Unrolling Inference:

The Recurrent Inference Machine

Max Welling

University of Amsterdam / Qualcomm

Canadian Institute for Advanced Research

ML @ UvA

AMLAB Amsterdam Machine Learning Lab

(12fte)

Microsoft Research (2 fte)

Machine Learning
in Amsterdam

SAP

(3fte)

QUVA

(12fte)

PHILIPS

(4fte)

STW connecting innovators

BOSCH

(10fte)

AML4HEALTH

Overview

- Meta learning
- Recurrent Inference Machine
- Application to MRI
- Application radio astronomy
- Conclusions

Learning to learn by gradient descent by gradient descent

2016

Marcin Andrychowicz[1], Misha Denil[1], Sergio Gómez Colmenarejo[1], Matthew W. Hoffman[1], David Pfau[1], Tom Schaul[1], Brendan Shillingford[1,2], Nando de Freitas[1,2,3]

[1]Google DeepMind [2]University of Oxford [3]Canadian Institute for Advanced Research

marcin.andrychowicz@gmail.com
{mdenil,sergomez,mwhoffman,pfau,schaul}@google.com
brendan.shillingford@cs.ox.ac.uk, nandodefreitas@google.com

- Train an optimizer to choose the best parameter updates by solving many optimization problems and learn the patterns.
- Unroll gradient optimizer, then abstract into a parameterized computation graph, e.g. RNN

$$\theta \leftarrow \theta + \eta_t \nabla_\theta F(\theta)$$

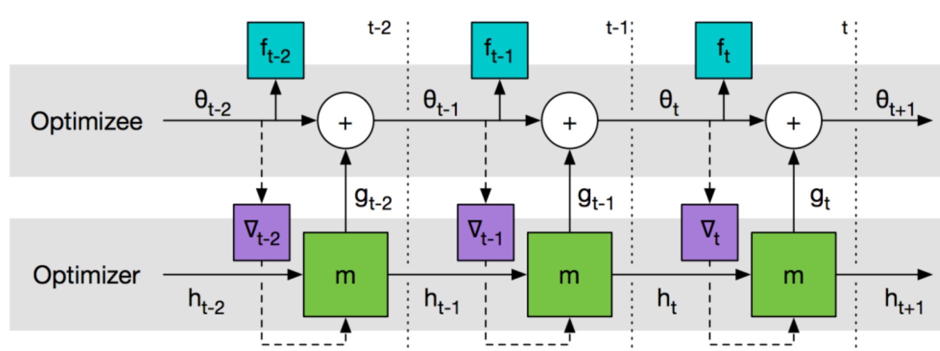


Figure 2: Computational graph used for computing the gradient of the optimizer.

Value Iteration Networks 2017

Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, and Pieter Abbeel

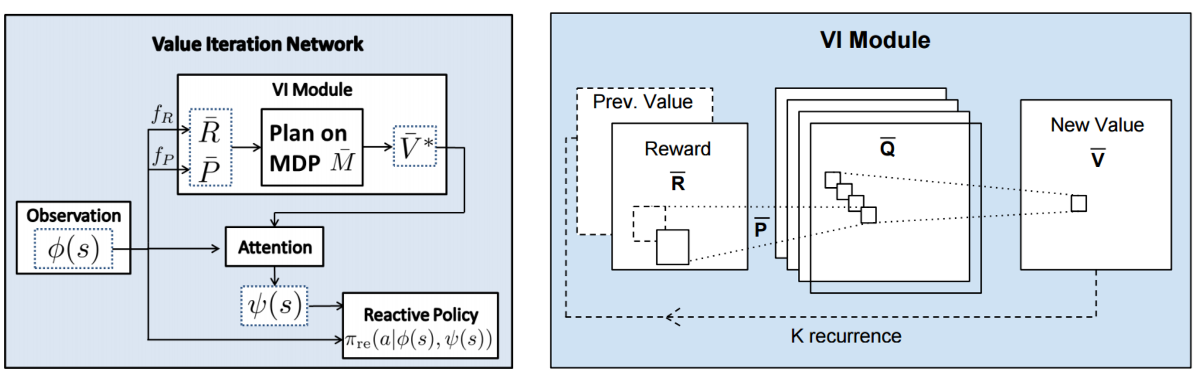
Dept. of Electrical Engineering and Computer Sciences, UC Berkeley

LEARNING TO REINFORCEMENT LEARN

JX Wang[1], Z Kurth-Nelson[1], D Tirumala[1], H Soyer[1], JZ Leibo[1], R Munos[1], C Blundell[1], D Kumaran[1,3], M Botvinick[1,2]
[1]DeepMind, London, UK
[2]Gatsby Computational Neuroscience Unit, UCL, London, UK
[3]Institute of Cognitive Neuroscience, UCL, London, UK

{wangjane, zebk, dhruvat, soyer, jzl, munos, cblundell, dkumaran, botvinick} @google.com

- Learning a planning algorithm to execute the best actions by solve many different RL.

