Smale's 18th Problem and the Barriers of Deep Learning

Matthew Colbrook
(University of Cambridge and École Normale Supérieure)

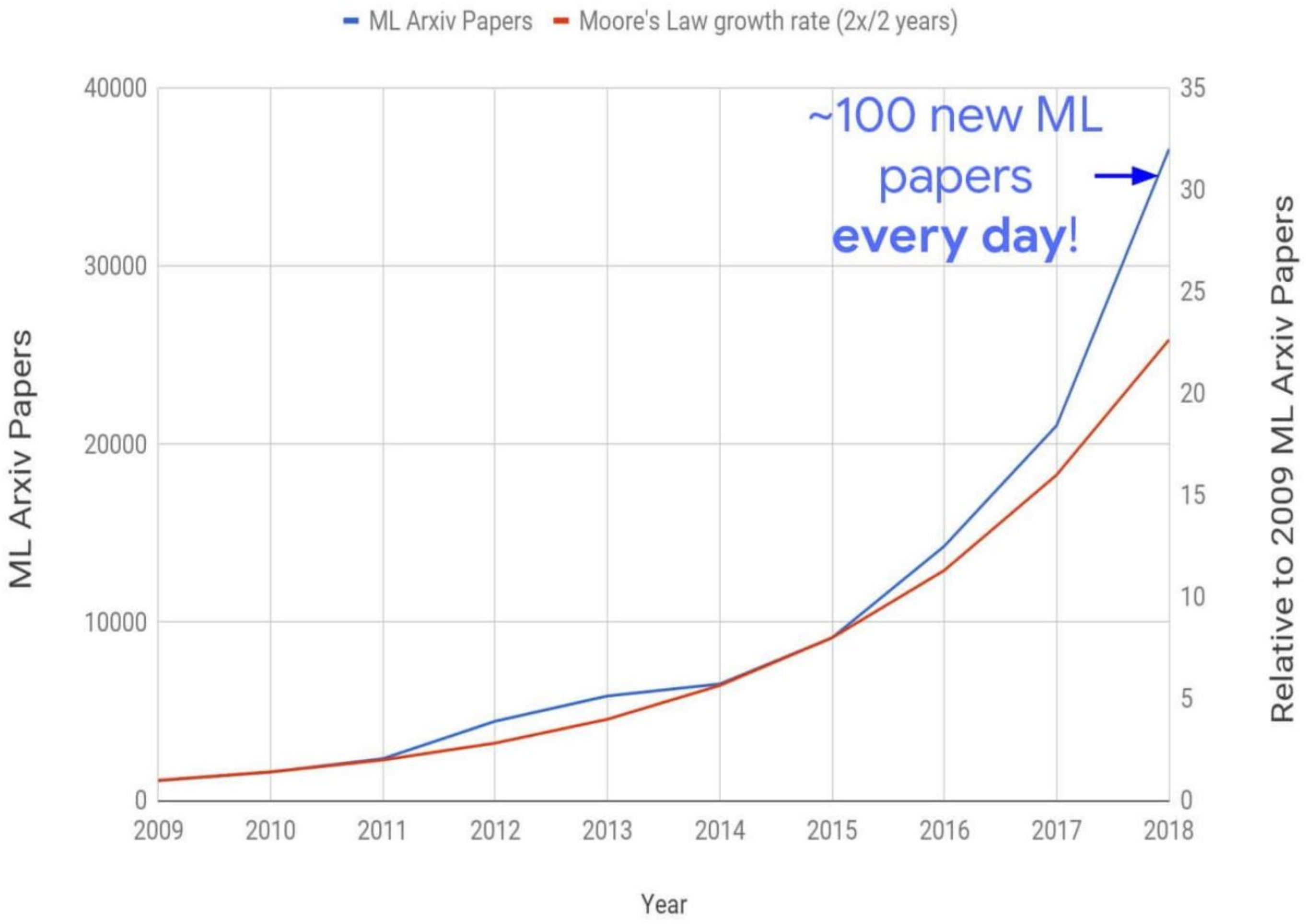
Smale's 18th problem*: *What are the limits of artificial intelligence?*

M. Colbrook, V. Antun, A. Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem*" (PNAS, 2022)

*Steve Smale's list of problems for the 21st century (requested by Vladimir Arnold), inspired by Hilbert's list.

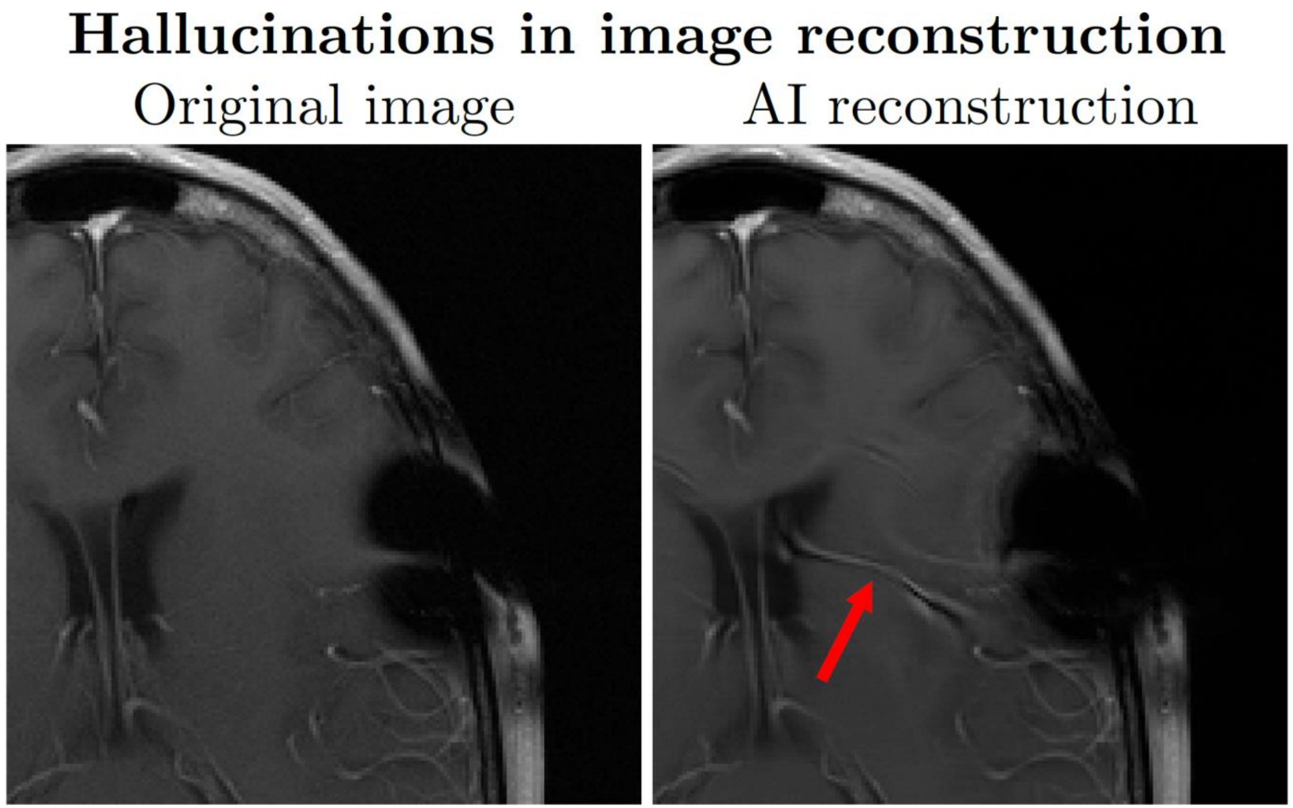
http://www.damtp.cam.ac.uk/user/mjc249/home.html: slides, papers, and code

