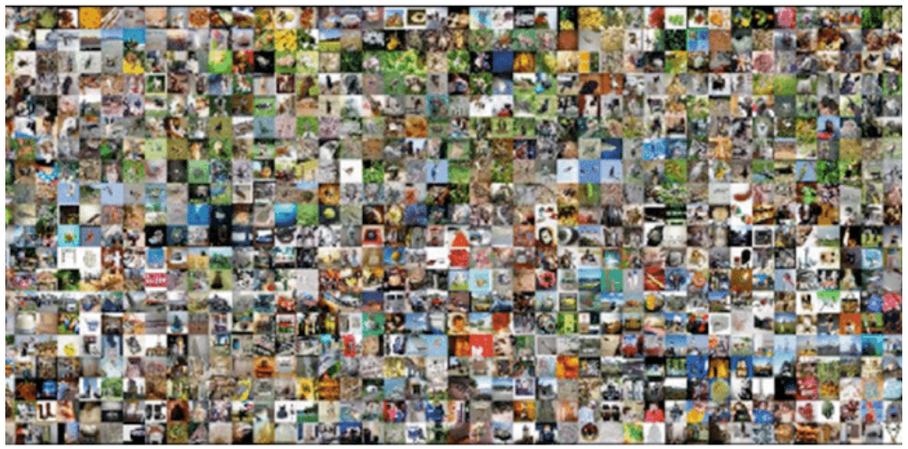
Overcoming the data bottleneck in AI for the sciences