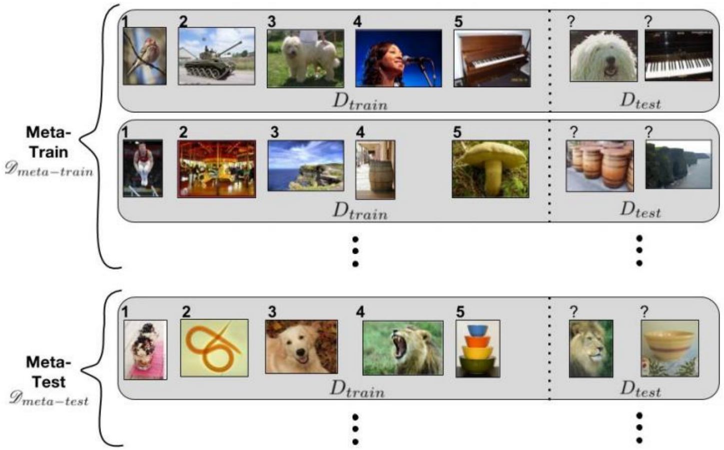


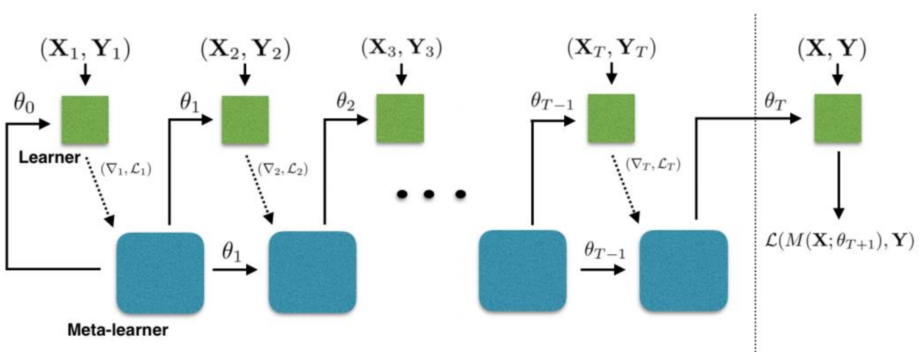
Optimization as a Model for Few-Shot Learning

2017

Sachin Ravi* and Hugo Larochelle
Twitter, Cambridge, USA
{sachinr,hugo}@twitter.com

- One shot learning: meta-learn a learning algorithm to classify from very few examples