A fun stat!

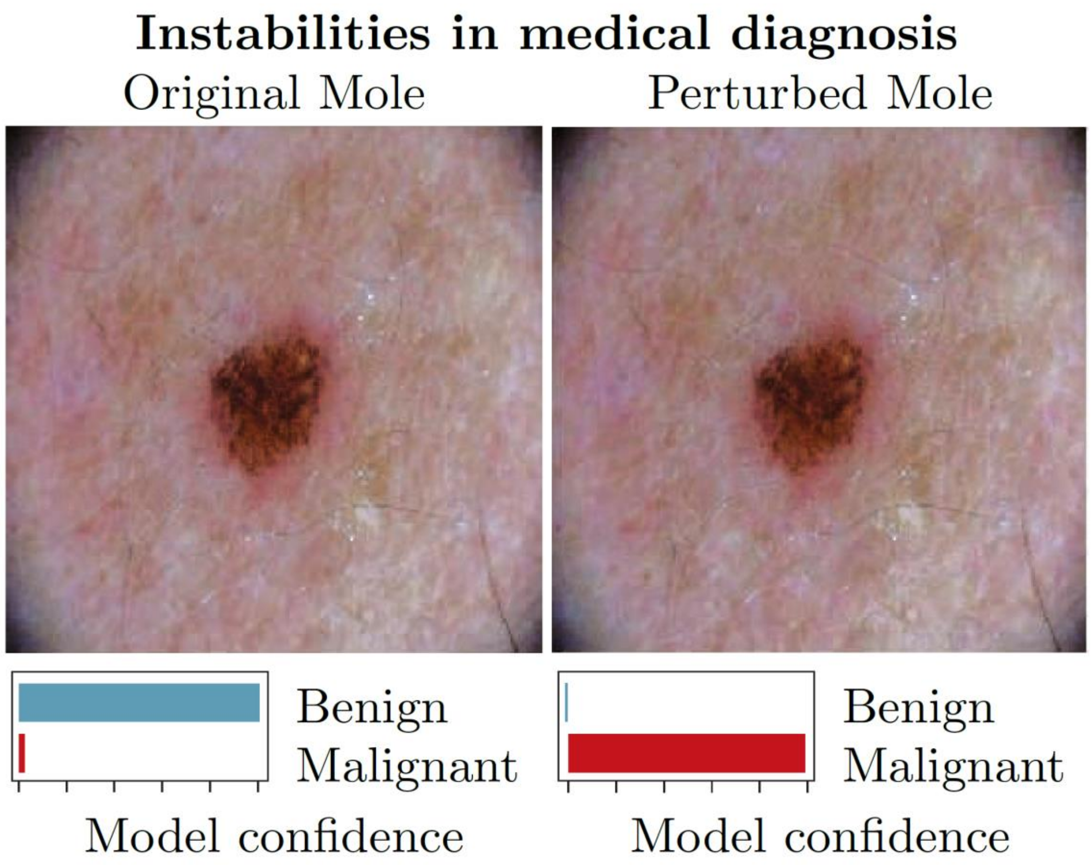


To keep up during first lockdown, would need to continually read a paper every 4 mins!

Problem: hallucinations and instability



"**AI generated hallucination**", from Facebook and NYU's ***FastMRI challenge*** 2020



From Finlayson et al., "*Adversarial attacks on medical machine learning*," **Science**, 2019.

When can we make AI robust and trustworthy?

Smale's 18th problem: “What are the limits of AI?”

“Very often, the creation of a technological artifact precedes the science that goes with it. The steam engine was invented before thermodynamics. Thermodynamics was invented to explain the steam engine, essentially the ***limitations*** *of it.* ***What we are after is the equivalent of thermodynamics for intelligence.****”* Yann LeCun

“2021 was the year in which the wonders of artificial intelligence stopped being a story. Many of this year's top articles grappled with the ***limits of deep learning*** *(today's dominant strand of AI).”*

IEEE Spectrum, 2021's Top Stories About AI (Dec. 2021)

Example of the limits of deep learning

Paradox: "Nice" linear inverse problems where a *stable* and *accurate* neural network for image reconstruction exists, but it can never be trained!

E.g., suppose we want to solve (holds for much more general problems)

$$\min_{x \in \mathbb{C}^N} \|x\|_{l^1} + \lambda \|Ax - y\|_{l_2}^2$$

$$A \in \mathbb{C}^{m \times N} \text{ (modality}, m < N), \qquad S = \{y_j\}_{j=1}^{R} \text{ (samples)}$$

Arises when given $y \approx Ax + e$.

Allow arbitrary precision of training data.

Enforce condition numbers bounded by 1.

Example of the limits of deep learning

Paradox: "Nice" linear inverse problems where a *stable* and *accurate* neural network for image reconstruction exists, but it can never be trained!