Zalan Fabian

MHI Finalist Talk 2021

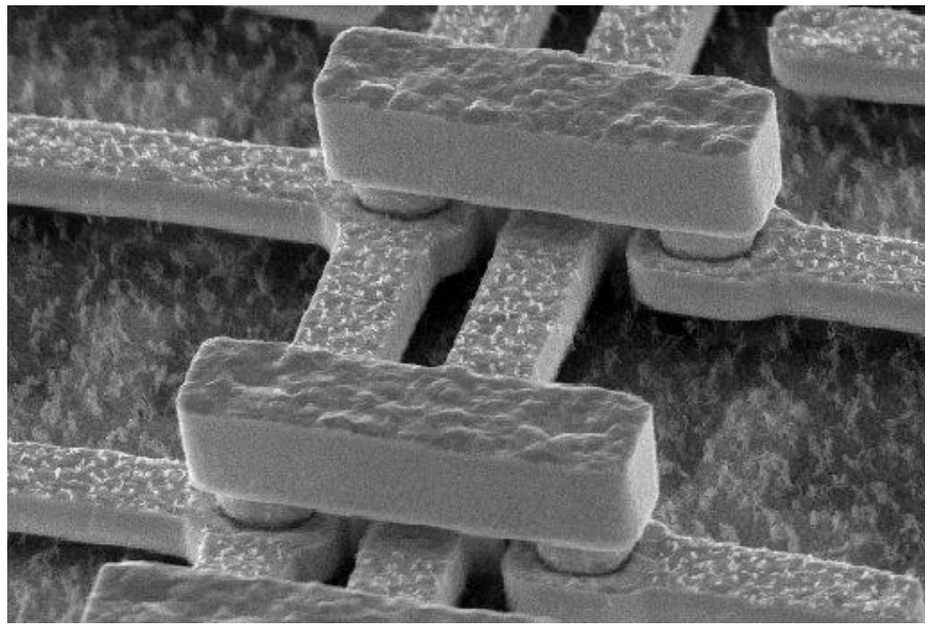
Deep learning is data-hungry

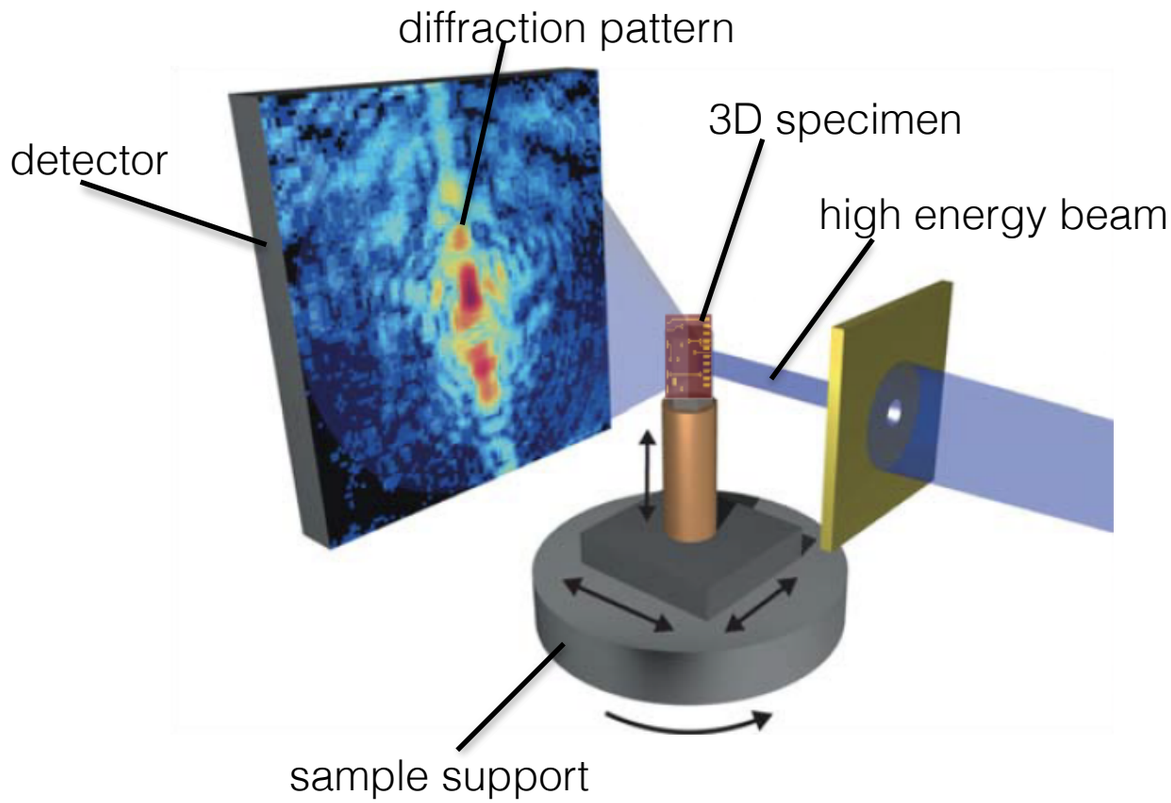


	'Internet data'
Data acquisition	readily available
Cost	very low
Amount of data	huge

How about scientific data?

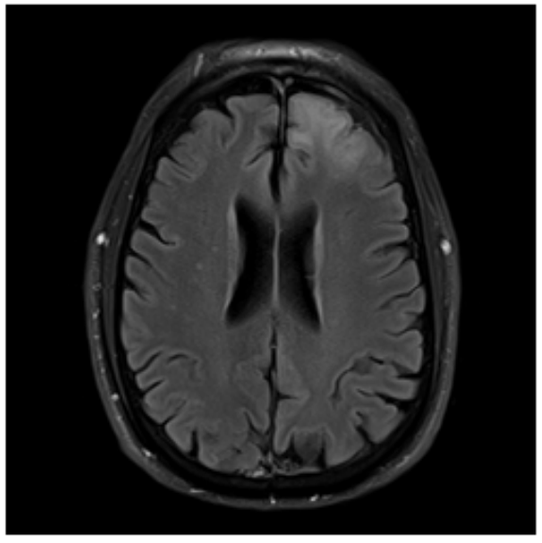
Case study: nano-scale imaging



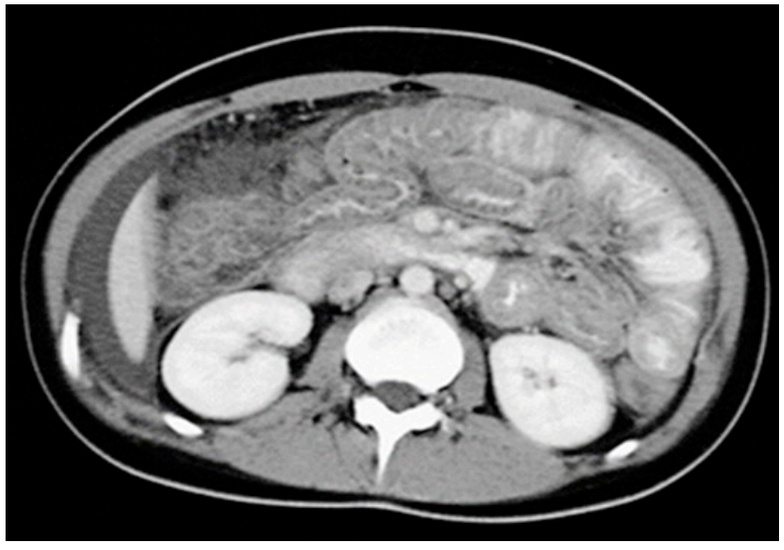


	3D CDI data
Data acquisition	1cm³ volume ≈ 2500 days
Cost	very high
Amount of data	some simulated data

