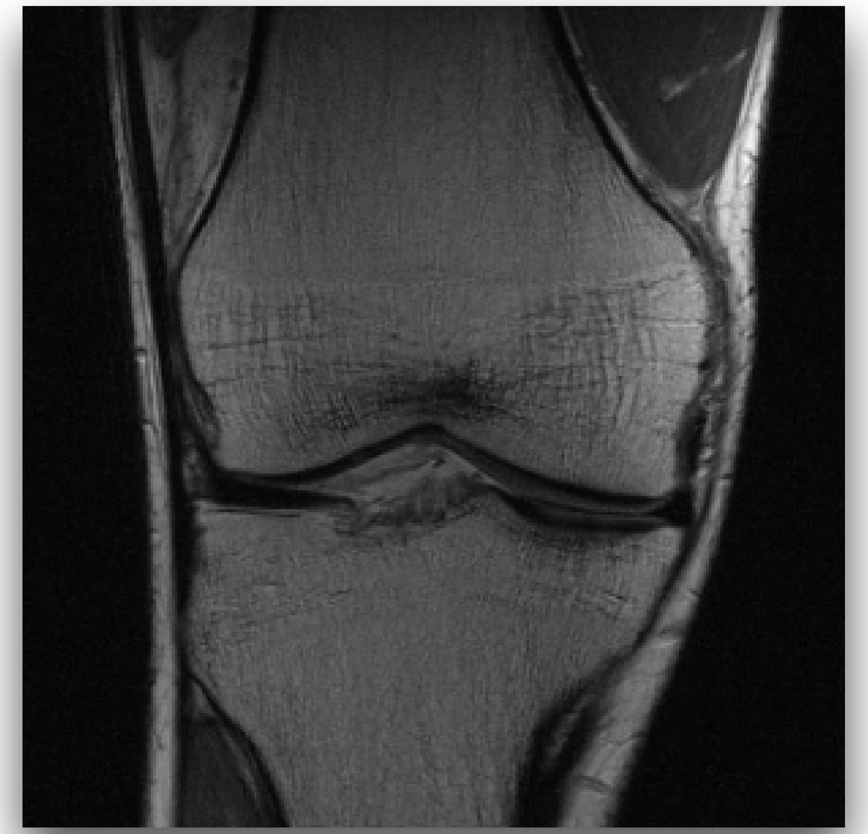
# Measuring and Enhancing Robustness in Deep Learning Based Compressive Sensing

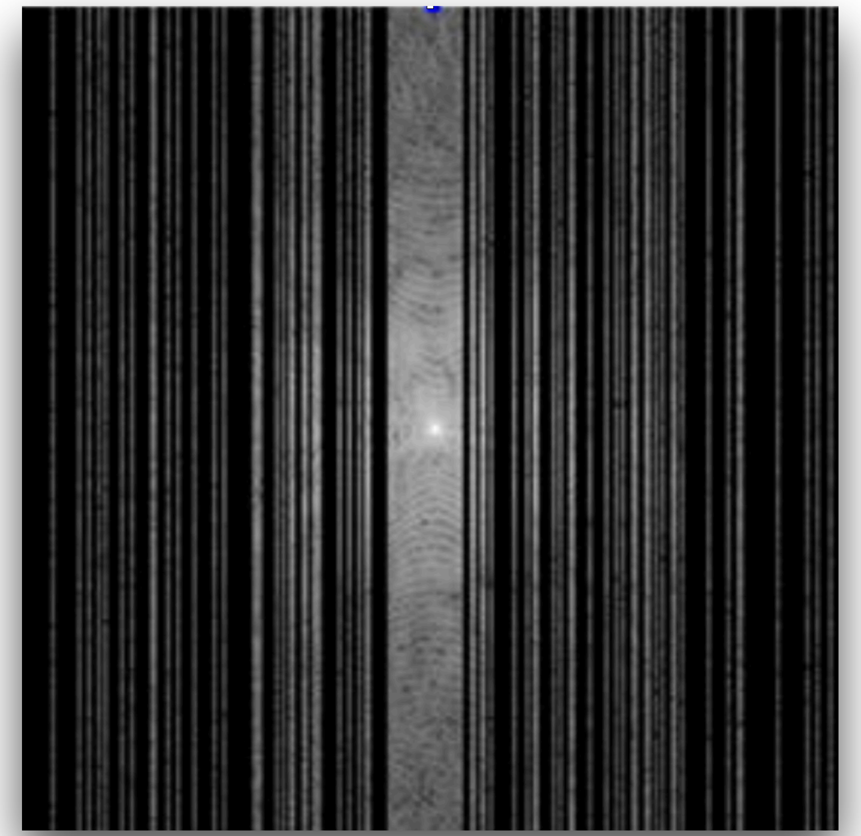
Reinhard Heckel
Technical University of Munich and Rice University

Joint work with Mohammad Zalbagi Darestani (Rice), Jiayu Liu (TUM), and Akshay Chaudhari (Stanford)

# Accelerated MRI



4x
undersampled



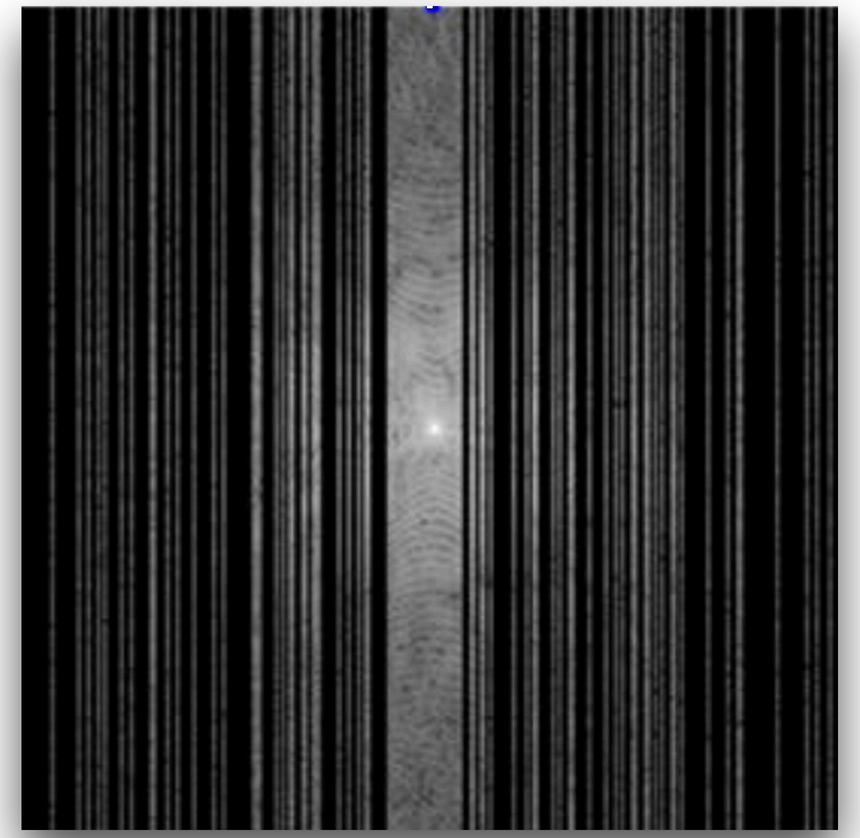
Goal:
estimate original image

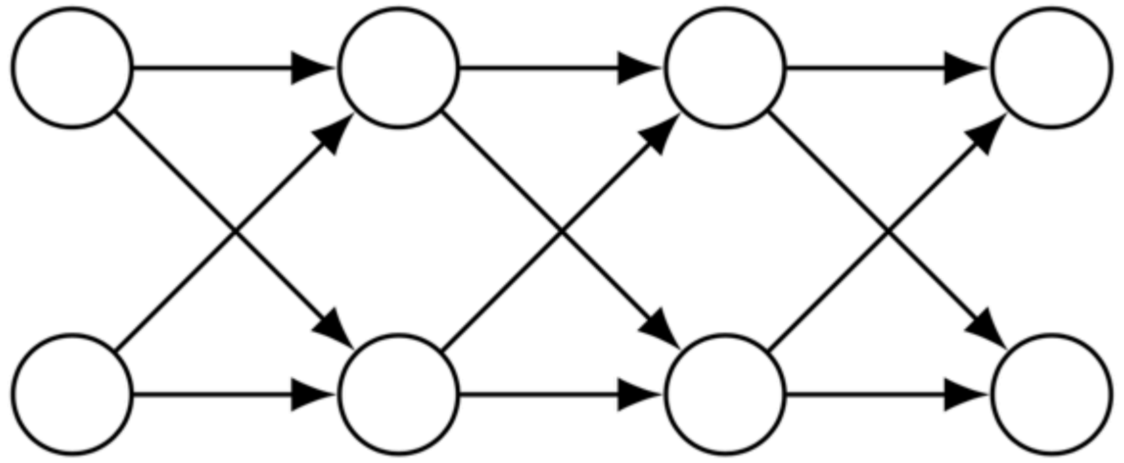
# Traditional approach: sparse recovery

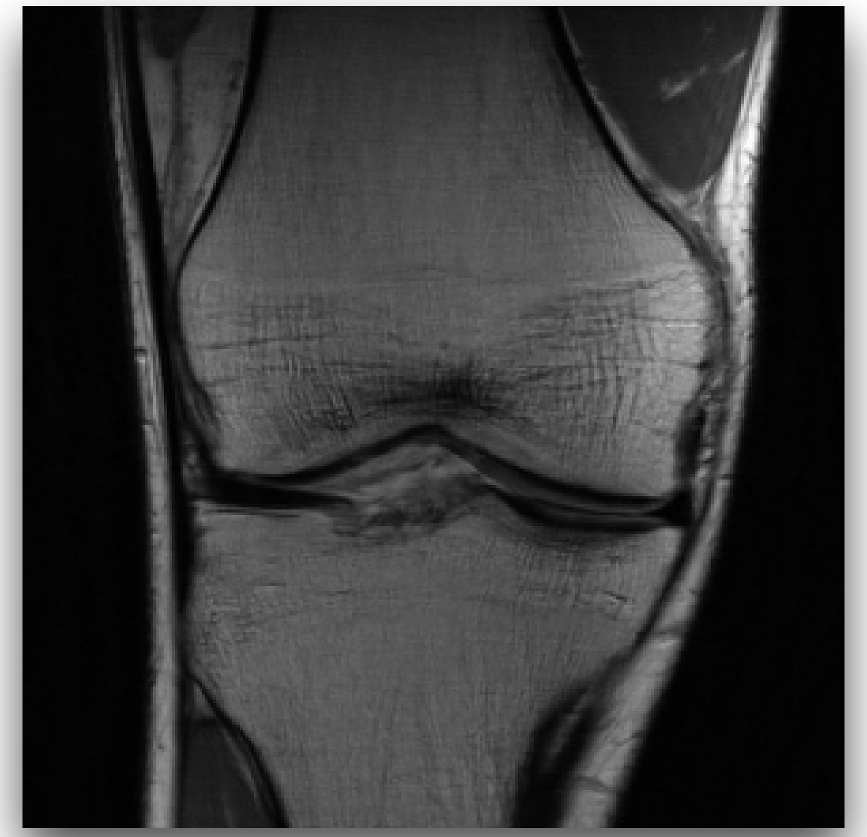
$$\hat{x} = \arg\min \|Ax - y\|_2^2 + \lambda \|x\|_{TV}$$