Theorem: Pick positive integers $n \geq 3$ and M. Class of problems such that:

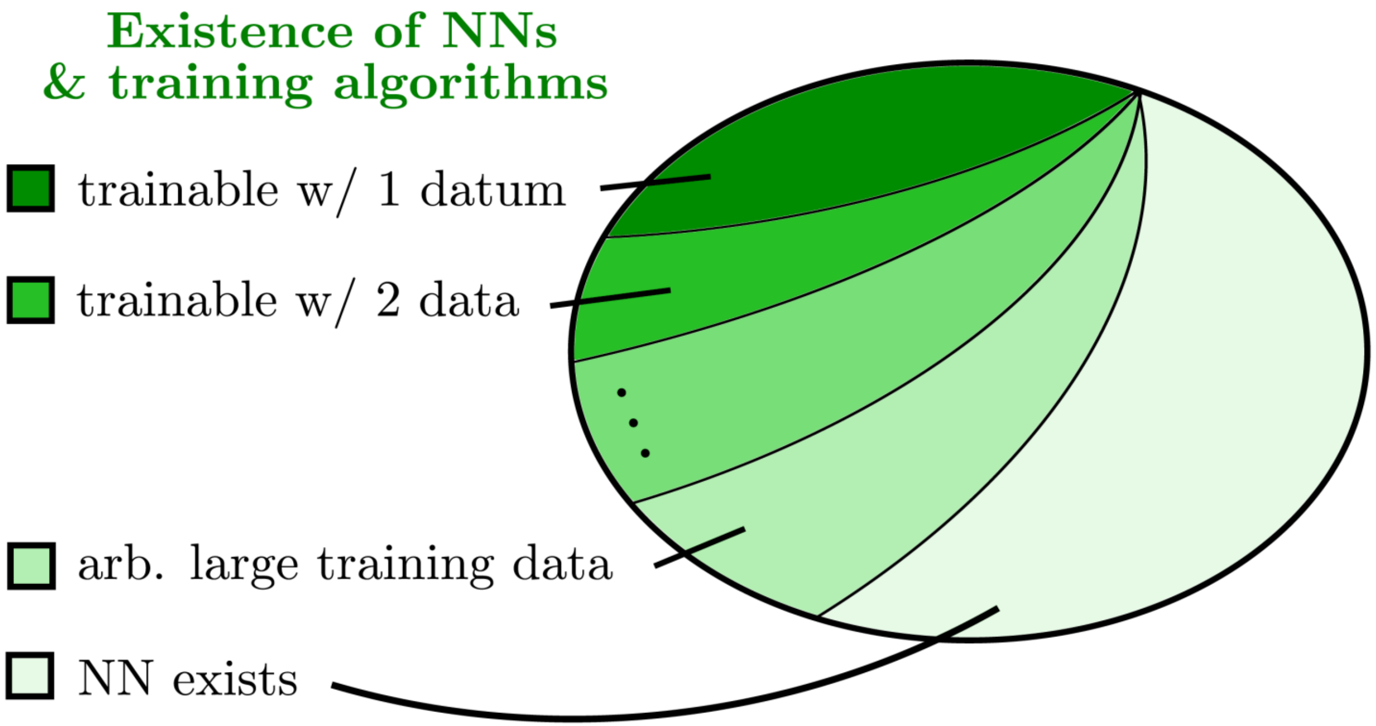
- (**Not trainable**) No algorithm (even random) can train a neural network with $\boldsymbol{n}$ **digits** of accuracy over the dataset with probability greater than 1/2.
- (**Not practical**) $\boldsymbol{n-1}$ **digits** of accuracy possible over the dataset, but any training algorithm requires **arbitrarily large training data**.
- (**Trainable and practical**) $\boldsymbol{n-2}$ digits of accuracy possible over the dataset via training algorithm using $\boldsymbol{M}$ **training data**.

Holds for any architecture, any precision of training data.

$\Longrightarrow$ **Classification theory telling us what can and cannot be done**

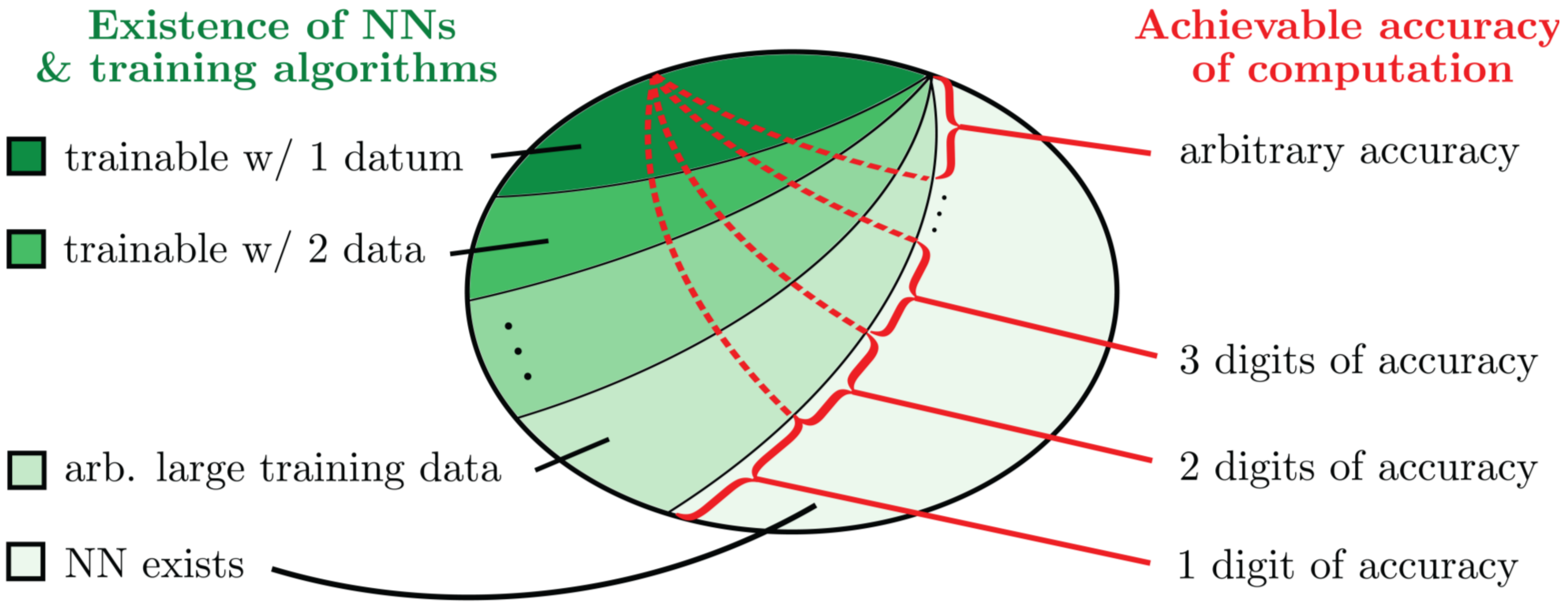
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.
- Antun, C., Hansen, *"Proving Existence Is Not Enough: : Mathematical Paradoxes Unravel the Limits of Neural Networks in Artificial Intelligence,"* **SIAM News**, May 2022.
- Choi, *"Some AI Systems May Be Impossible to Compute,"* **IEEE Spectrum**, March 2022.