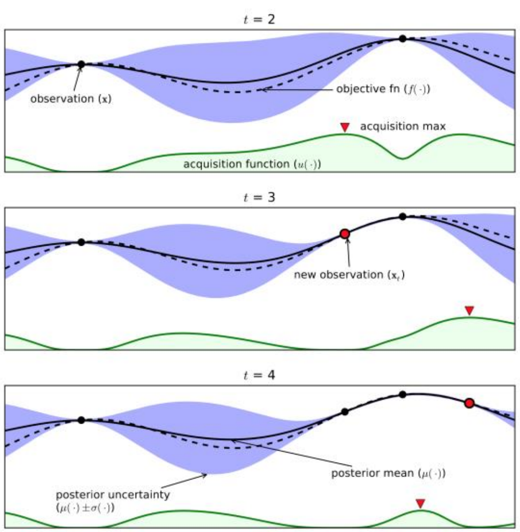


NEURAL ARCHITECTURE SEARCH WITH REINFORCEMENT LEARNING

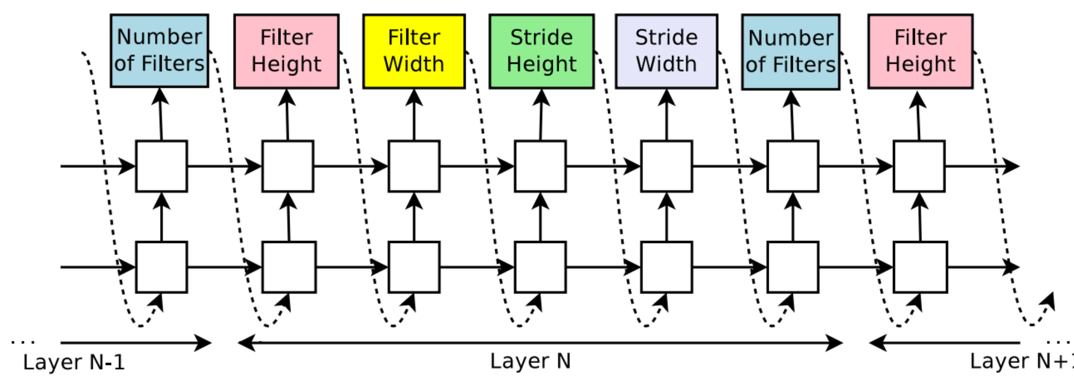
2017

Barret Zoph,* Quoc V. Le
Google Brain
{barretzoph,qvl}@google.com

- Learning a NN architecture using active learning / reinforcement learning



Bayesian optimization

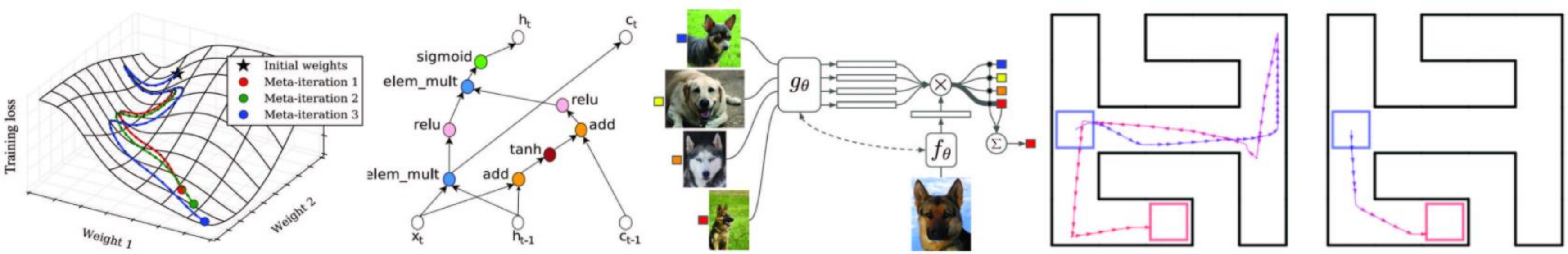


BAIR

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Learning to Learn

Chelsea Finn Jul 18, 2017



hyperparameter optimization
Maclaurin et al. '15

learned recurrent cell
Zoph & Le '17

few-shot image classifier
Vinyals et al. '16

learning to quickly navigate new mazes
Duan et al. '16

Various recent meta-learning approaches.

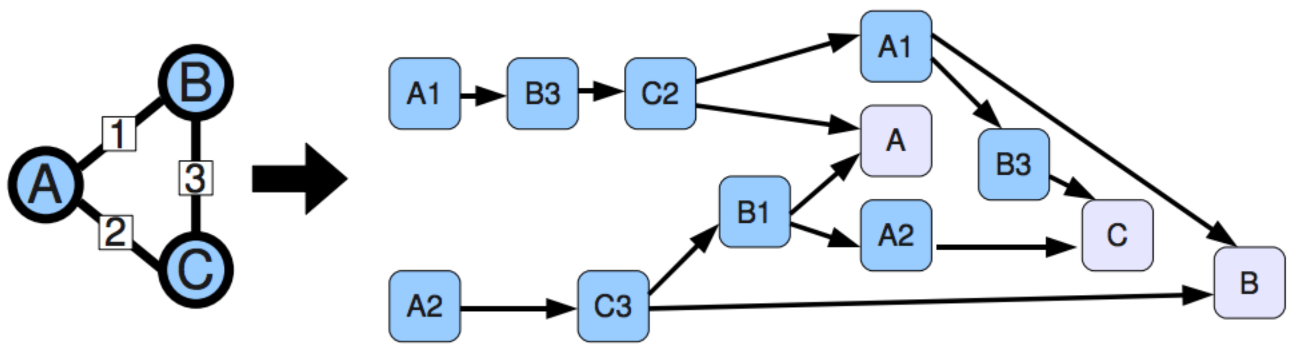
The Recipe

- Study the classical iterative optimization algorithm
- Unroll the computation tree and cut it off at T steps (layers)
- Generalize / parameterize the individual steps
- Create targets at the last layer
- Backpropagate through the ”deep network” to fit the parameters

- Execute the network to make predictions

Learning to Infer

- Unroll a known iterative inference scheme (e.g. mean field, belief propagation)
- Abstract into parameterized computation graph for fixed nr. iterations, e.g. RNN
- Learn parameters of RNN using meta-learning (e.g. solving many inference problems)

