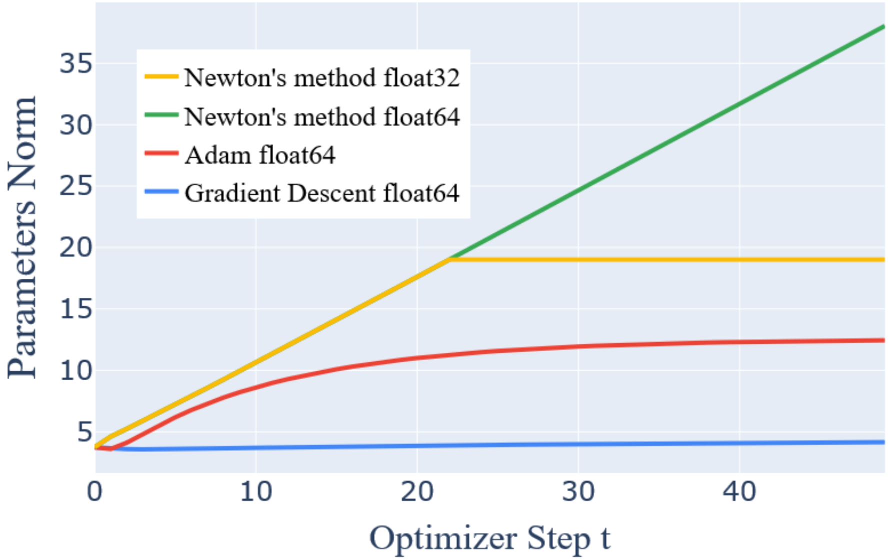


Moving along Accuracy-Robustness Pareto front
on Cifar-10 by tuning hyper-parameters of the loss.

Do they **generalize** well ?

GENERALIZATION

DEEL
DEpendable & Explainable Learning

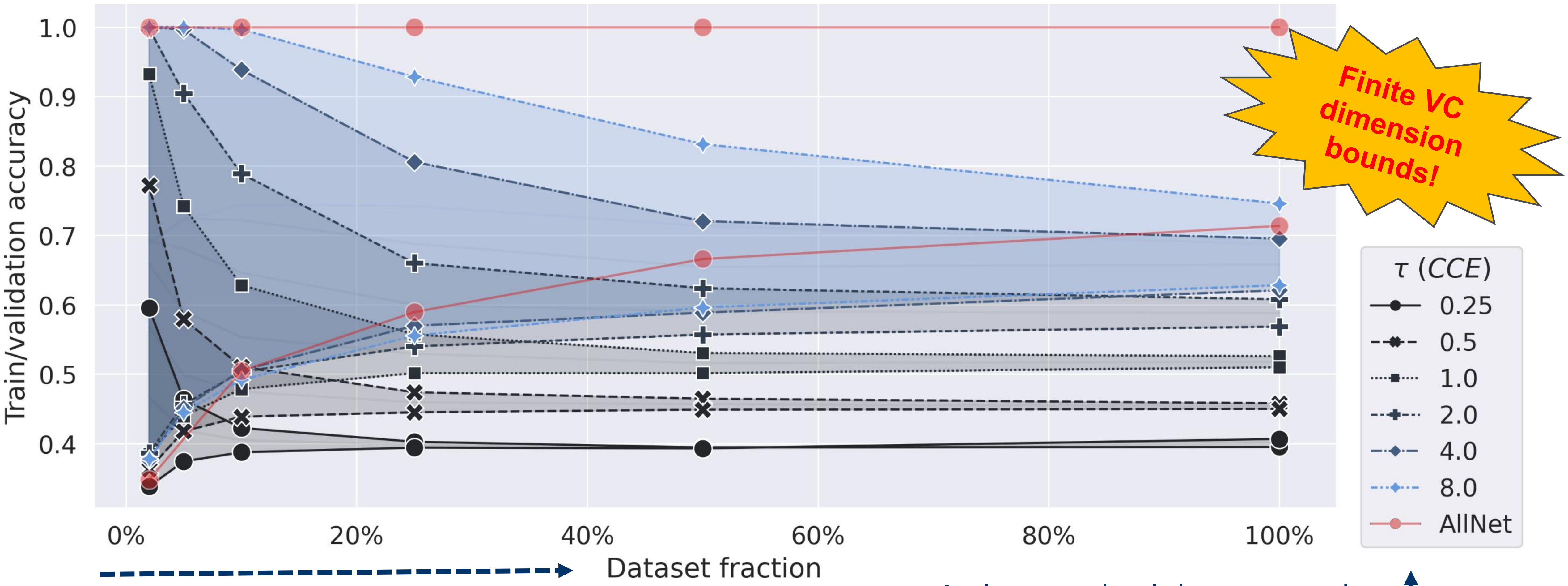


Theorem: 1-Lipschitz networks are ***consistent estimators***; adding more samples closes the train/test gap.

Conventional networks optimization leads to uncontrollable growth of their Lipschitz constant.

GENERALIZATION

DEEL
DEpendable & Explainable Learning



We trained the same network each time on a larger fraction of CIFAR10

And reported train/test accuracies for various loss parameters

Conclusions

Why should you care ?

1) If you want networks with robustness certificates, generalization guarantee, fine grained control of accuracy/robustness tradeoff.

2) If you want to understand training dynamic and generalization of conventional networks.

DEEL

DEpendable & Explainable Learning

THANK YOU !