The world of neural networks



Given a problem and conditions, where does it sit in this diagram?

The world of neural networks



Given a problem and conditions, where does it sit in this diagram?

Example counterpart theorem

Certain conditions: stable neural networks trained with exponential accuracy.
E.g., *approximate Łojasiewicz-type inequality*:

$$(1) \quad \min_{x\in\mathbb{C}^N} f(x) \quad \text{s.t.} \quad \|Ax-y\| \leq \varepsilon$$

$$\mathrm{dist}(x, \mathrm{solution}) \leq \alpha([f(x)-f^*] + [\|Ax-y\|-\varepsilon] + \delta)$$

Fast **I**terative **RE**started **NET**works (FIRENETs)
(unrolled primal-dual with novel restart scheme)

Theorem: Training algorithm that, under above assumption, produces *stable* neural networks φ_n of width $O(N)$, depth $O(n)$, guaranteed worst bound

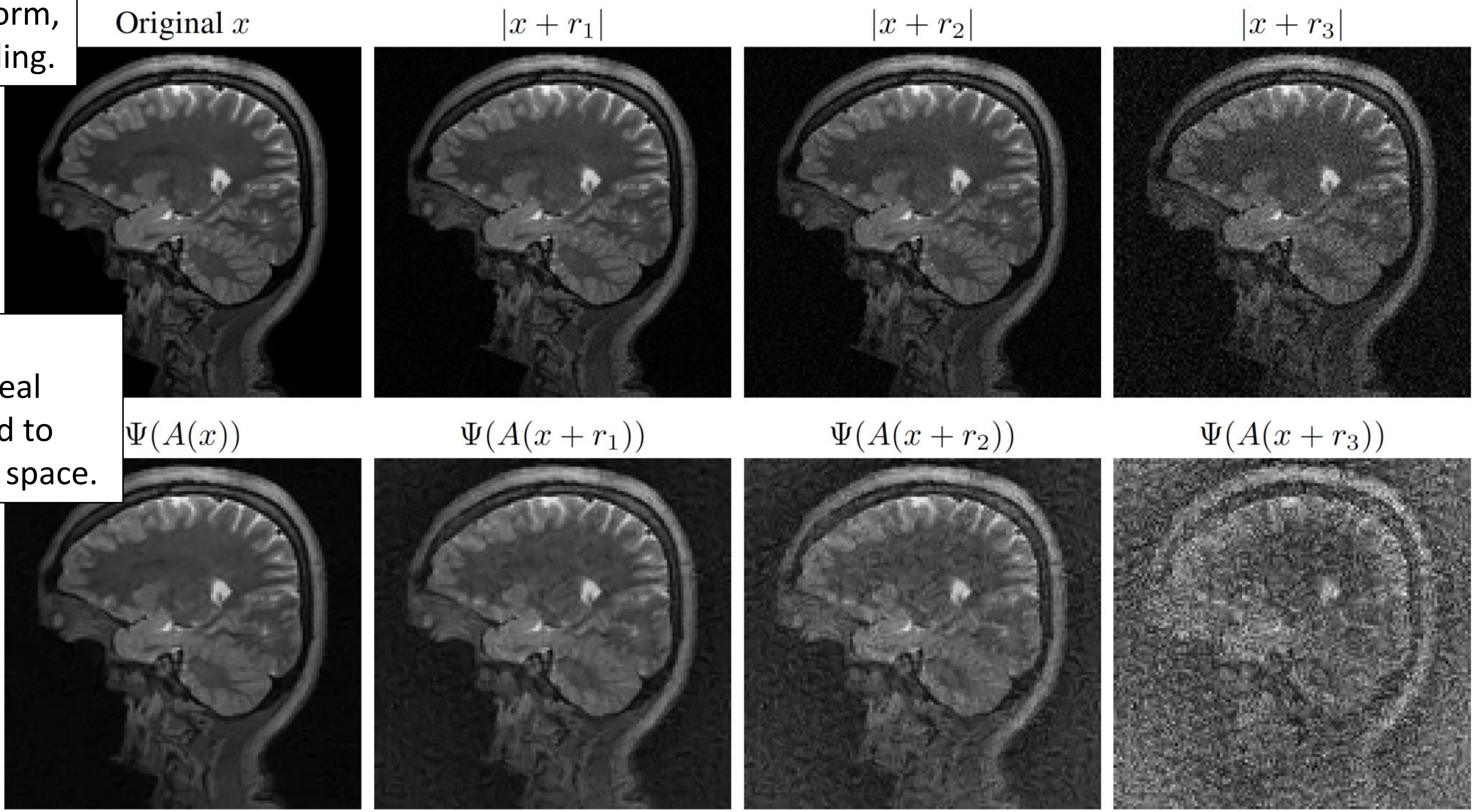
$$\mathrm{dist}(\varphi_n(y), \mathrm{solution}) \lesssim e^{-n} + \delta$$

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.
- C., "*WARPd: A linearly convergent first-order method for inverse problems with approximate sharpness conditions,*" **SIIMS**, 2022.