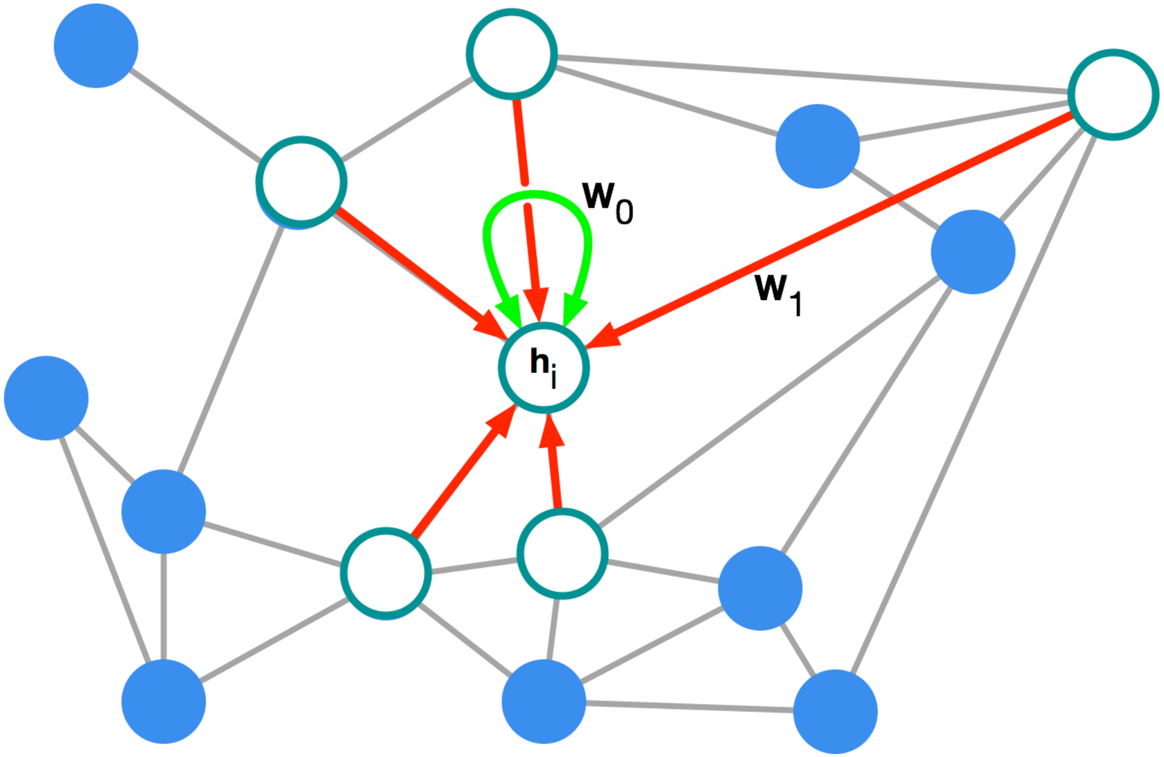


Learning Message-Passing Inference Machines for Structured Prediction

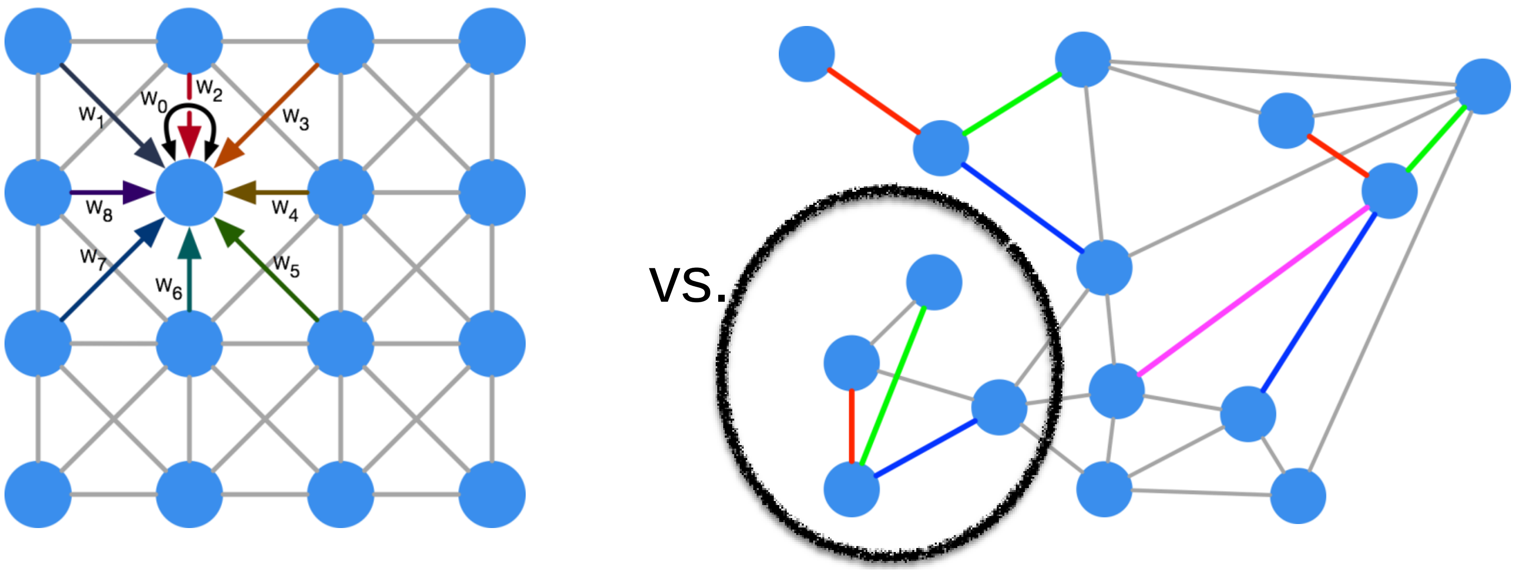
Stéphane Ross Daniel Munoz Martial Hebert J. Andrew Bagnell
The Robotics Institute, Carnegie Mellon University
stephaneross@cmu.edu, {dmunoz, hebert, dbagnell}@ri.cmu.edu