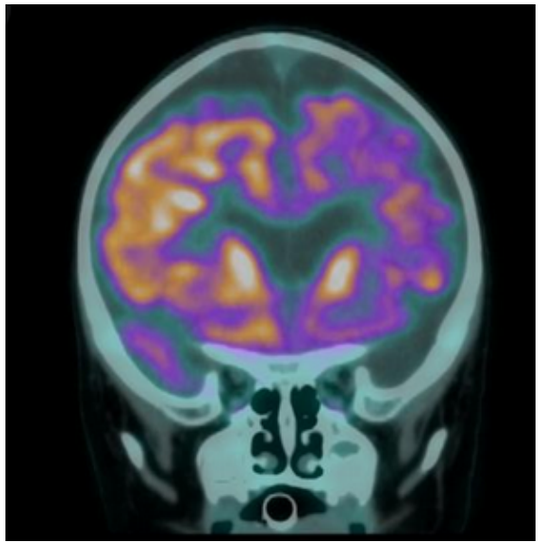
Case study: medical imaging



MRI



CT



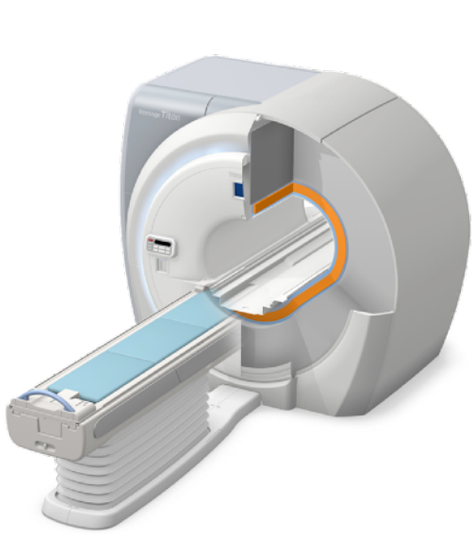
PET

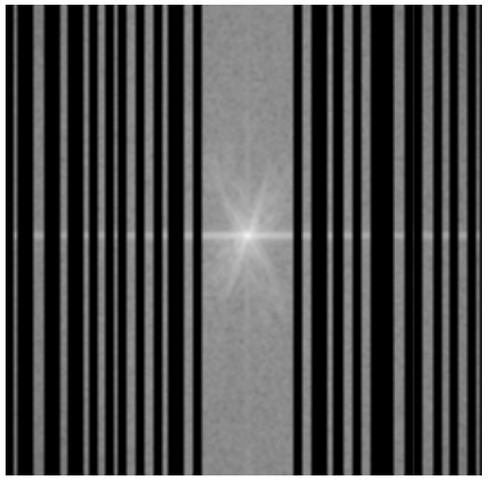
	Medical imaging data
Data acquisition	slow, harmful radiation
Cost	high
Amount of data	small public datasets

How to address the data bottleneck?

Reducing training data for MRI reconstruction

MRI reconstruction

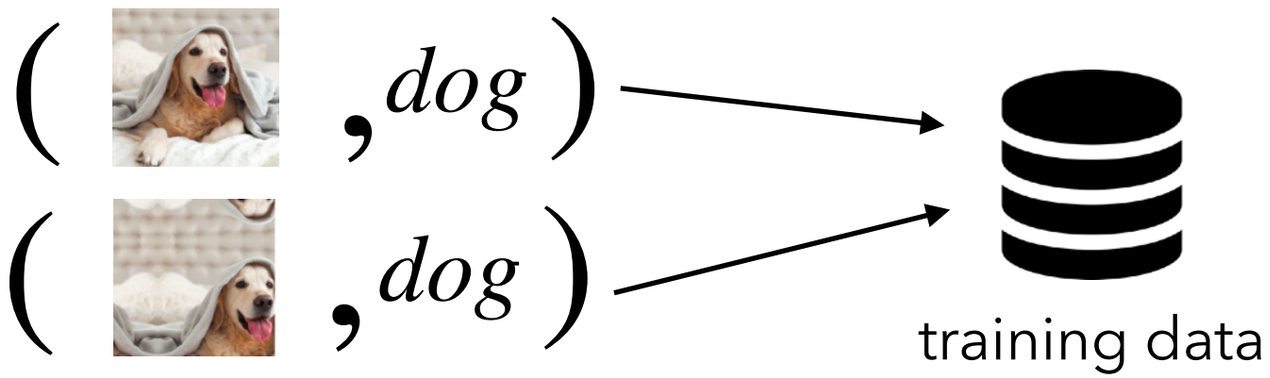




?

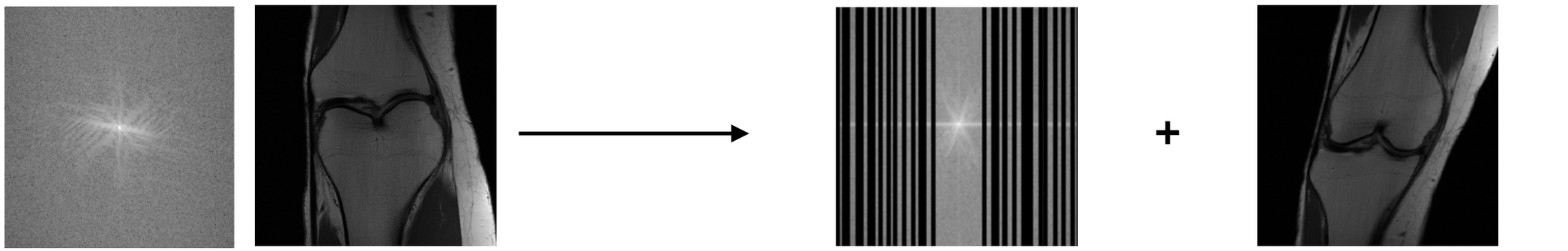
$\mathcal{F}$-transform + undersampling + noise