Example of severe instability

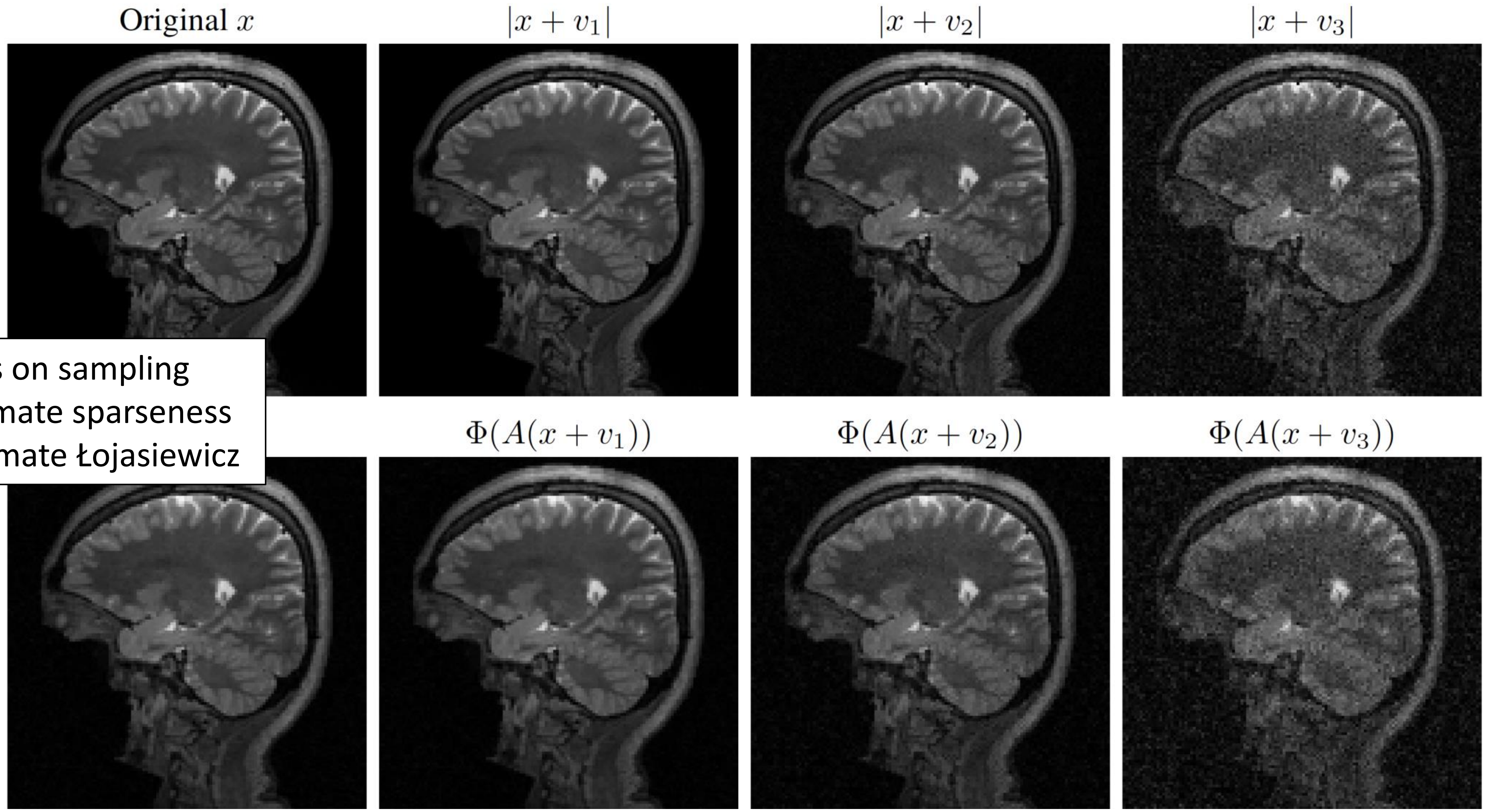
MRI: discrete 2D Fourier transform, 60% subsampling.

Perturbations computed in real space, mapped to measurement space.



- Zhu et al., "*Image reconstruction by domain-transform manifold learning,*" **Nature,** 2018.
- Antun et al., "*On instabilities of deep learning in image reconstruction and the potential costs of AI,*" **PNAS**, 2020.