Graph Convolutions

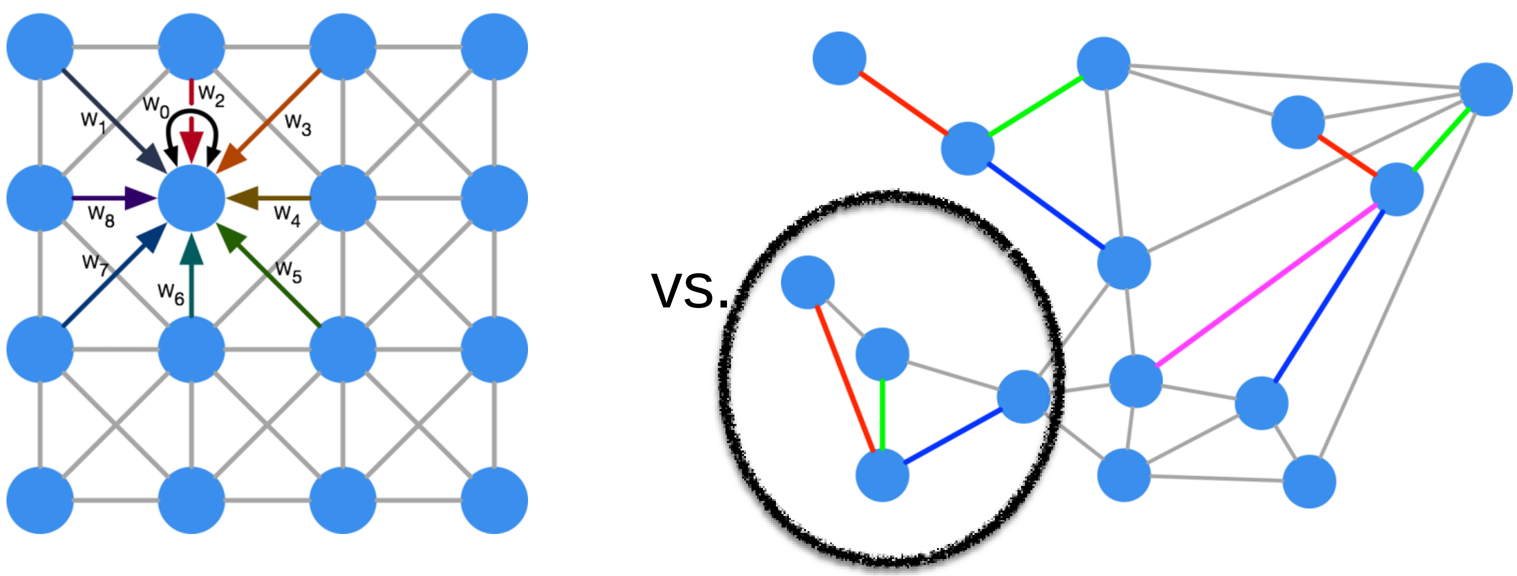
Thomas Kipf



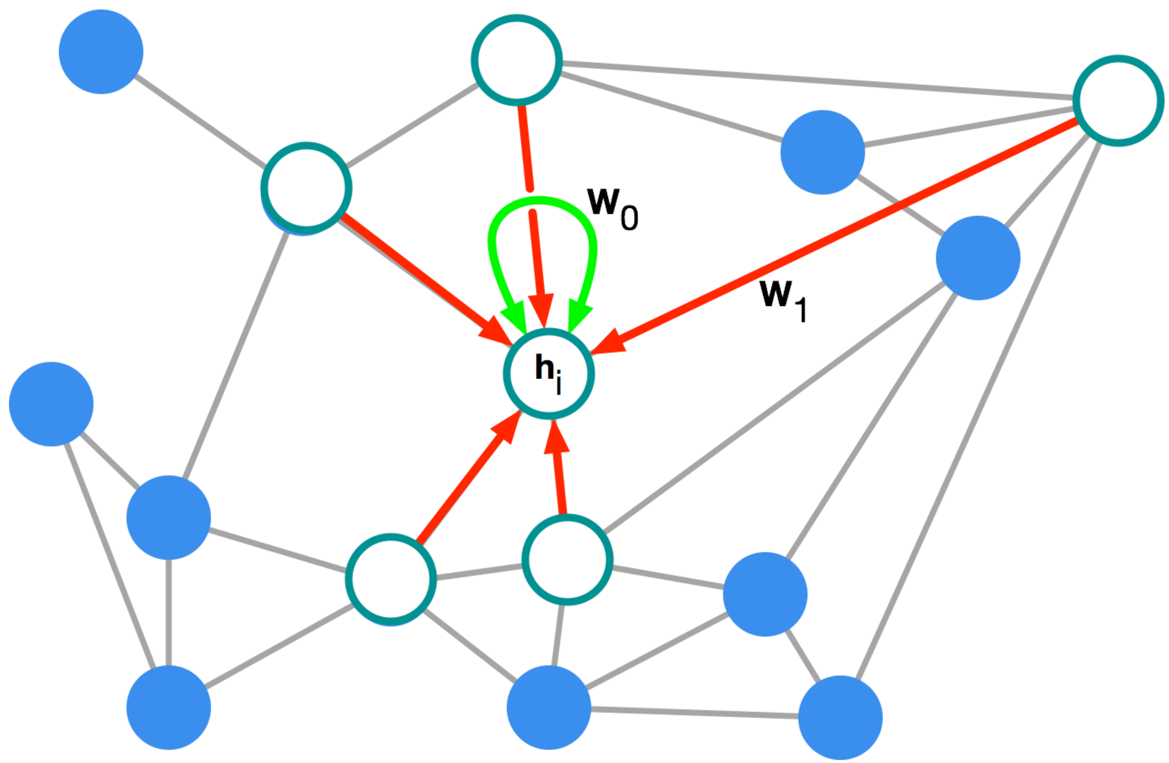
Convolutions vs Graph Convolutions



Convolutions vs Graph Convolutions



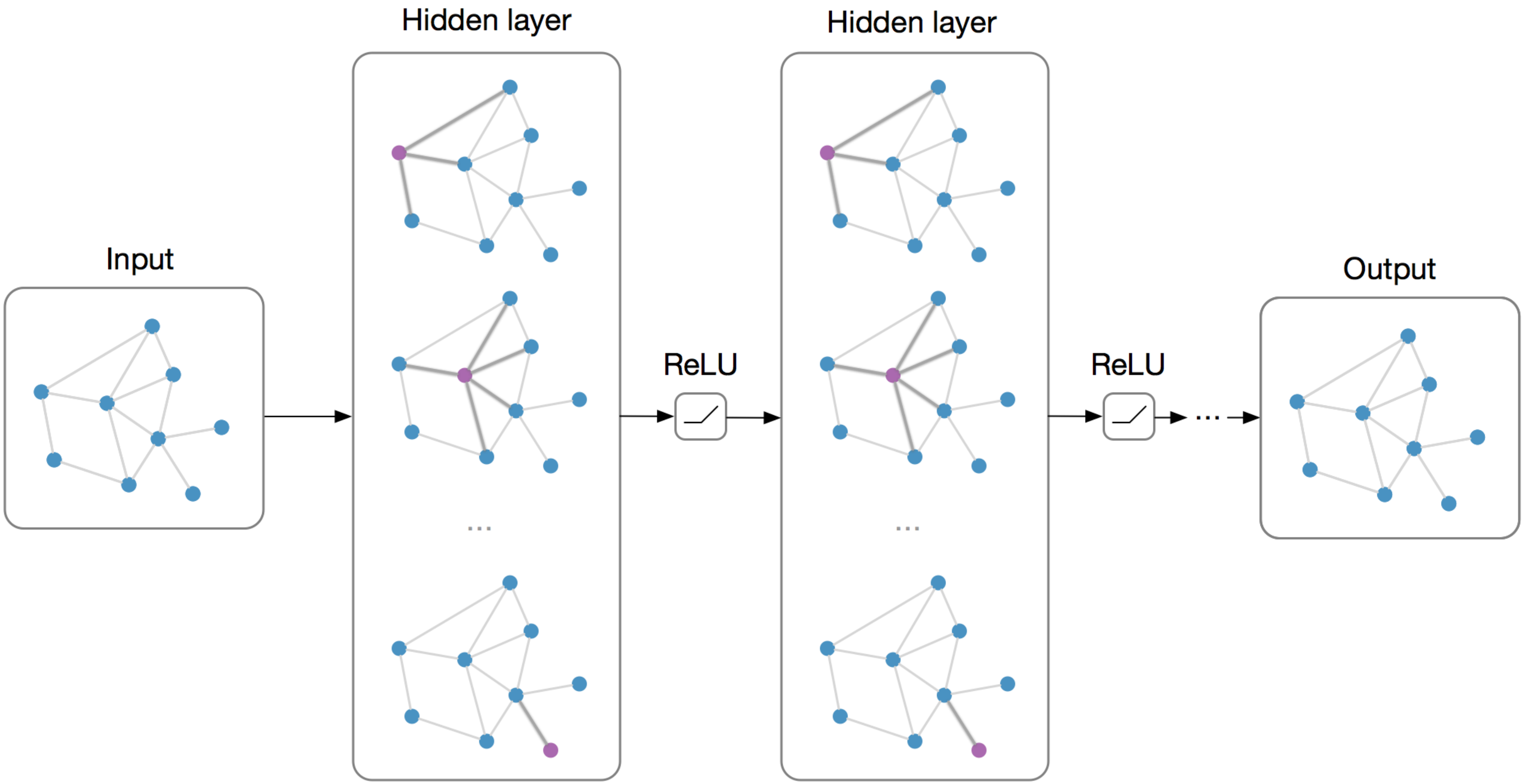
Graph Convolutions



$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{h}_i^{(l)}\mathbf{W}_0^{(l)} + \sum_{j\in\mathcal{N}_i}\frac{1}{c_{ij}}\mathbf{h}_j^{(l)}\mathbf{W}_1^{(l)}\right)$$