*Lustig et al. ``Sparse MRI..", 2007*

# Learning-based: Training end-to-end

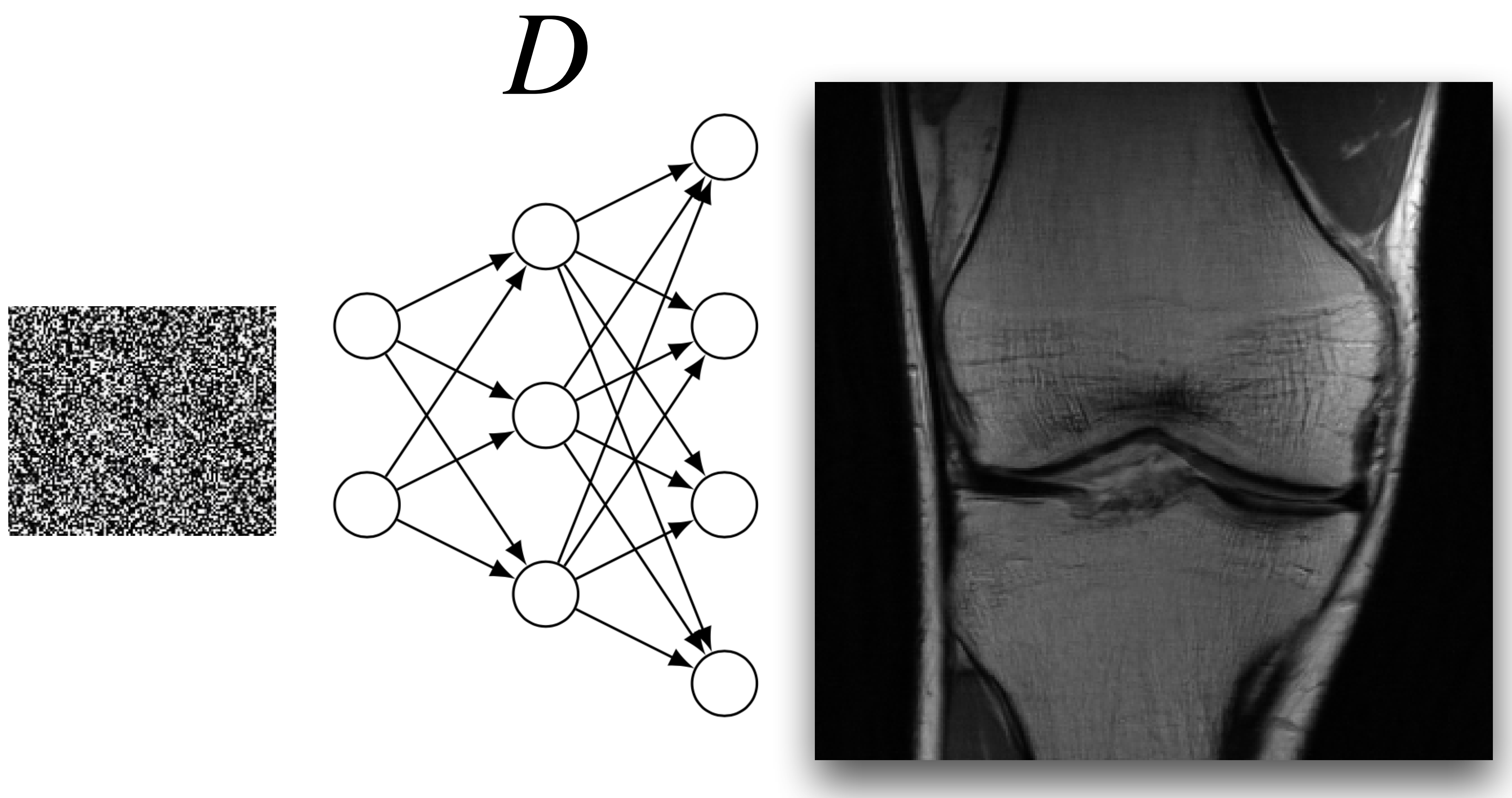






Jin et al. ``Deep convolutional neural network..", 2017
Hammernick et al., ``Learning a variational network..", 2018
Sriram et al. ``End-to-end variational networks..", 2020

# Recovery with un-trained network

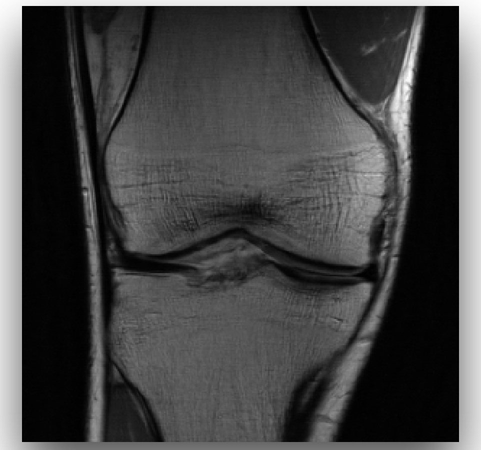
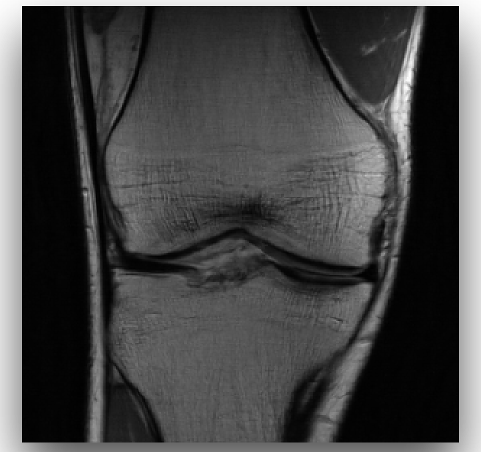



$$\hat{\theta} = \arg\min_{\theta} \|AD(\theta) - y\|_2^2$$

$$\hat{x} = D(\hat{\theta})$$

*Ulyanov et al., ``Deep image prior", 2018*
*Heckel and Hand., ``Deep decoder..", 2019*
*Van Veen et al., ``Compressed sensing with DIP..", 2018*

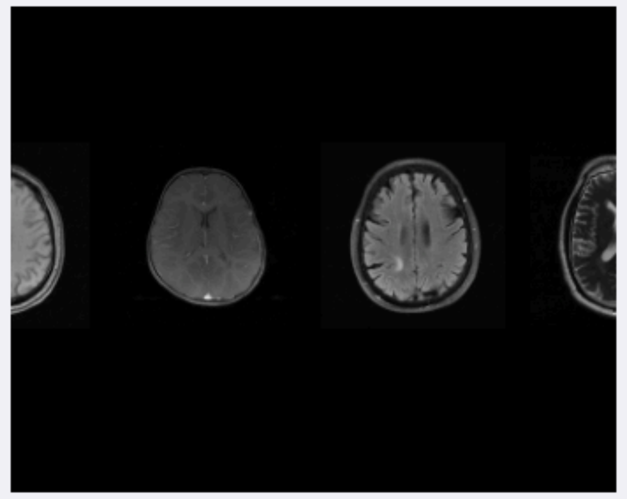
# Performance comparison

| | image quality | |
|---|---|---|
| sparsity based | ok |  |
| un-trained network | good |  |
| end-to-end network (U-net, VarNet) | very good |  |

# fastMRI

Accelerating MR Imaging with AI

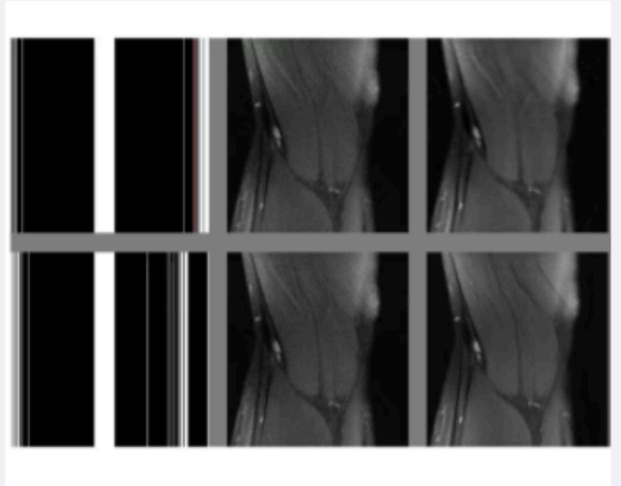
## Latest News & Updates



05-10-2021

The 2020 brain challenge paper has been acccepted in IEEE-TMI.

Read More



10-05-2

Using rei
personali

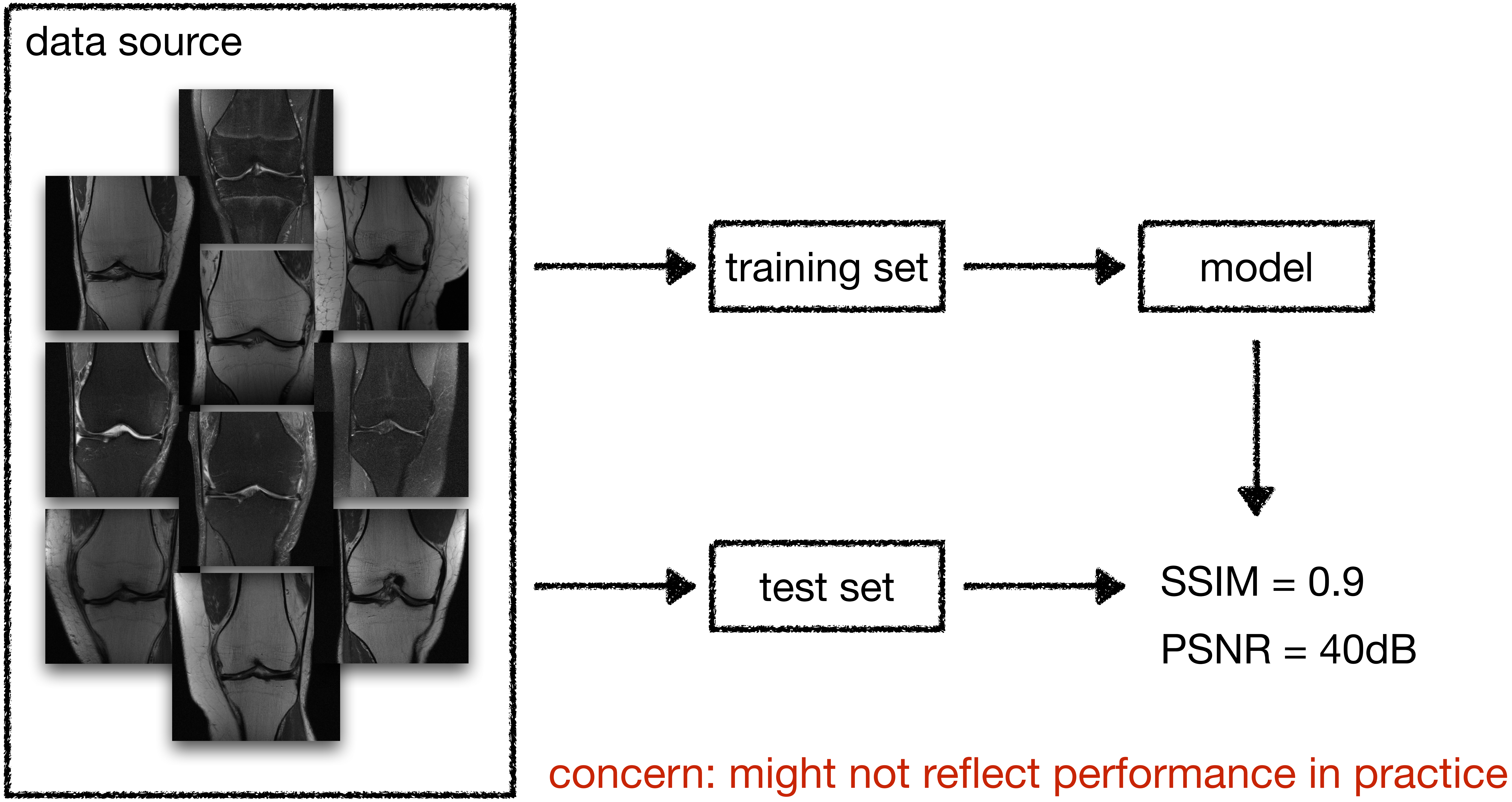
Read Mor

## What is fastMRI?

fastMRI is a collaborative research project between Facebook AI Research (FAIR) and NYU Langone Health. The aim is to investigate the use of AI to make MRI scans up to 10 times faster.