FIRENET: provably stable (even to adversarial examples) and accurate

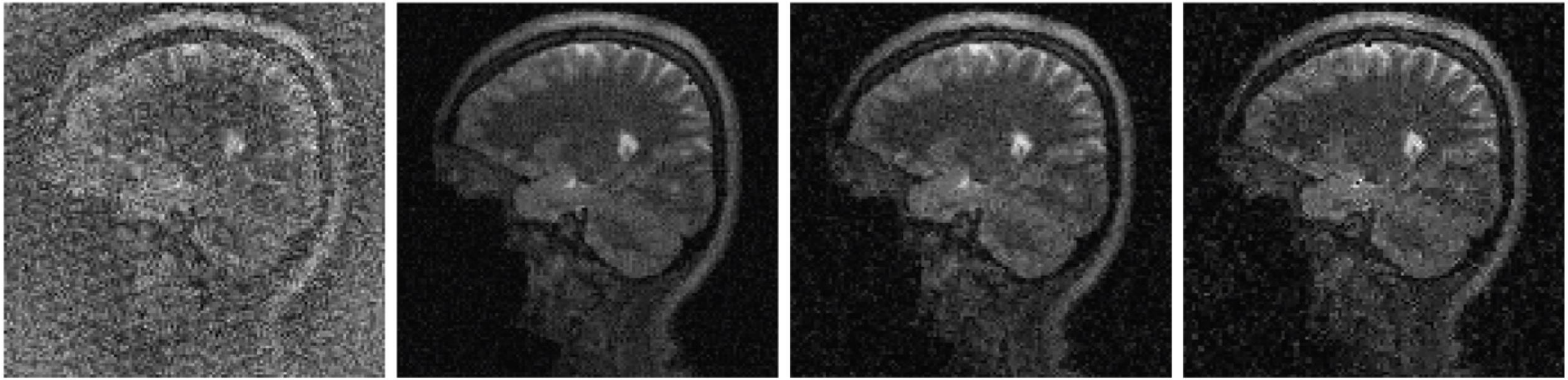


Assumptions on sampling and approximate sparseness give approximate Łojasiewicz

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

Stabilising unstable neural networks

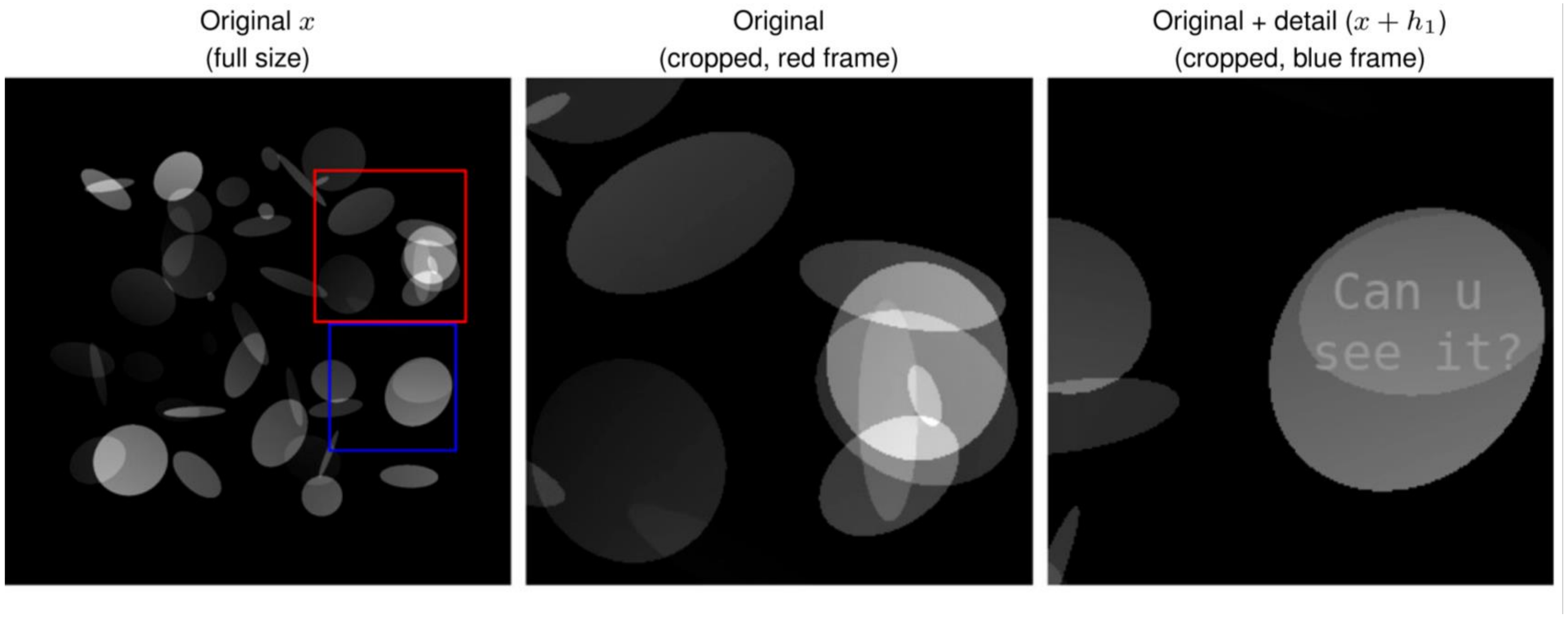
$\Psi(\tilde{y}),\ \tilde{y} = Ax + e_3$ | $\Phi\left(\tilde{y}, \Psi(\tilde{y})\right)$ | FIRENET rec. from $y = Ax + \tilde{e}_3$ | AUTOMAP+FIRENET rec. from $y = Ax + \hat{e}_3$



Key pillars: stability and accuracy

MRI: discrete 2D Fourier transform, 15% subsampling.

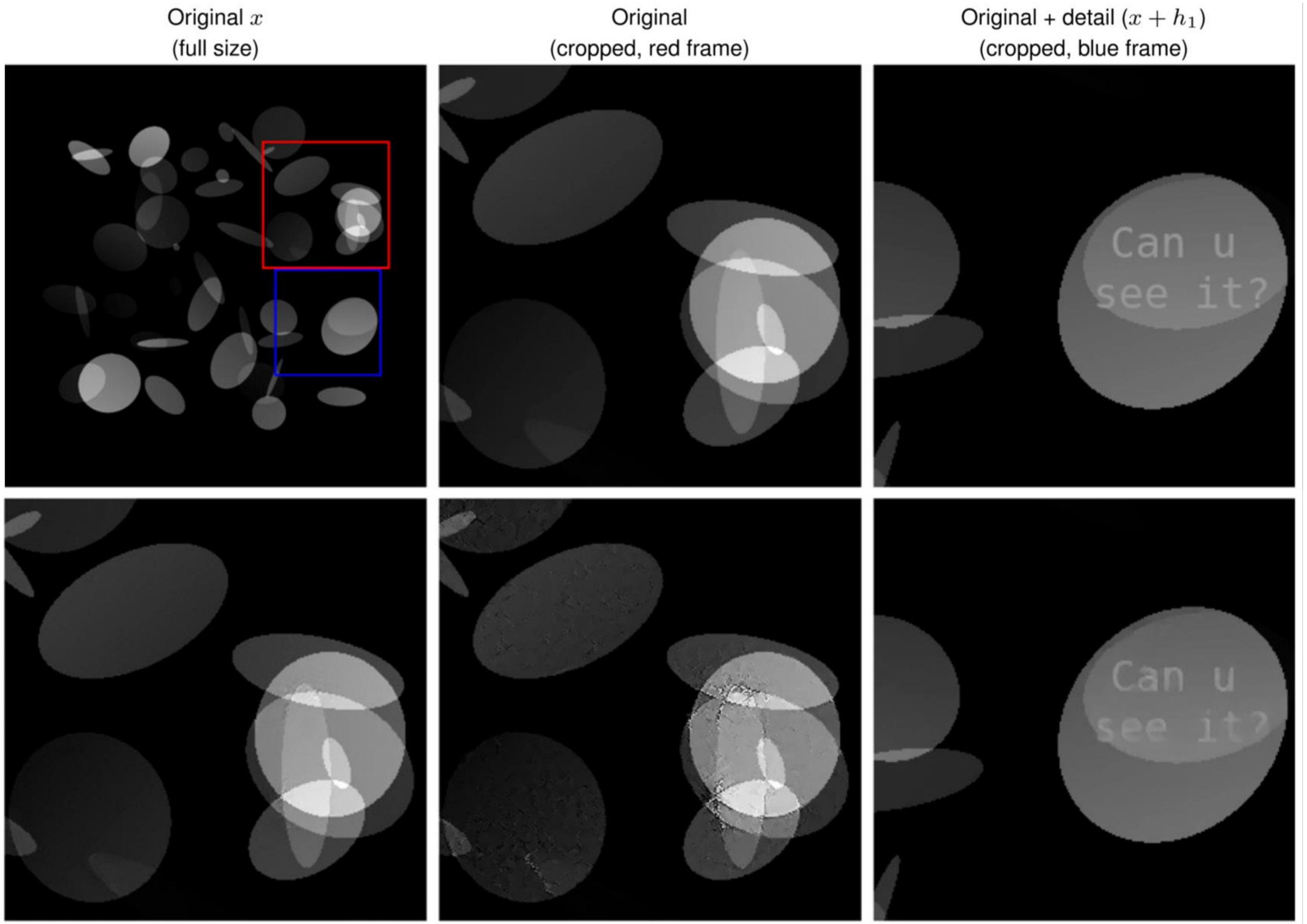
All networks trained on 5000 images of ellipses



- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

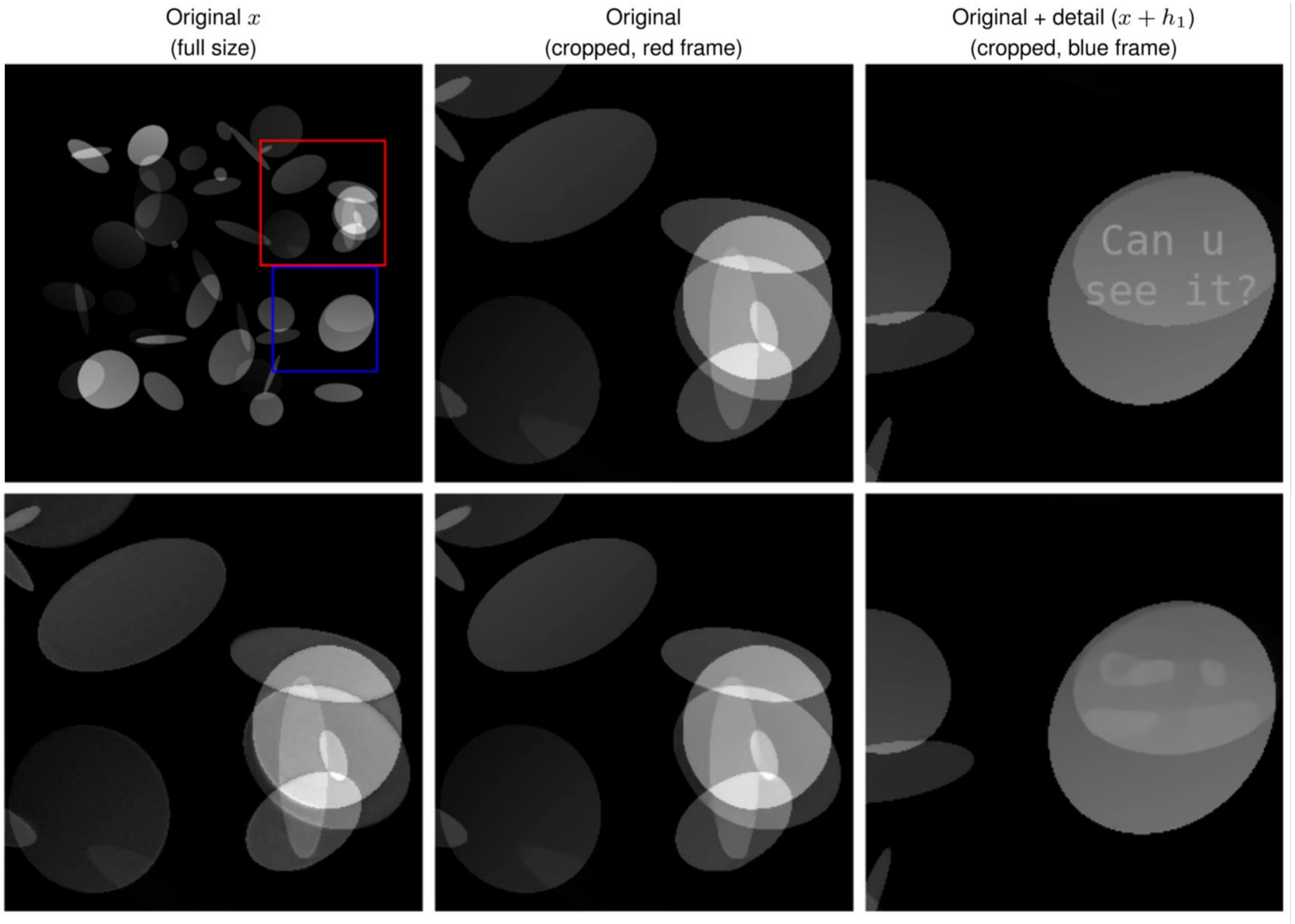
U-Net with no noise: accurate but unstable

U-Net: standard neural network architecture for imaging. Approx 4 million parameters.



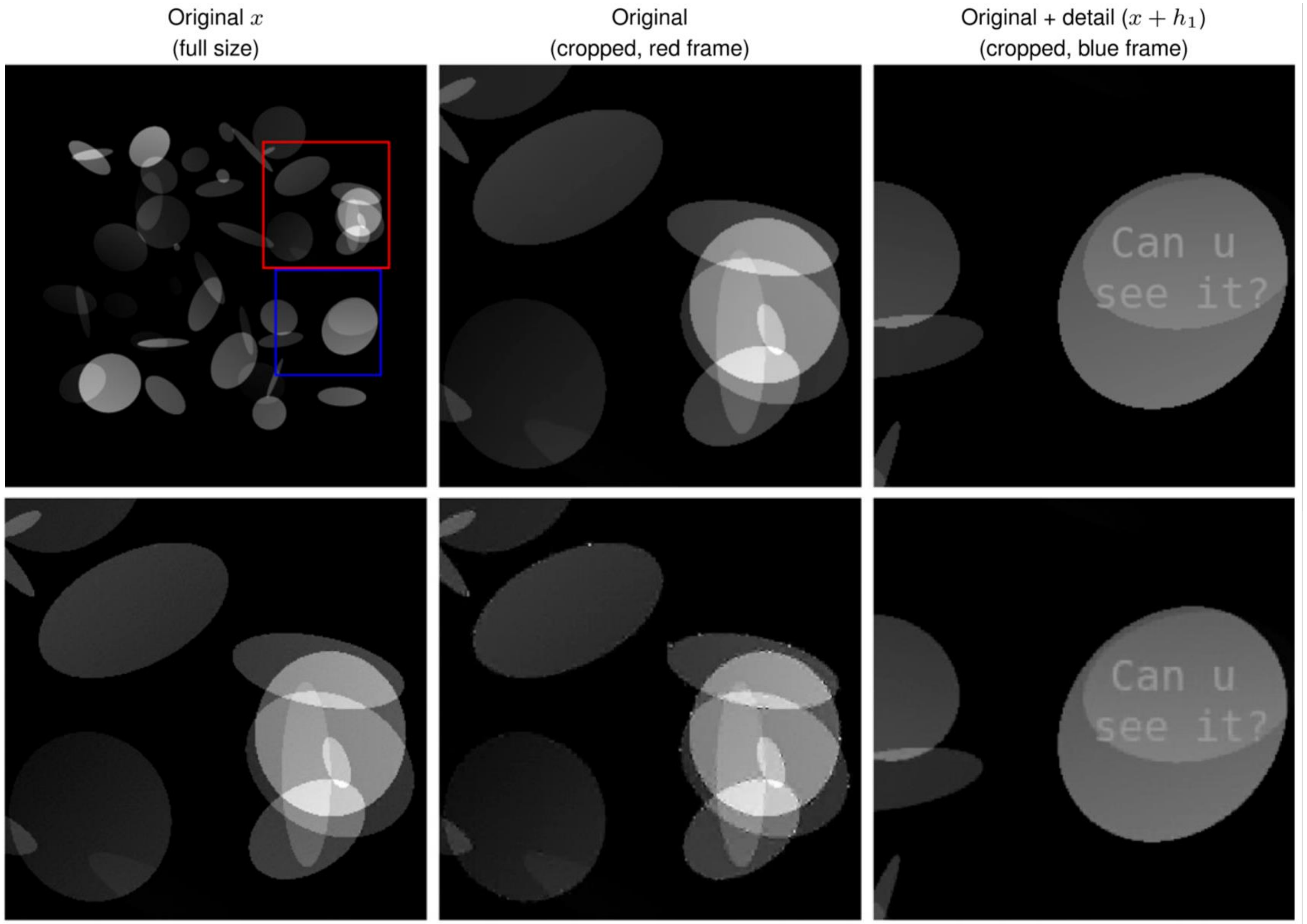
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