Data augmentation in classification: straightforward



Data augmentation in regression: non-trivial

1. Output is **not** invariant to transformations



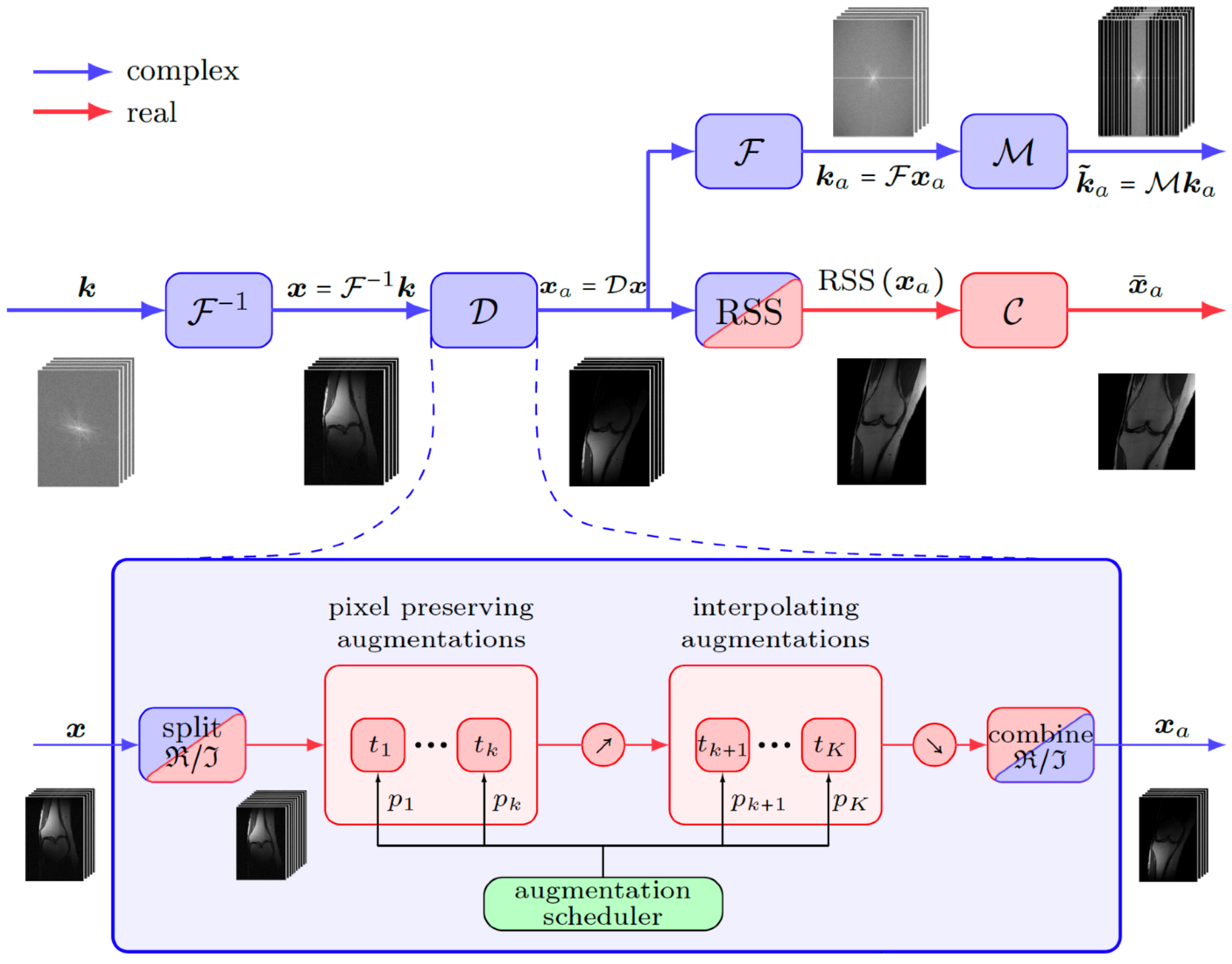
2. Distribution shift due to noise

$$x = x^* + n \longrightarrow x_{aug.} = \mathscr{D}x^* + \mathscr{D}n$$

augmented signal

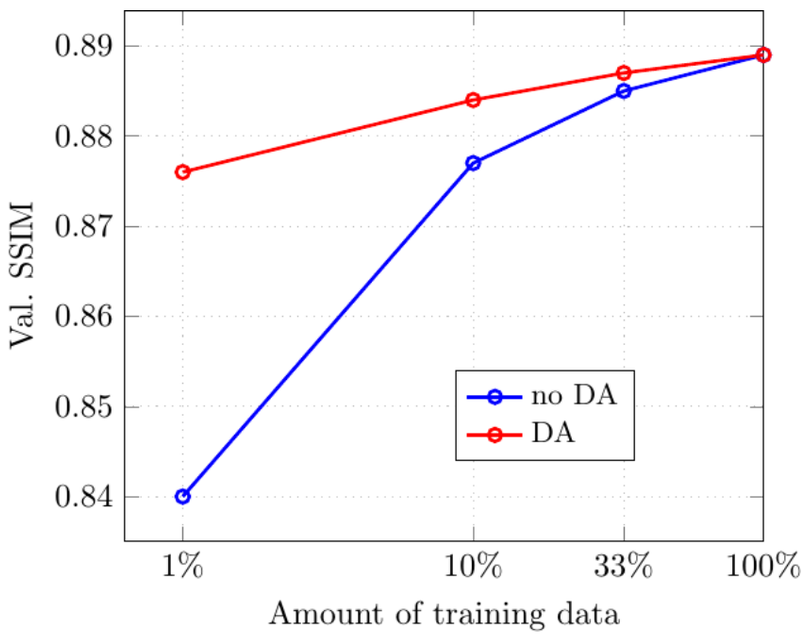
augmented noise!