By producing accurate images from under-sampled data, AI image reconstruction has the potential to improve the patient's experience and to make MRIs accessible for more people.

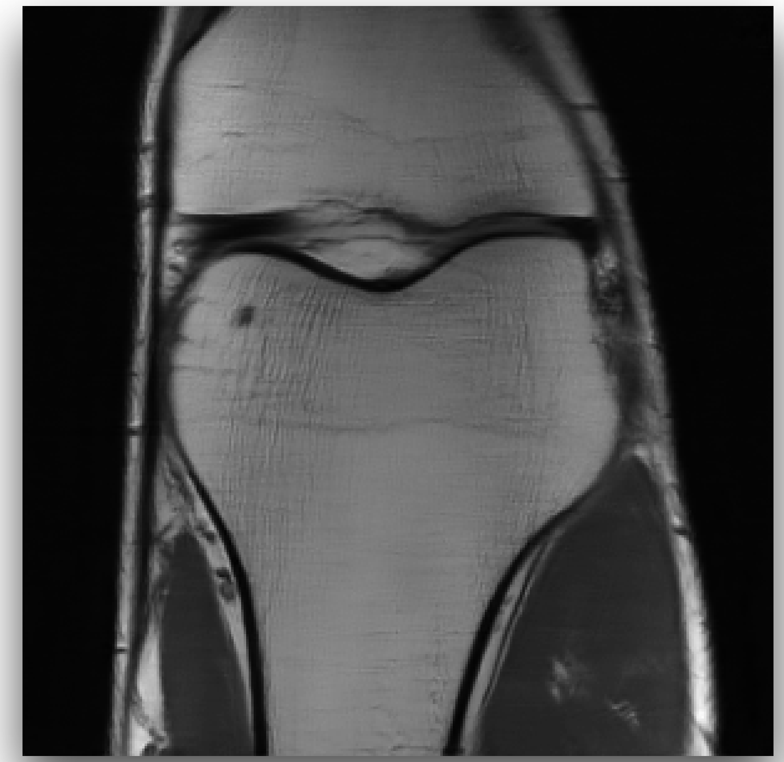
To enable the broader research community to participate in this important project, NYU Langone Health has released fully anonymized raw data and image datasets. Visit our github repository, which contains baseline reconstruction models and PyTorch data loaders for the fastMRI dataset.


data source
training set
model
test set
SSIM = 0.9
PSNR = 40dB
concern: might not reflect performance in practice