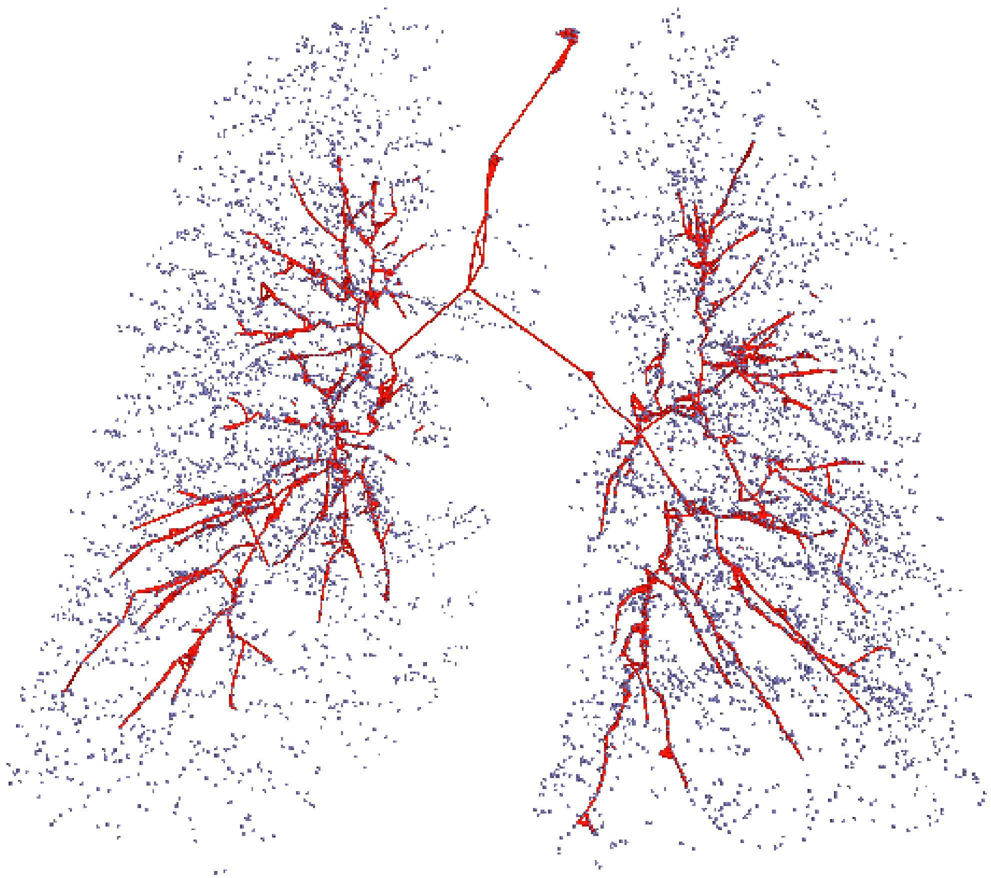
Graph Convolutional Networks

Kipf & Welling ICLR (2017)



Application to Airway Segmentation

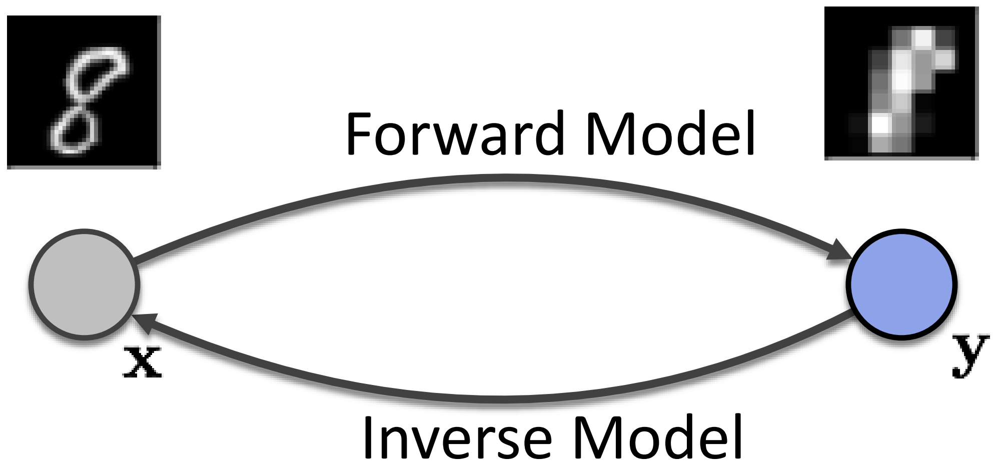
(work in progress, with Rajhav Selvan & Thomas Kipf)