MRAugment pipeline

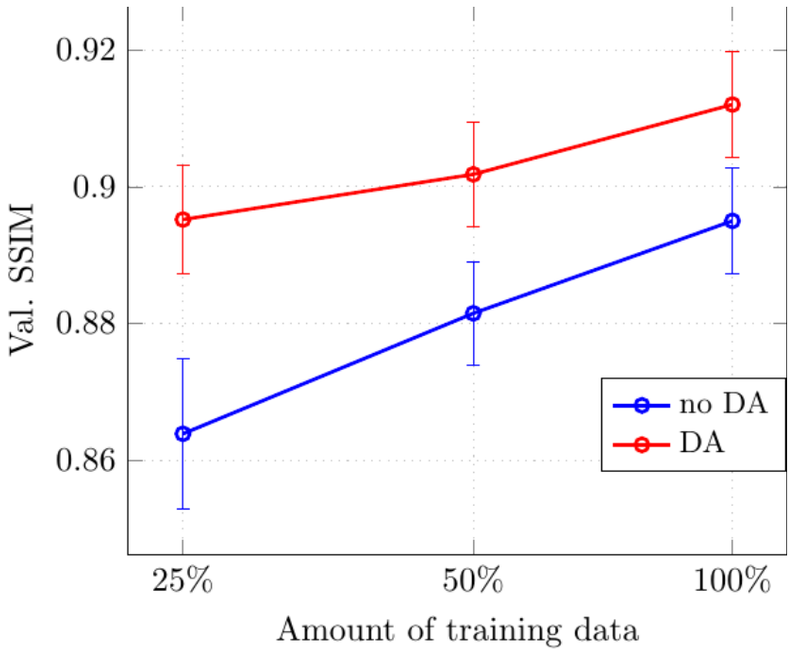


Results on various datasets

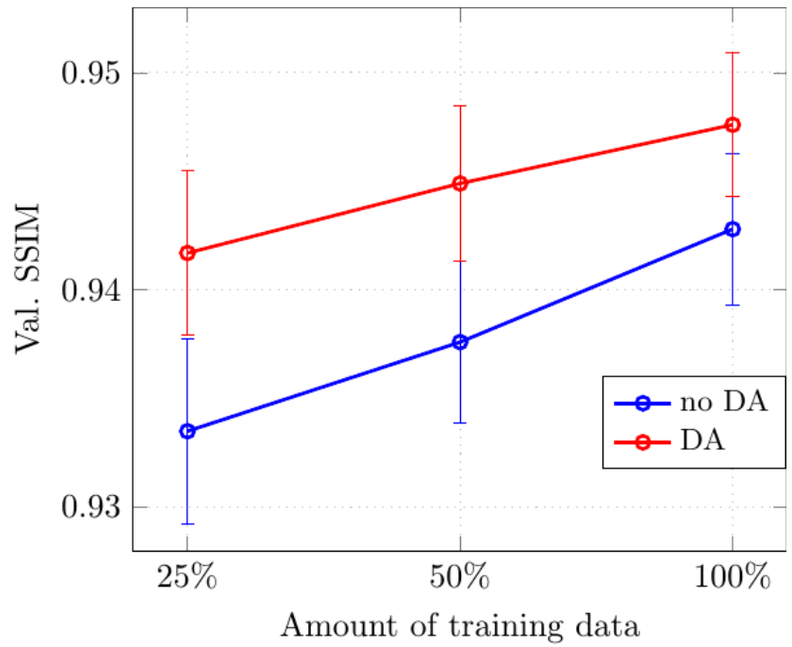
fastMRI knees



Stanford 2D FSE



Stanford 3D FSE knees



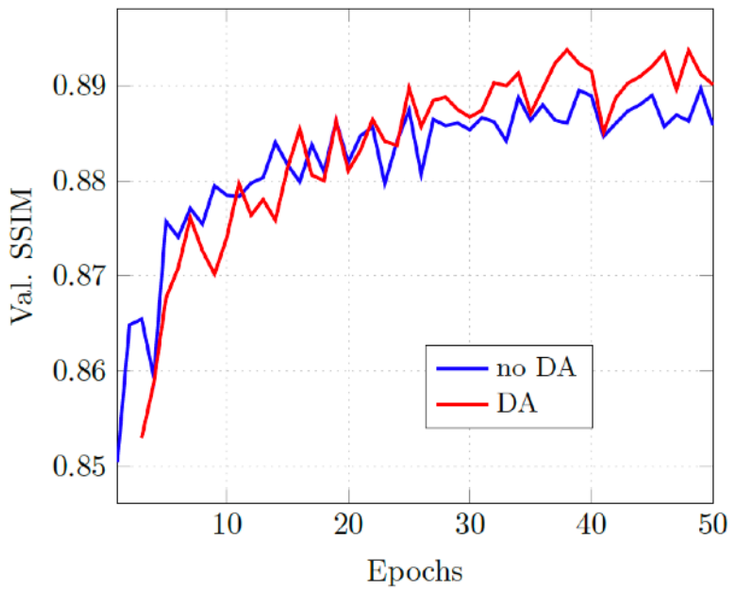
Robustness experiments

Unseen scanners

2% train	no DA	DA