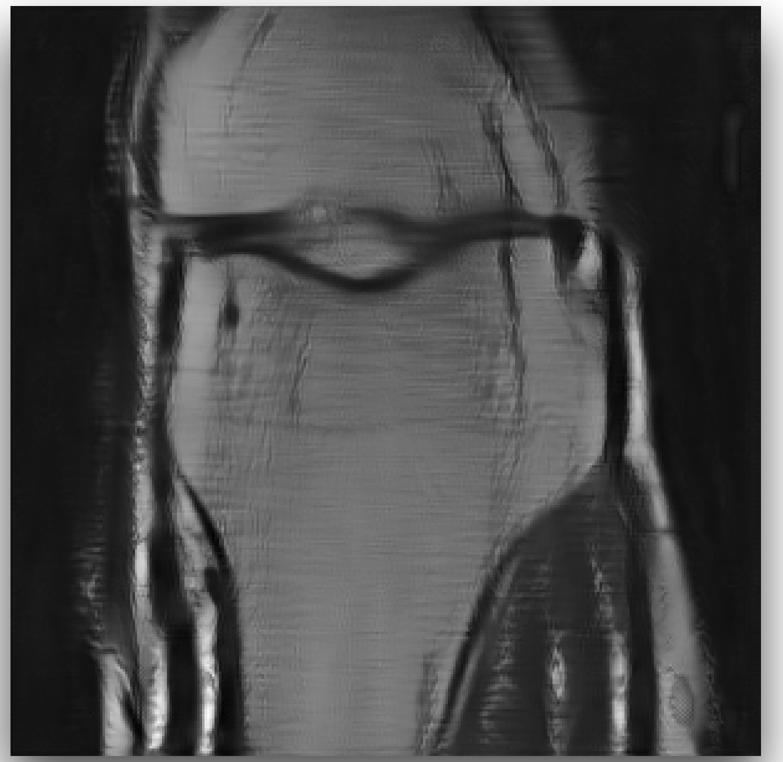
# Robustness concerns

## adversarial robustness

clean recon

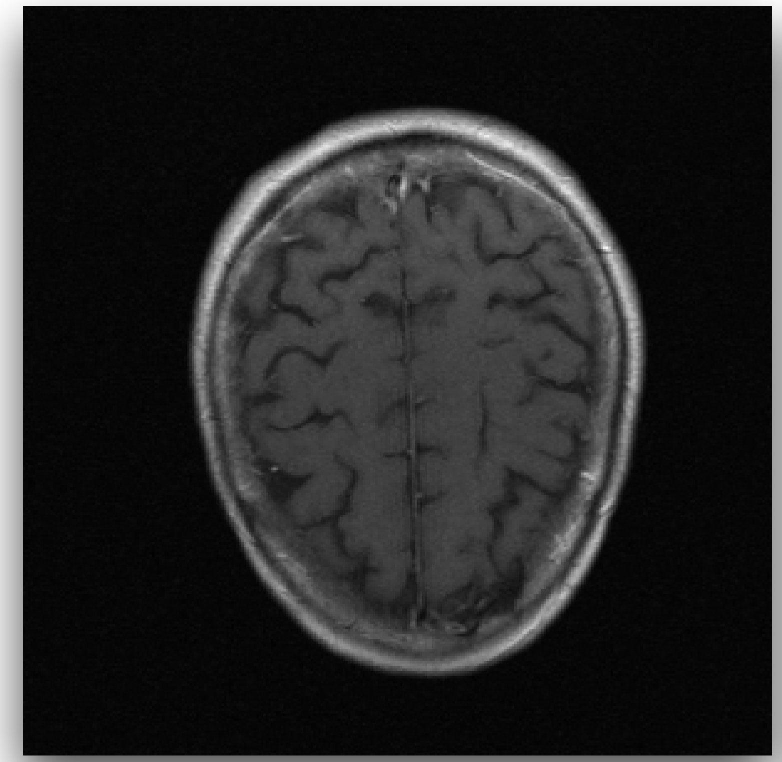


perturbed recon

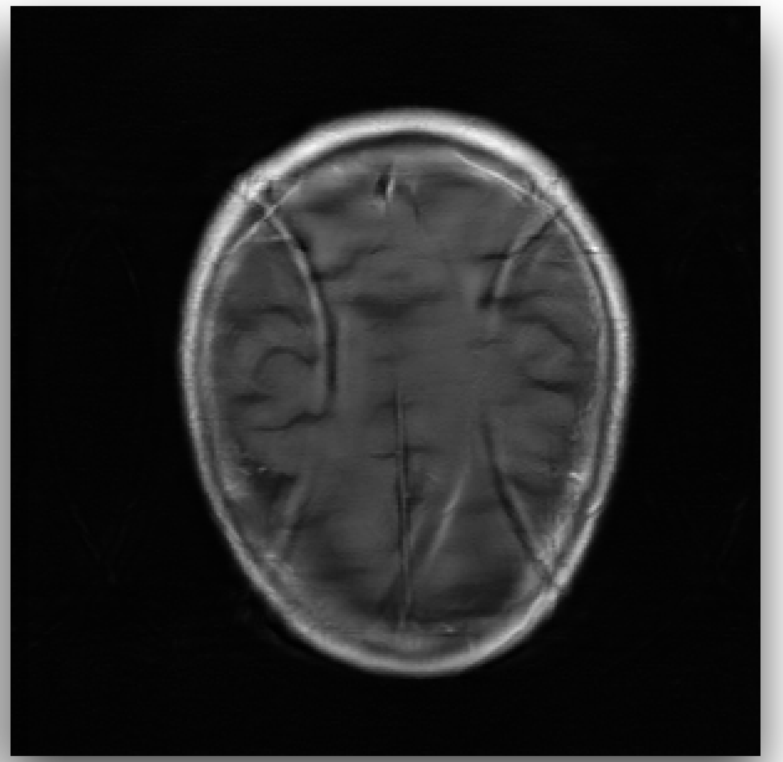


## distribution shifts

trained on brains

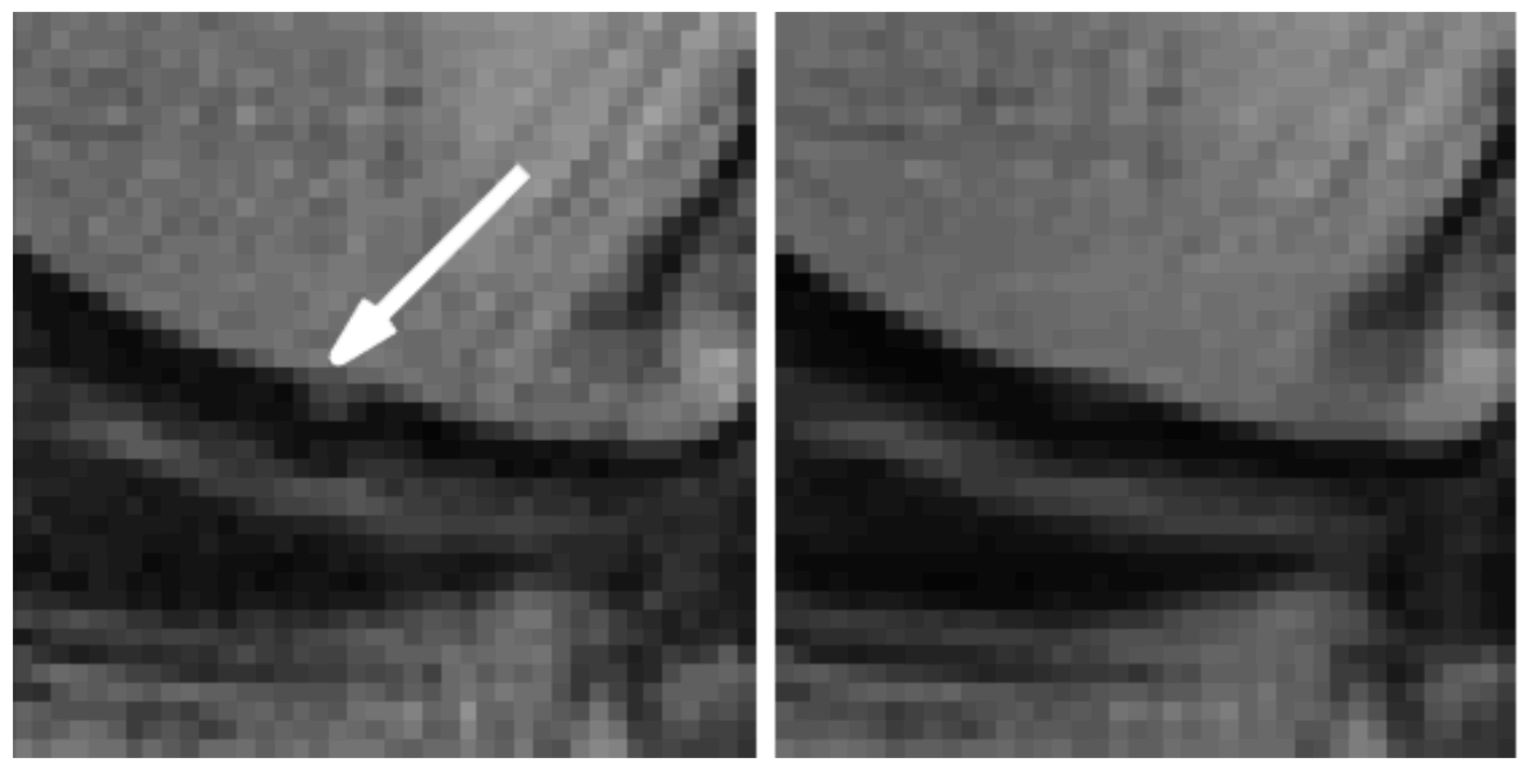


trained on knees



## fine details

original* reconstruction*



**Knoll et al. ``Advancing machine learning for MR image reconstruction", 2020*

# i: Adversarial robustness

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or D

Home | Articles | Front Matter | News | Podcasts | Authors | S

COLLOQUIUM ON THE SCIENCE OF DEEP LEARNING

Check for updates

# On instabilities of deep learning in image reconstruction and the potential costs of AI

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen

+ See all authors and affiliations

PNAS December 1, 2020 117 (48) 30088-30095; first published May 11, 2020; https://doi.org/10.1073/pnas.1907377117

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved March 12, 2020 (received for review June 4, 2019)

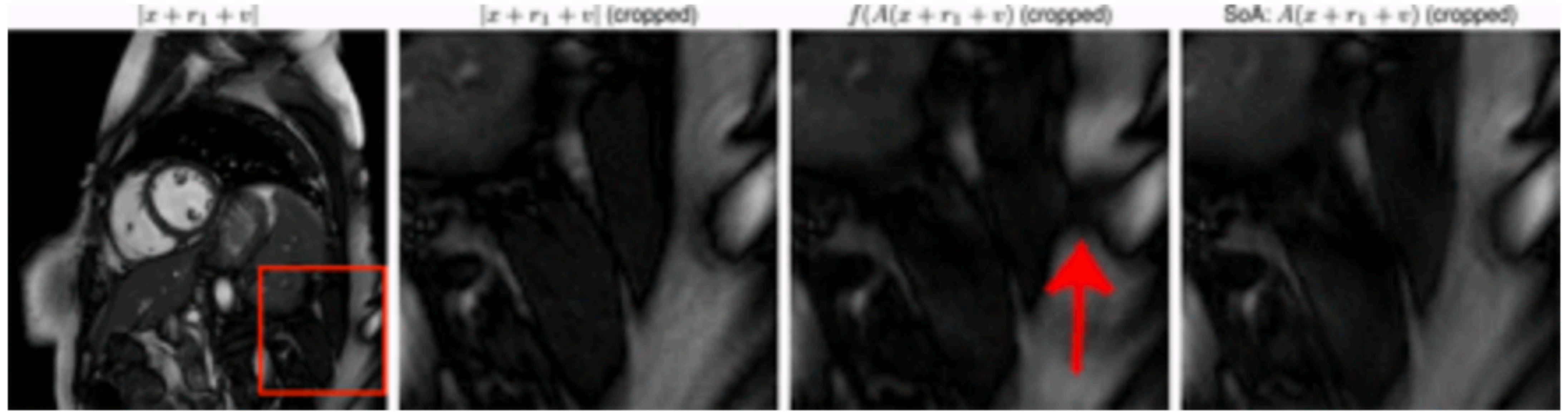
``Deep learning typically yields unstable methods for image reconstruction''

``There are cases where DL attains lower errors than sparse regularization, but in doing so it is unstable.''

# AI-Based Image Reconstruction Techniques Could Lead to Misdiagnosis

May 13, 2020
Whitney J. Palmer

*Algorithms can create errors in multiple imaging systems, according to new tests.*



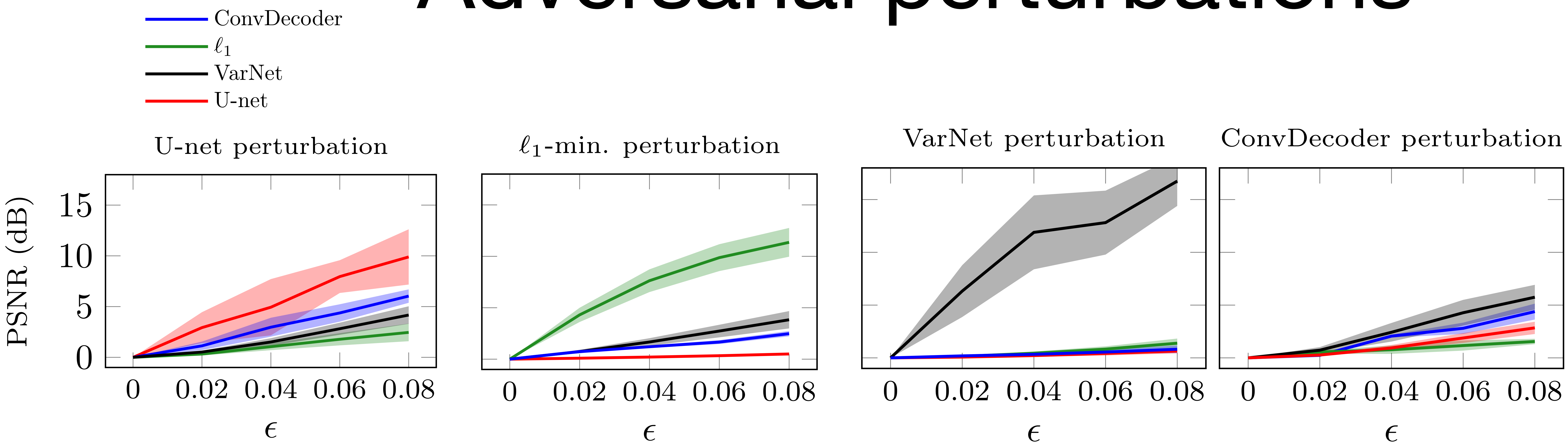

Artificial intelligence (AI) and machine learning might not be as reliable in medical imaging as previously hoped. A new study suggests these tools are "highly unstable" in medical image reconstruction.

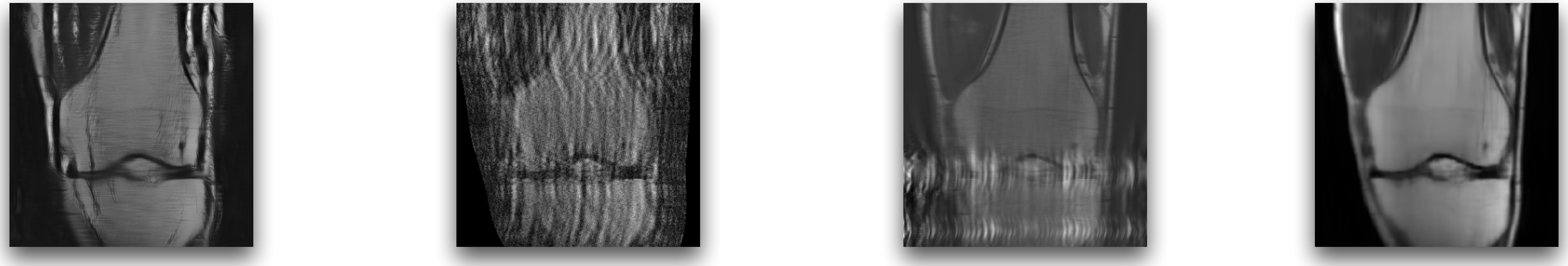
In a study published on May 11 in the [Proceedings of the National Academy of Sciences](), investigators from the University of Cambridge and Simon Fraser University, developed tests to be used on medical image reconstruction algorithms that are based on AI and deep learning. What they found indicates potential problems for radiology.

$\Psi$ : reconstruction algorithm (DNN, l1-minimization, DeepDecoder)

Adversarial perturbation: $$\hat{z} = \arg \max_{\|z\|_2 \leq \epsilon} \|\Psi(Ax^*) - \Psi(Ax^* + z)\|_2^2$$