U-Net with noise: stable but inaccurate



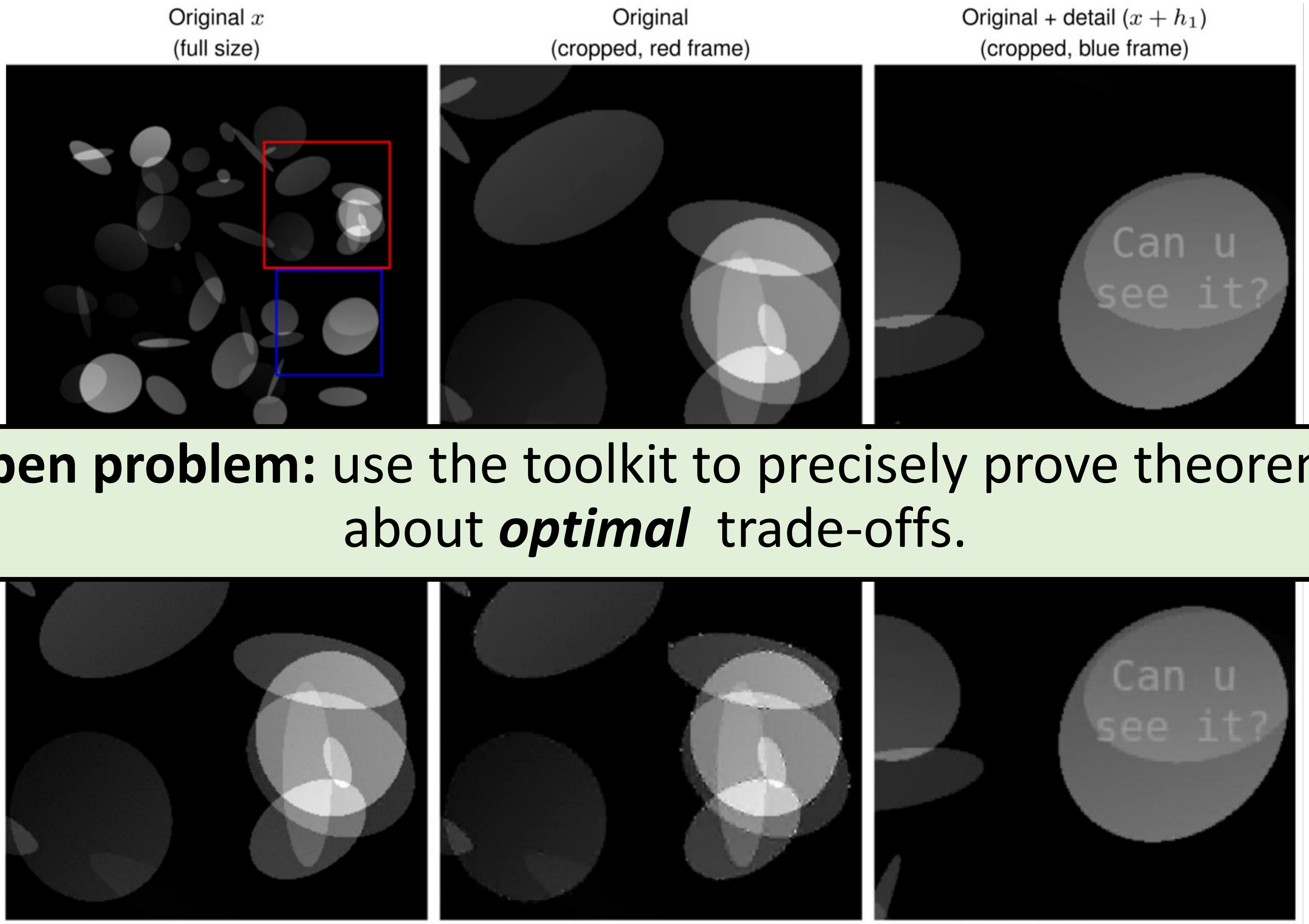
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

FIRENET: balances stability and accuracy?



- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

FIRENET: balances stability and accuracy?



Open problem: use the toolkit to precisely prove theorems about ***optimal*** trade-offs.

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

Summary

Need for foundations in AI/deep learning!

- **Paradox:** Nice linear inverse problems where stable and accurate neural network exists but cannot be trained!
- Trainability depends on
  - Accuracy desired.
  - Amount of training data.
- Specific conditions $\Rightarrow$ FIRENETs exp. convergence
  + withstand adversarial attacks.
- Trade-off between stability and accuracy in deep learning.