$3T \rightarrow 3T$	0.8646	**0.9049**
$3T \rightarrow 1.5T$	0.8241	**0.8551**
$1.5T \rightarrow 3T$	0.8174	**0.8913**

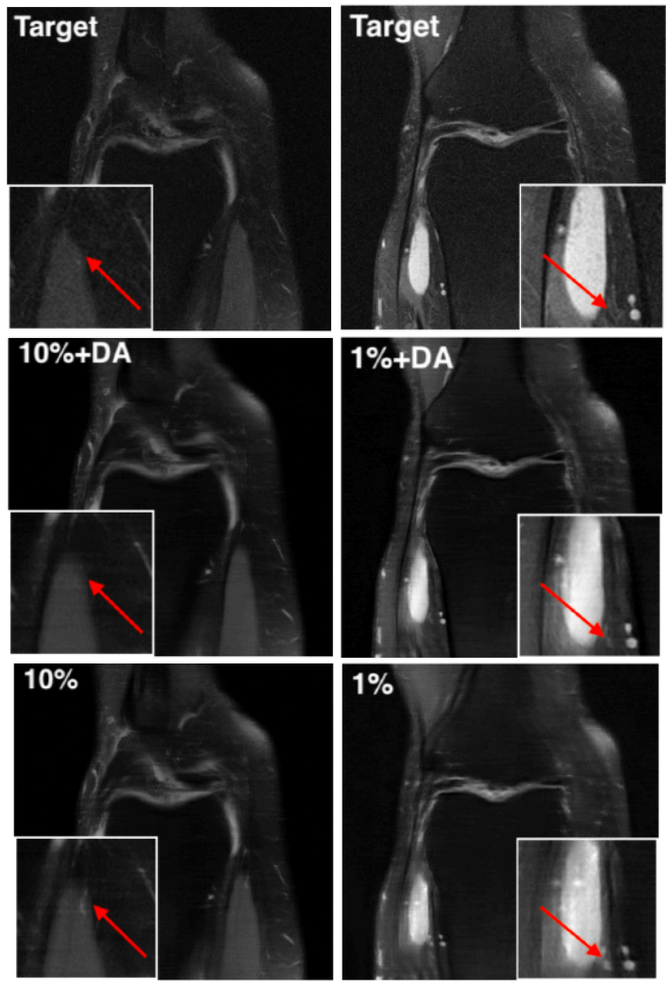
100% train	no DA	DA
$3T \rightarrow 3T$	0.9177	**0.9185**
$3T \rightarrow 1.5T$	0.8686	**0.8690**
$1.5T \rightarrow 3T$	0.9043	**0.9062**