# Adversarial perturbations

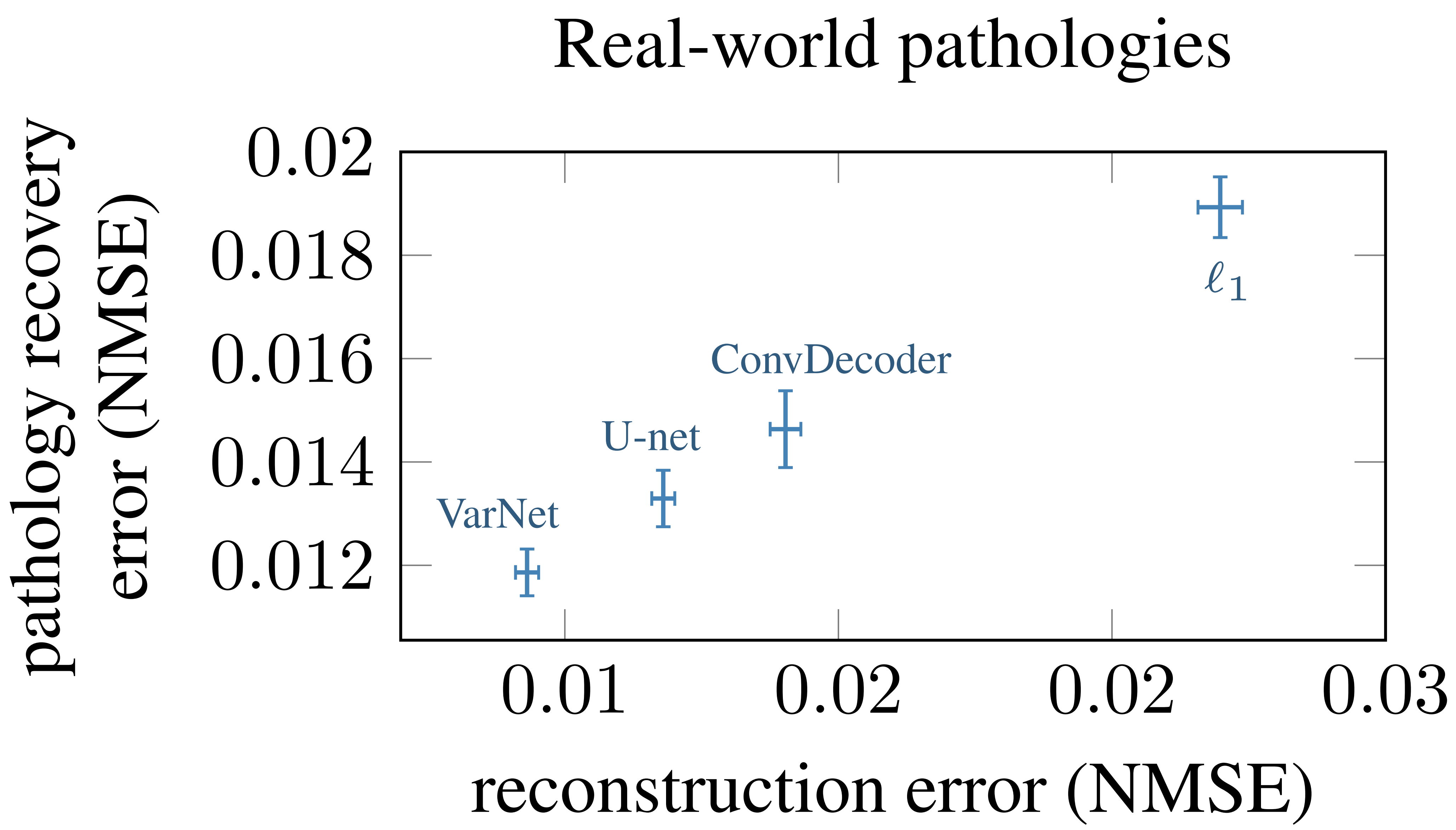


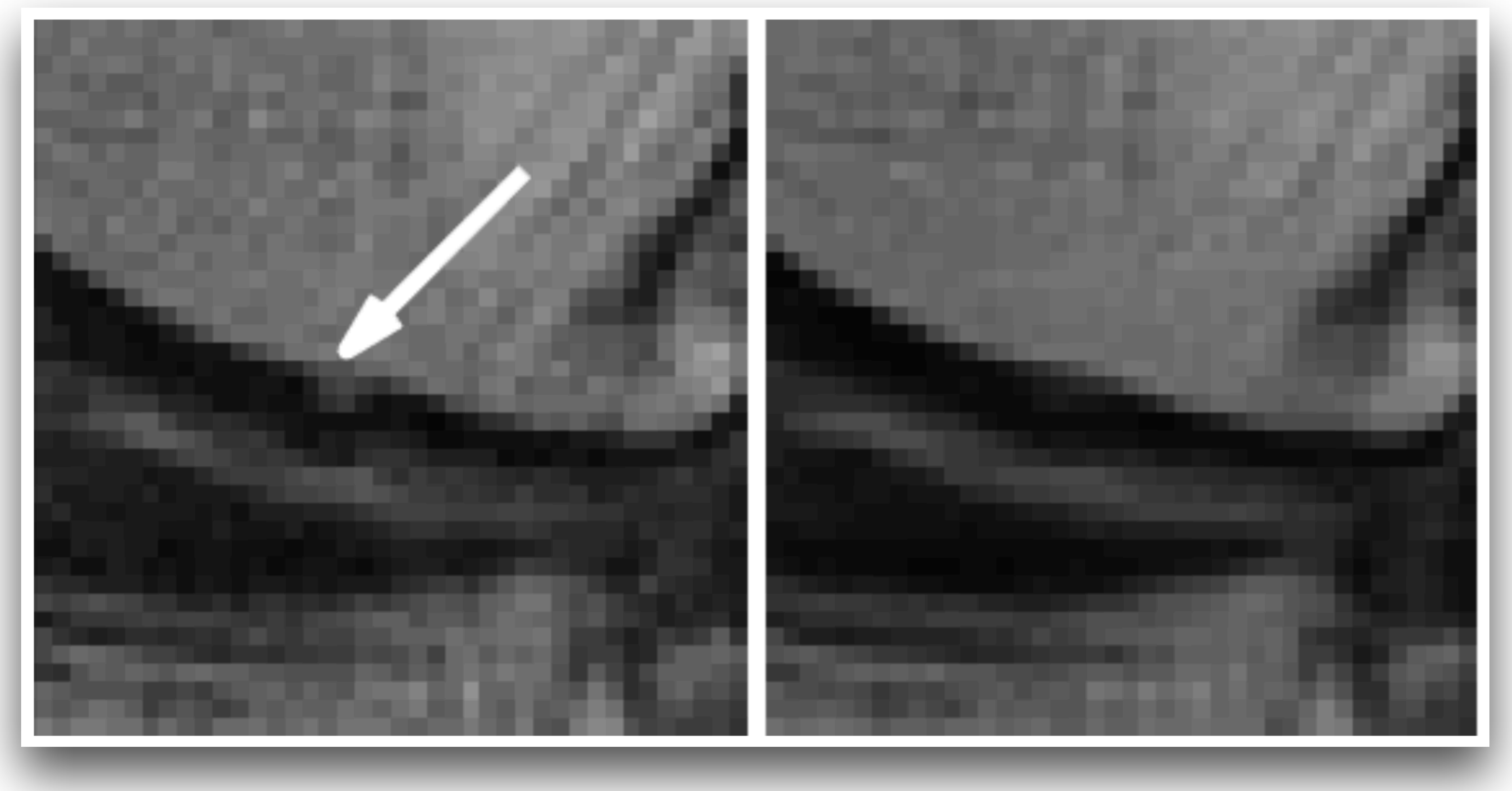



Un-trained methods are as unstable as trained ones!

# ii: Recovery of small features

# Recovery of small features

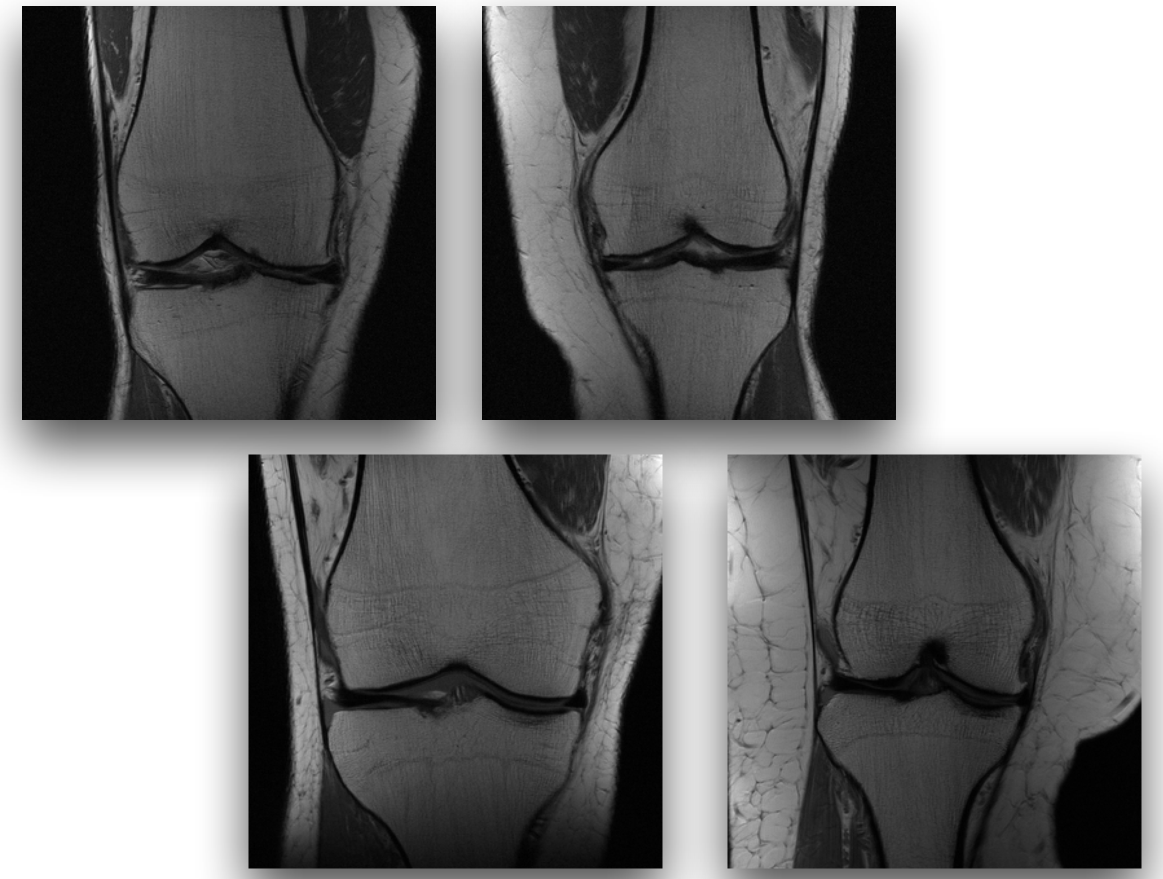


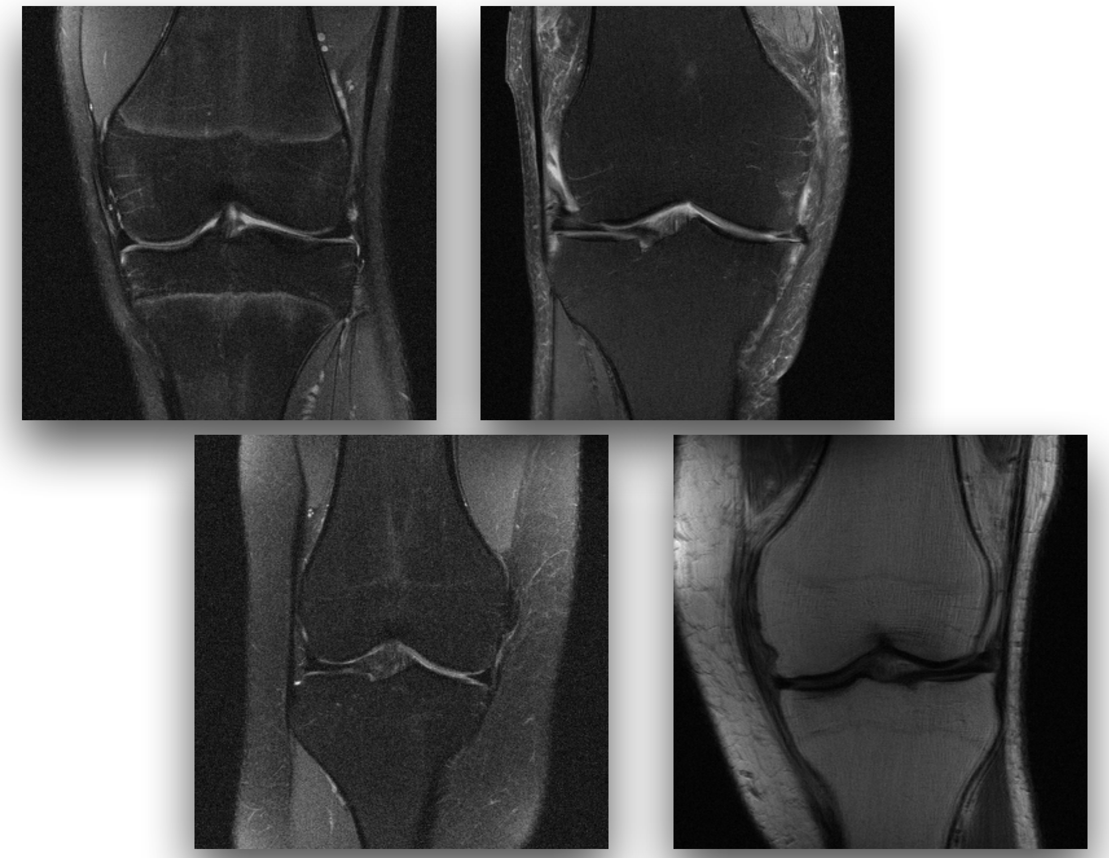
# iii: Robustness to distribution shifts

# Dataset shift

FastMRI (NYU)



Stanford dataset



training set

FastMRI test set

Stanford test set

*Differences:*

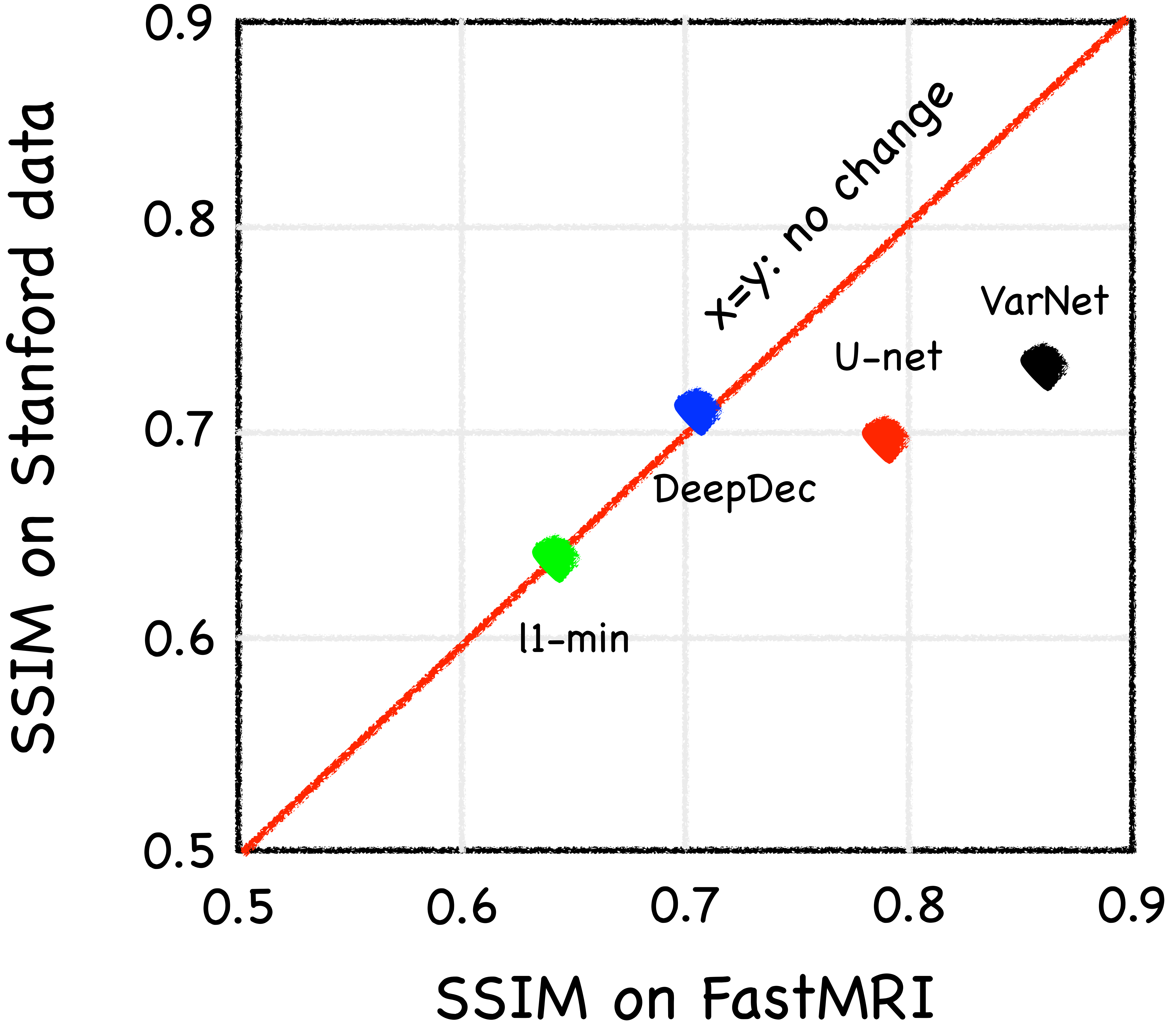
*- frequency resolution: 320x320 vs 640x360*

*- slice thickness, lower SNR*

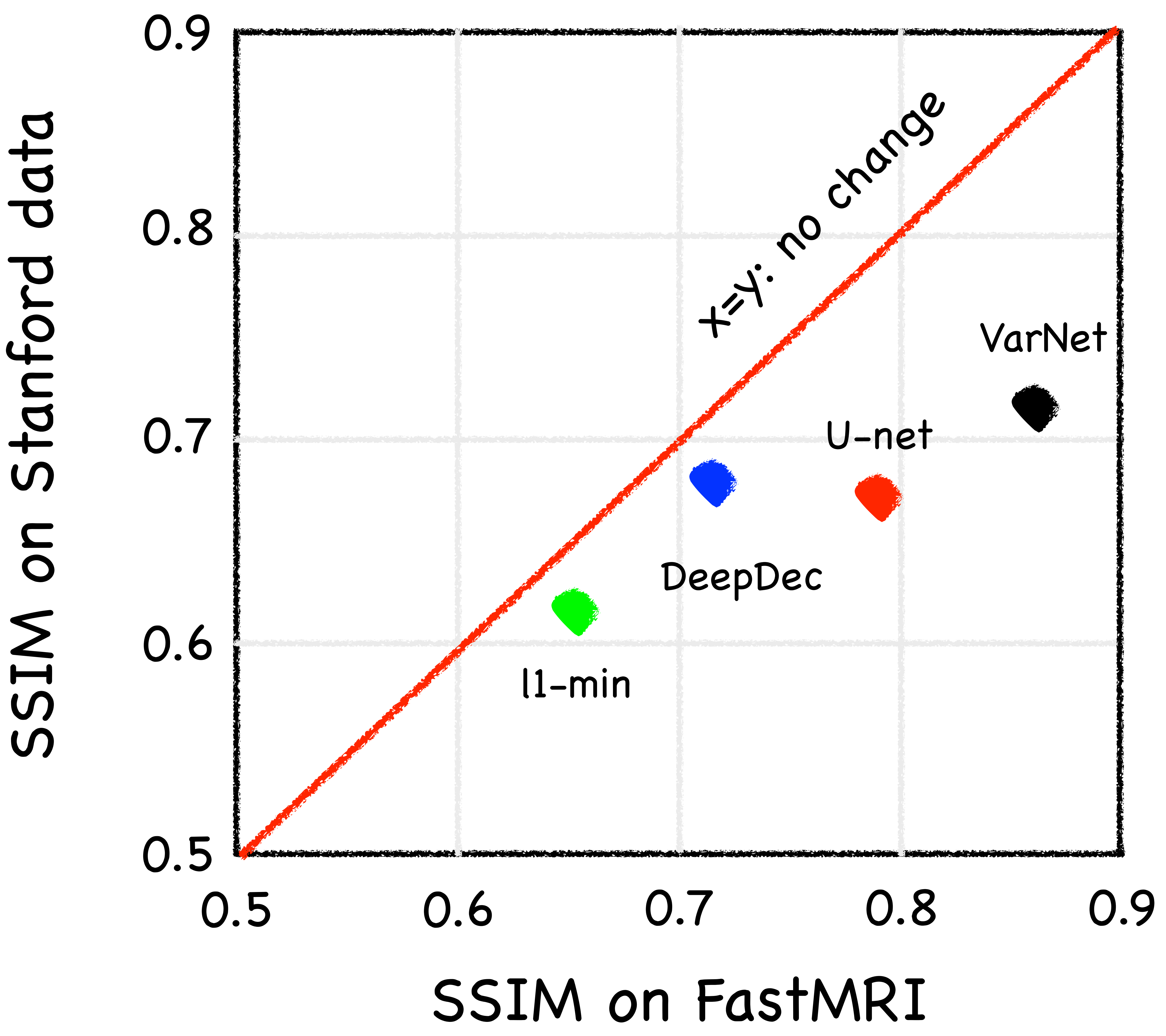
# Reconstruction methods

| | training / tuning |
|---|---|
| sparsity based | 1 parameter |
| un-trained network | 5 parameters |
| end-to-end network (U-net, VarNet) | 50 000 000 parameters |