Unseen anatomies



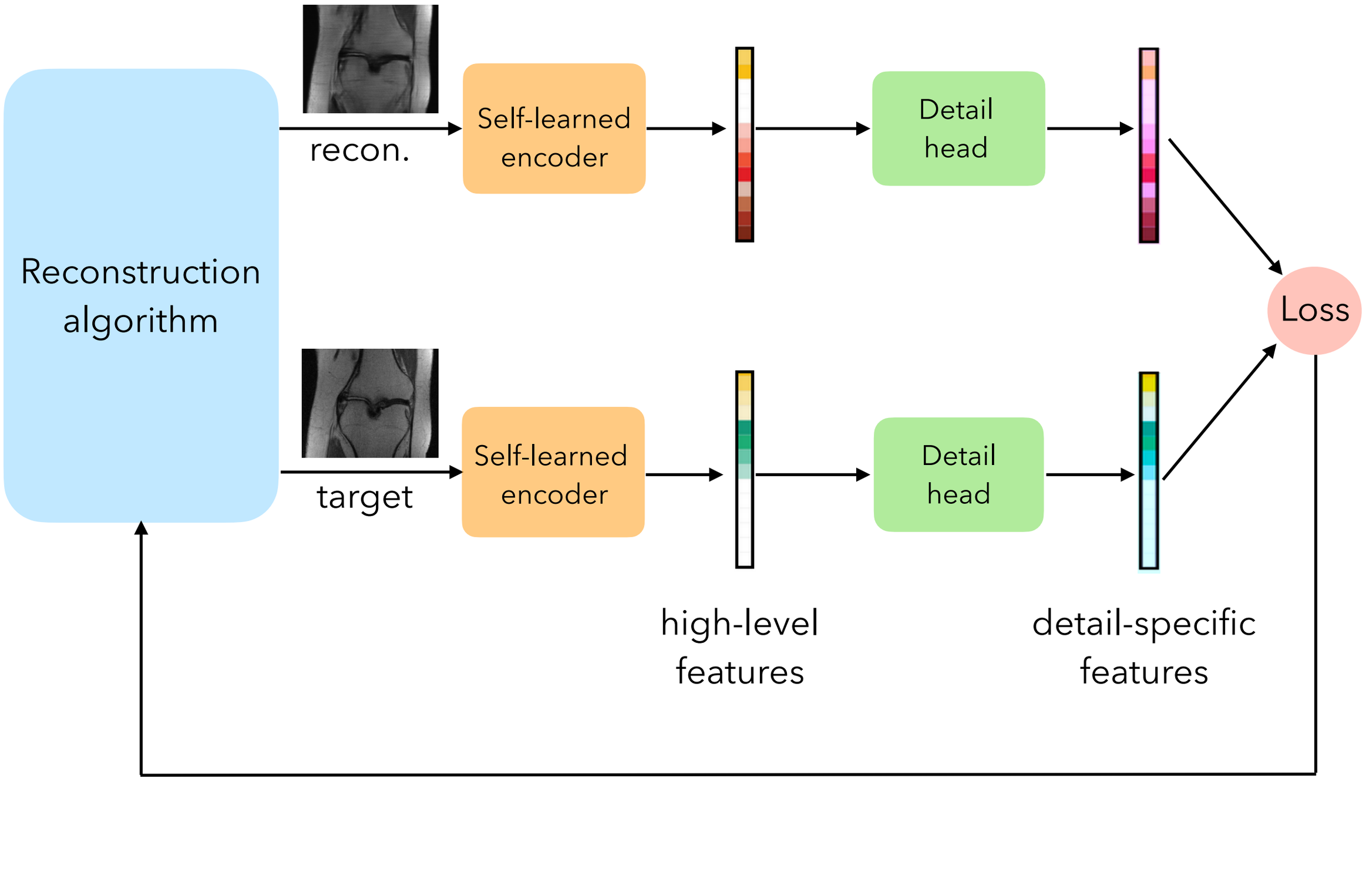
knee → brain

Hallucinations

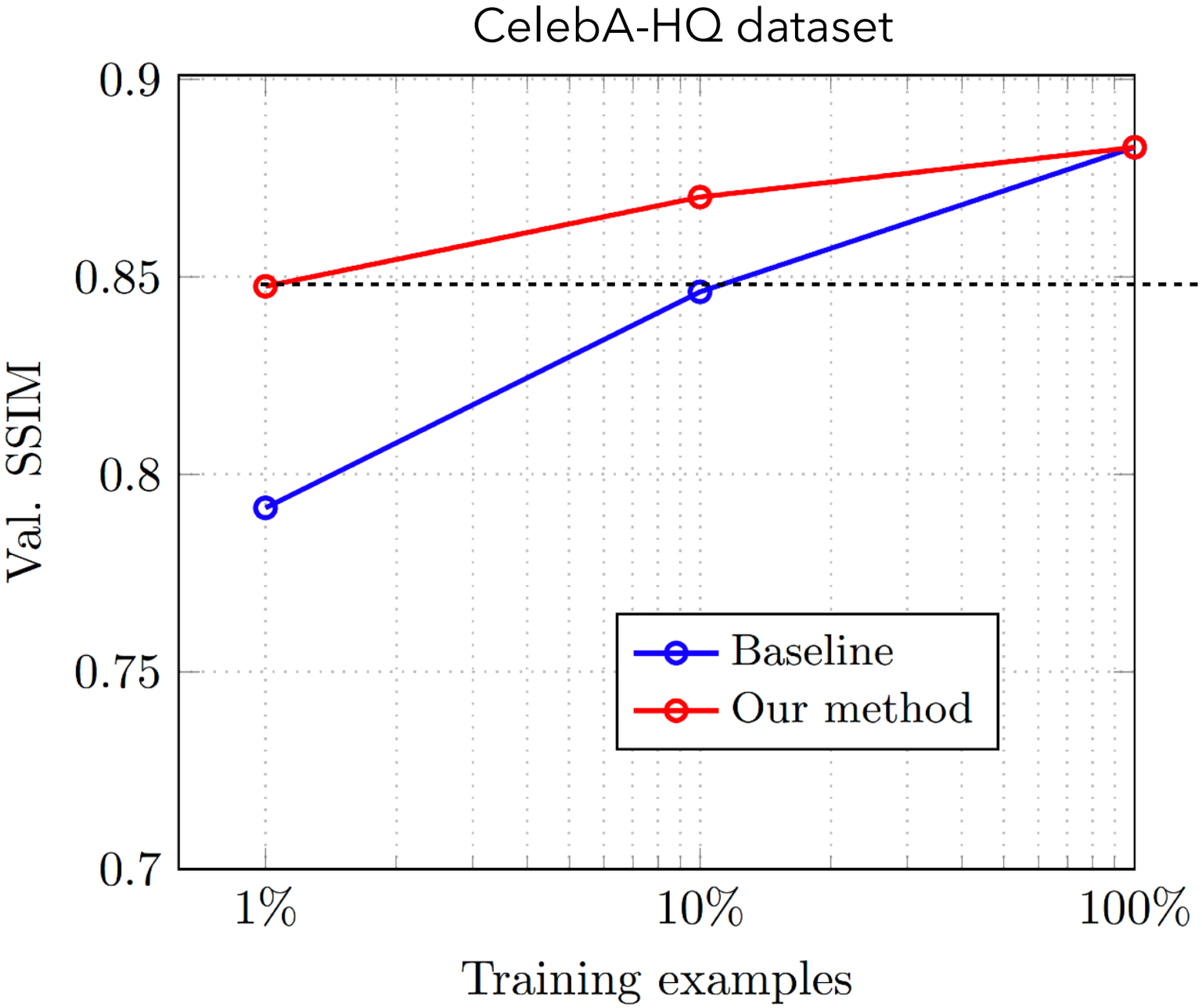


Leveraging self-learned models for data reduction

Detail encoding



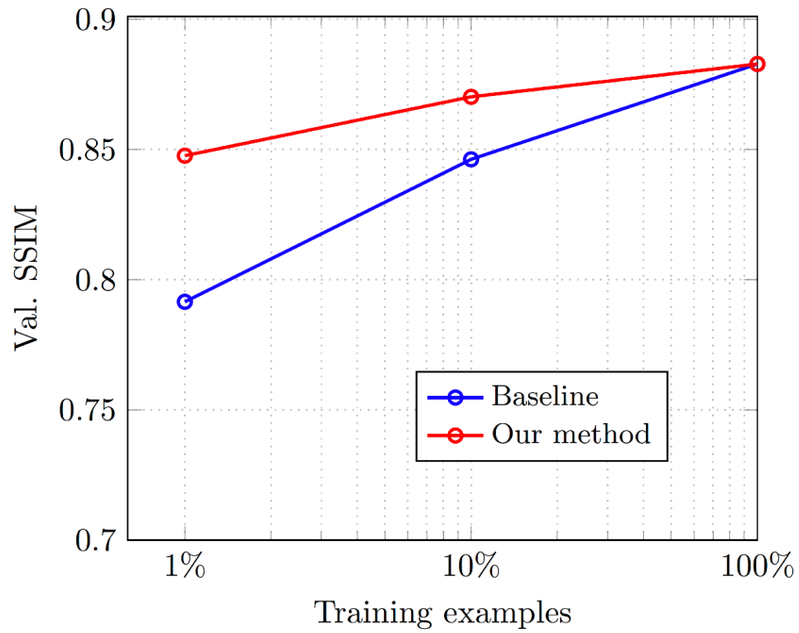
Detail encoding results



10x data reduction!