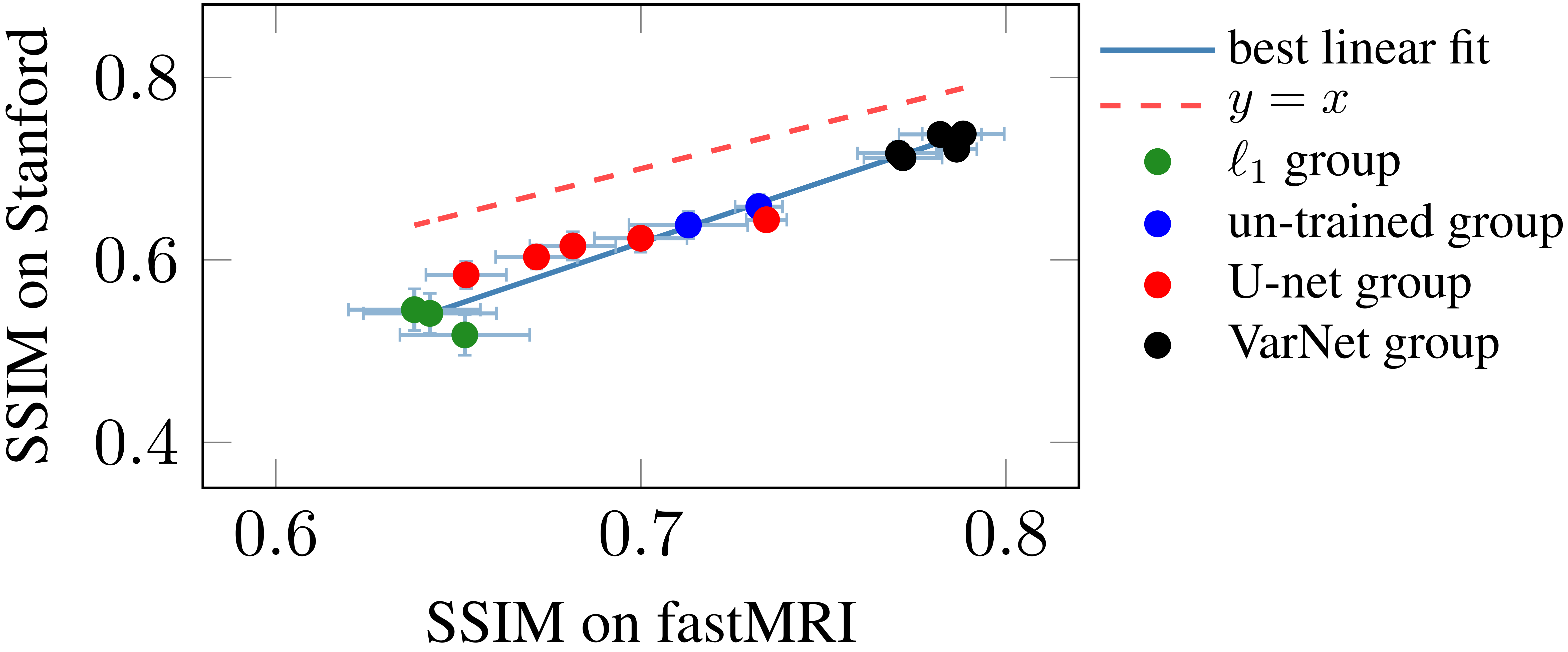
# What we might expect



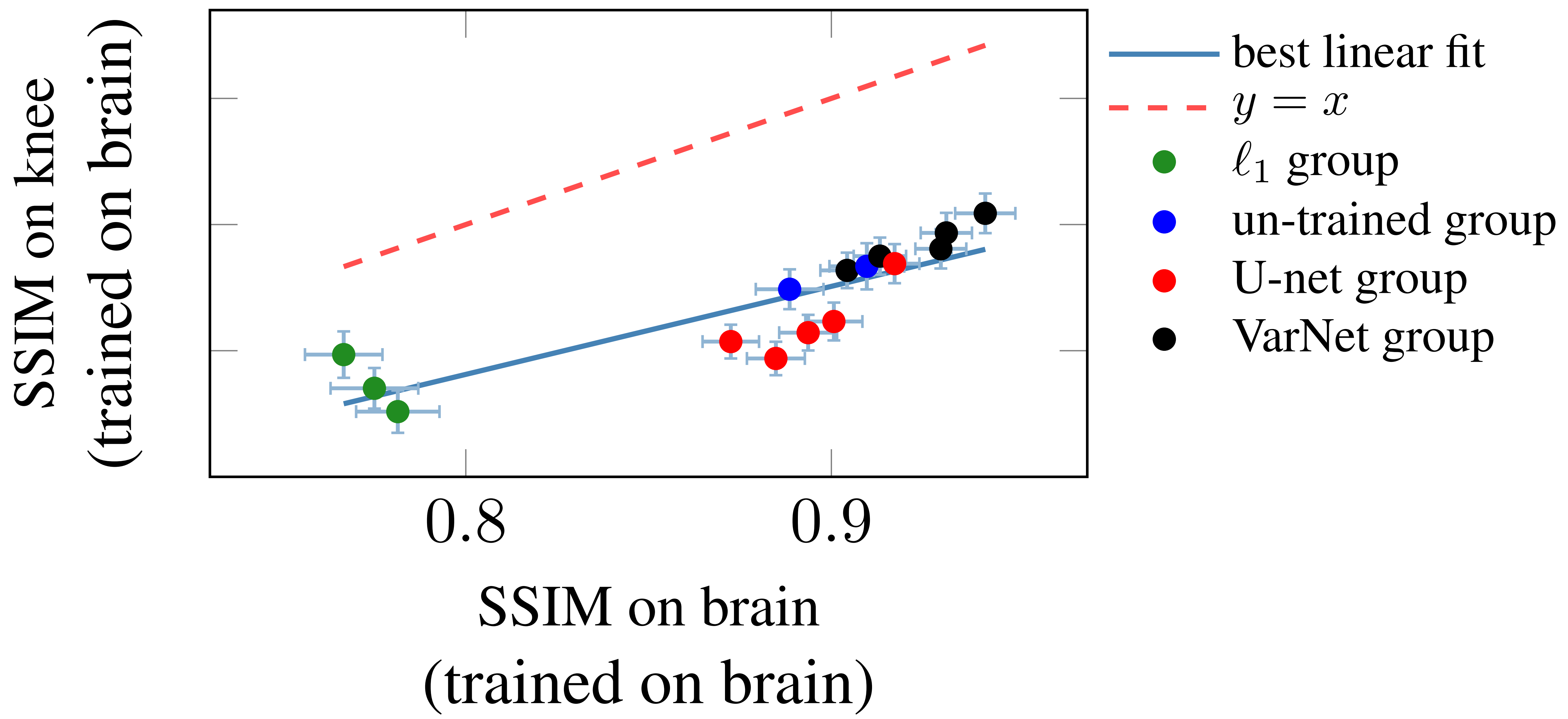
# What we might expect

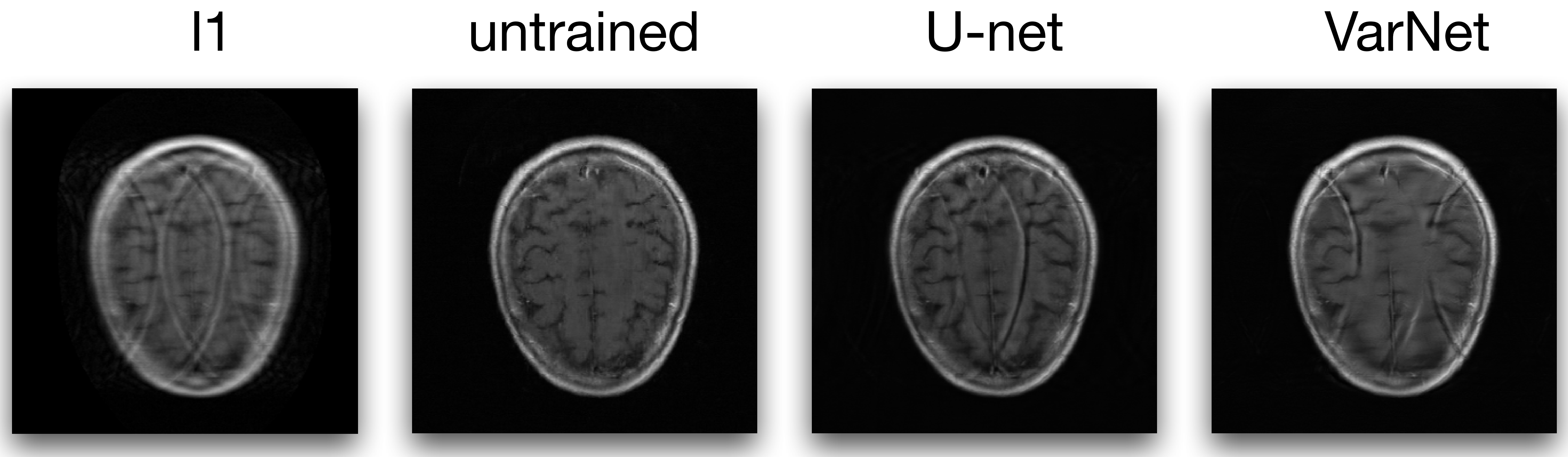


# Dataset shift

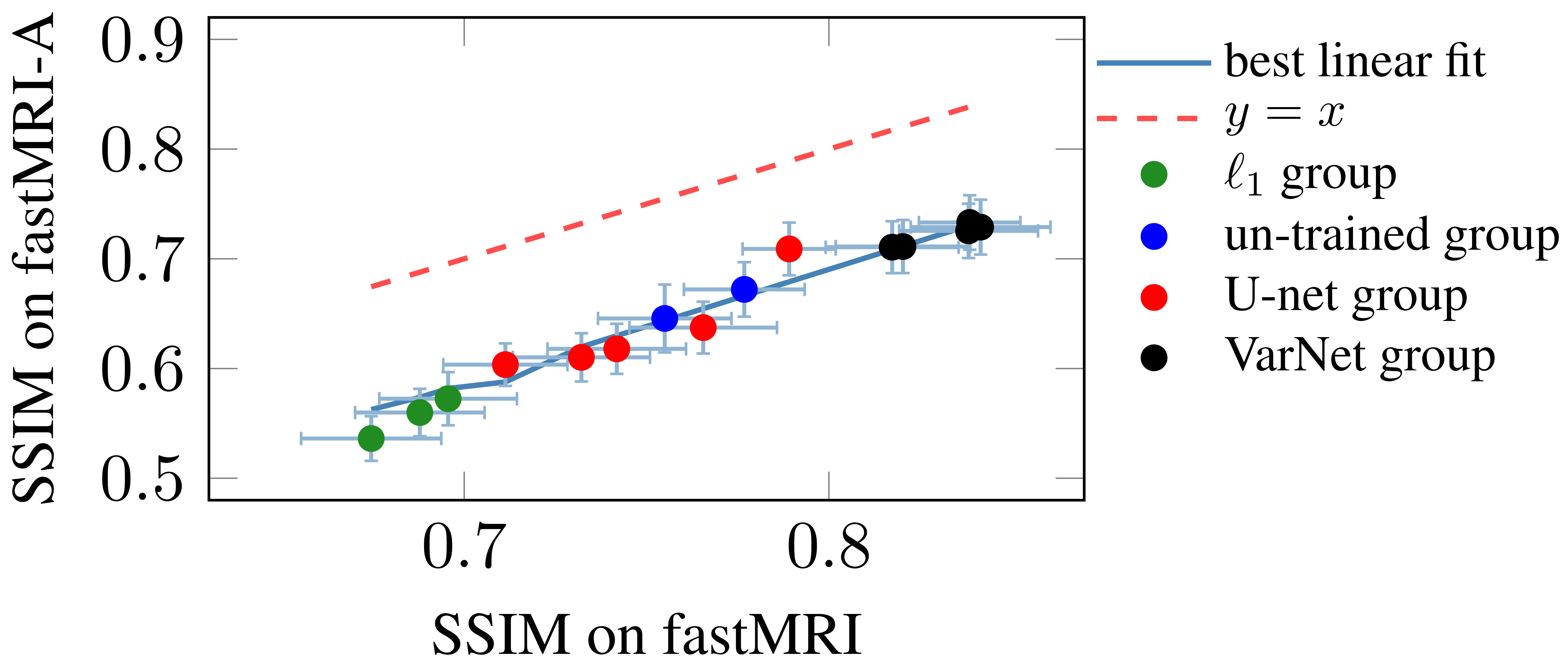


# Anatomy shift

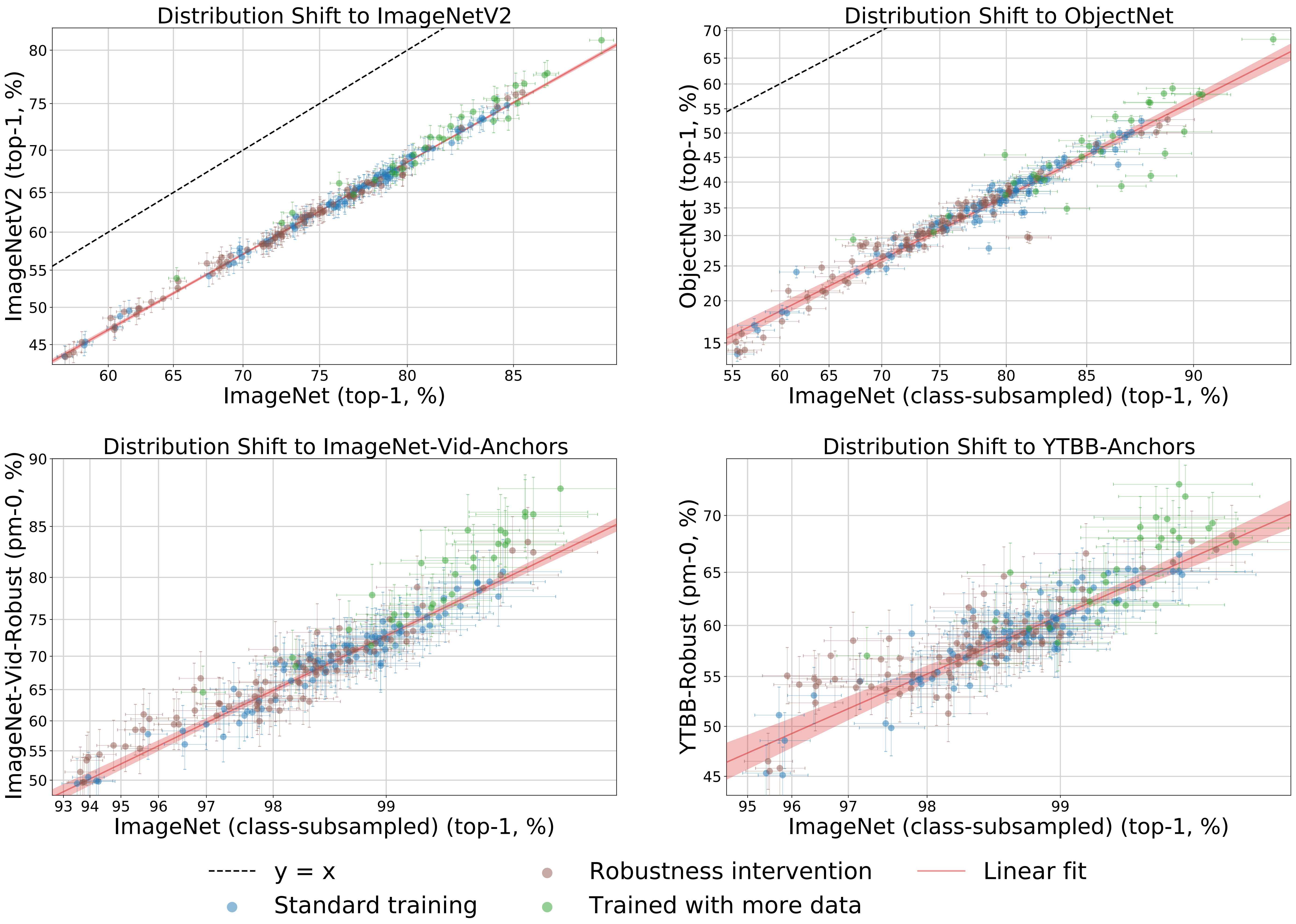


# Adversarially filtered shift



# For classification problems, ``natural distribution shifts are an open research problem''




*Taori, Dave, Shankar, Carlini, Recht, and Schmidt, ``Measuring Robustness to Natural Distribution Shifts in Image Classification''*

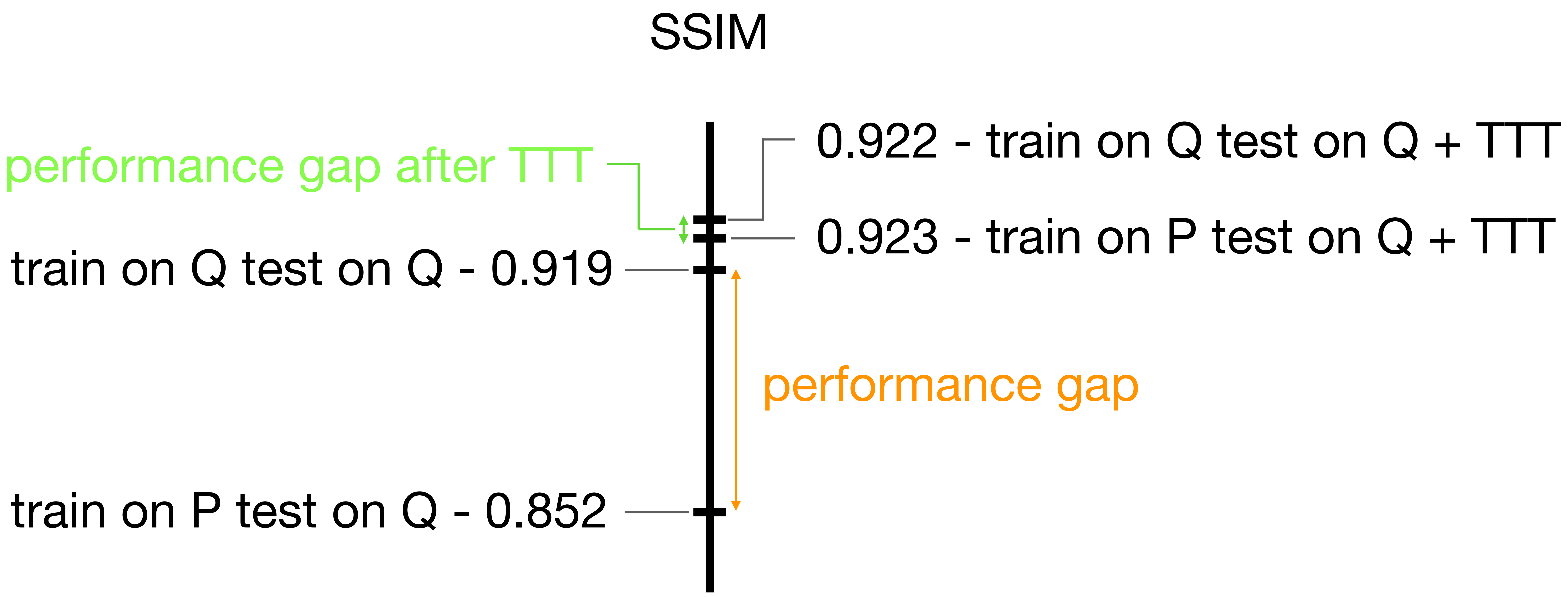
# Test time training

**Training:** $$\text{minimize}_\theta \sum_i \underbrace{\|x_i - f_\theta(y_i)\|_1}_{\text{supervised loss}} + \underbrace{\|y_i - Af_\theta(y_i)\|_1}_{\text{self-supervised loss}}$$

**Inference:** 1/ Test Time Training: $\text{minimize}_\theta \|y - Af_\theta(y)\|$

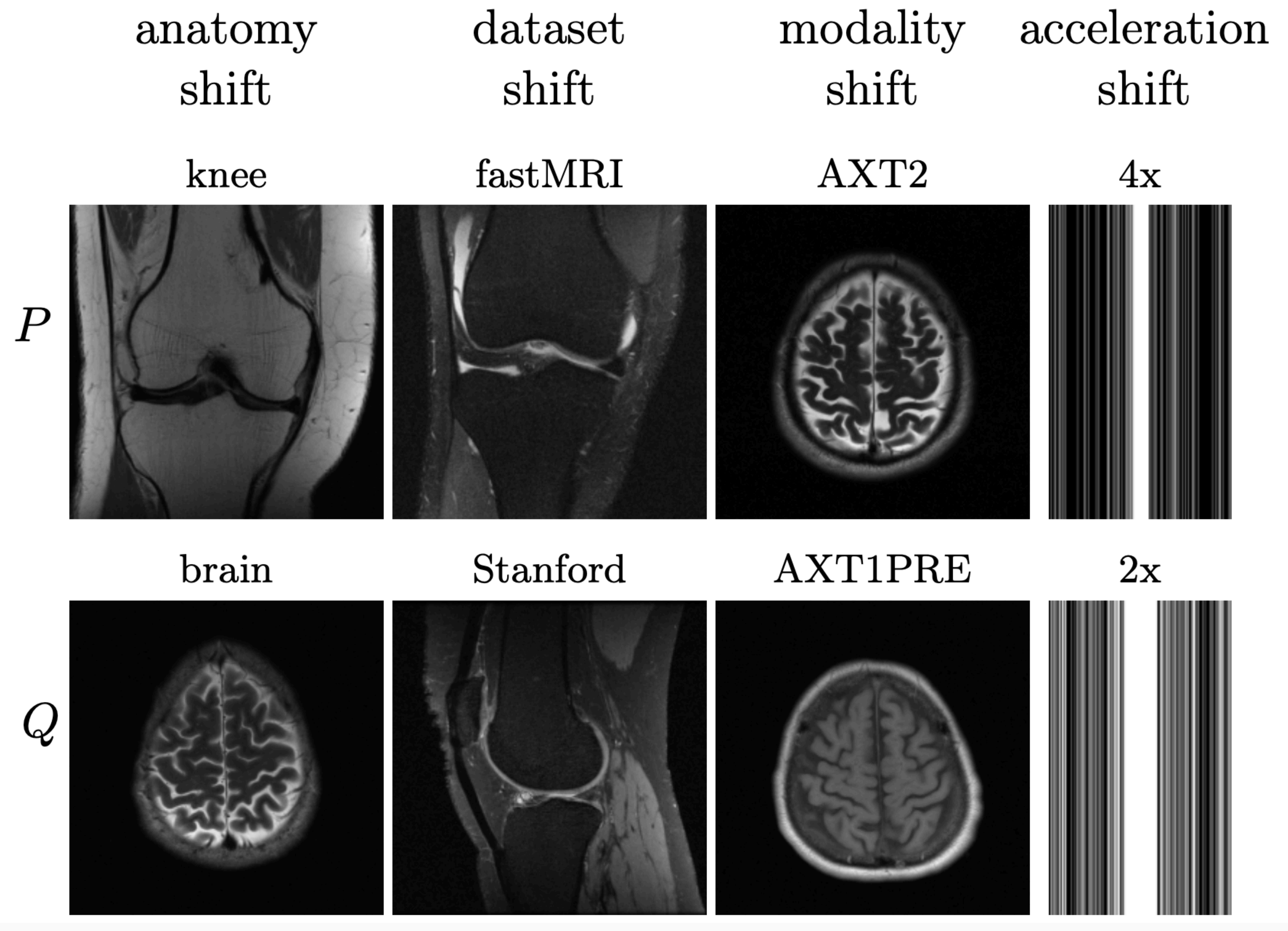
2/ Reconstruct: $\hat{x} = f_\theta(y)$

# Closing the distribution shift performance gap for anatomy shift




Test Time Training closes 99% of performance gap!

# Closing the distribution shift performance gap




Gap closed for VarNet: 99% 87% 96% 97%

# Conclusions

- No evidence that DNNs are fundamentally more susceptible to **adversarial perturbations** than sparsity based method
- **Recovery of fine details** is strongly correlated with overall reconstruction quality
- Strong linear correlation of in-distribution and **out-of-distribution generalization**
- Accuracy is a good measure for performance
- Test Time Training closes the distribution shift performance gap

# References


- M. Zalbagi Darestani, A. Chaudhari, and R. Heckel, ``Measuring Robustness in Deep Learning Based Compressive Sensing'' ICML 2021
- M. Zalbagi Darestani and R. Heckel, ``Accelerated MRI with Un-trained Neural Networks", IEEE Transactions on Comp. Imaging, 2021
- M. Zalbagi Darestani, J. Liu, and R. Heckel, ``Test-Time Training Can Close the Natural Distribution Shift Performance Gap in Deep Learning Based Compressed Sensing'' arXiv 2022
- R. Heckel and P. Hand, ``Deep Decoder: Un-trained Non-Convolutional Networks", ICLR 2018
- See also: Taori et al. ``Measuring Robustness to Natural Distribution Shifts in Image Classification", NeurIPS 2020


# Thank you!

# Abstract
# Measuring Robustness in Deep Learning Based Compressive Sensing

Traditional algorithms for reconstructing images from few and noisy measurements are handcrafted. Today, algorithms in form of deep networks learned on training data outperform traditional, handcrafted algorithms in computational cost and image quality.
However, recent works have raised concerns that deep-learning-based image reconstruction methods are sensitive to perturbations and are less robust than traditional, handcrafted, methods: Neural networks may be sensitive to small, yet adversarially-selected perturbations, may perform poorly under distribution shifts, and may fail to recover small but important features in an image. To understand the sensitivity to such perturbations, we measured the robustness of a variety of deep network based and traditional methods.
Perhaps surprisingly, in the context of accelerated magnetic resonance imaging, we find no evidence that deep learning based algorithms are less robust than classical, un-trained methods. Even for natural distribution shifts, we find that classical algorithms with a single hyper-parameter tuned on a training set compromise as much in performance than a neural network with 50 million parameters. Our results indicate that the state-of-the-art deep-learning-based image reconstruction methods provide improved performance than traditional methods without compromising robustness.