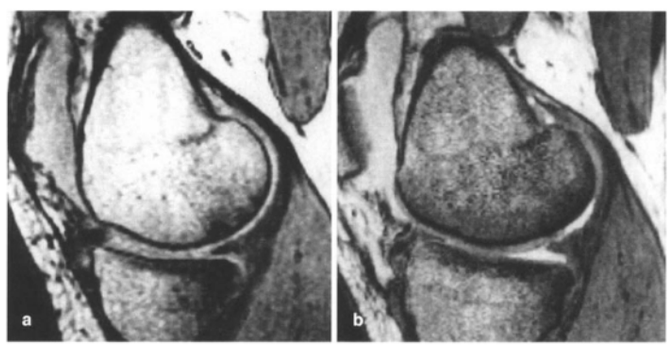
Future work

- Closing the gap

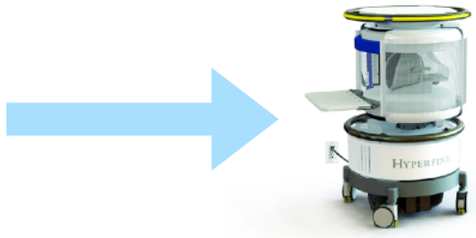


- 10x reduction in low-data regime ✓
- achieve 100% performance with 1-10% data

- Low-field MRI



0.2T 1.5T



Thank you for your attention!

https://github.com/MathFLDS/MRAugment