Inverse Problems

w/ Patrick Putzky

Quantity of interest Measurement



Forward Model $\mathbf{y} = g(\mathbf{x}) + n$

Inverse Model $\hat{\mathbf{x}} = h(\mathbf{y})$

The Usual Approach

$$L(X) = \log P_A(Y|X) + \log P_\theta(X)$$
$$Y = A \cdot X + \eta$$

prior (learn)

observations

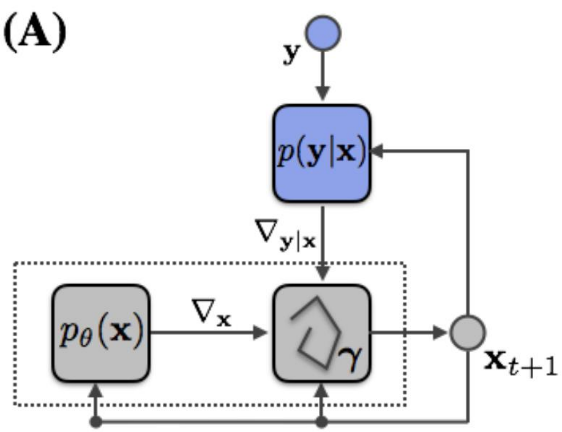
generative model (known)

$$X_{t+1} = X_t + \alpha_t(\nabla_X \log P_A(Y|X_t) + \nabla_X \log P_\theta(X_t))$$

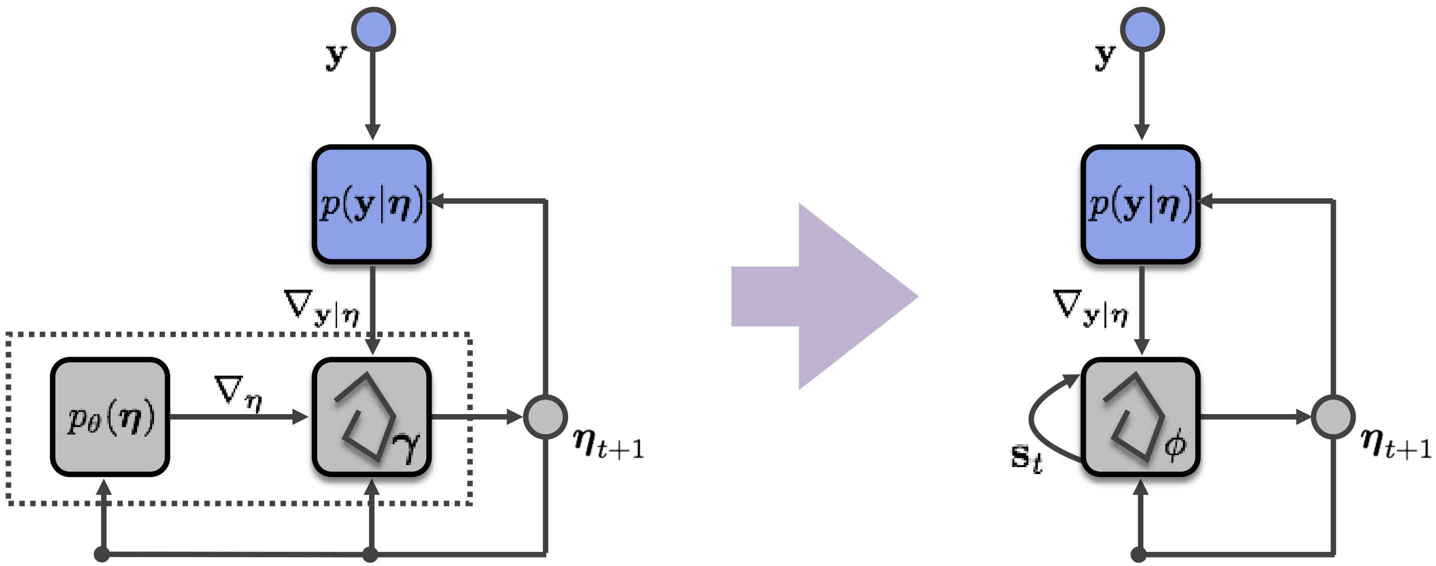
advantage: model P(X) and optimization are separated.

disadvantage: accuracy suffers because
model and optimization interact...

(A)



Learning Inference: Recurrent Inference Machine



$$\boldsymbol{\eta}_{t+1} = \boldsymbol{\eta}_t + \gamma_t(\nabla_{\mathbf{y}|\boldsymbol{\eta}} + \nabla_{\boldsymbol{\eta}})$$

$$\boldsymbol{\eta}_{t+1} = \boldsymbol{\eta}_t + h_\phi(\nabla_{\mathbf{y}|\boldsymbol{\eta}}, \boldsymbol{\eta}_t, \mathbf{s}_t)$$

- Abstract and parameterize computation graph into RNN
- Integrate prior P(X) in RNN
- Add memory state s
- Meta learn the parameters of the RNN

Recurrent Inference Machine (RIM)

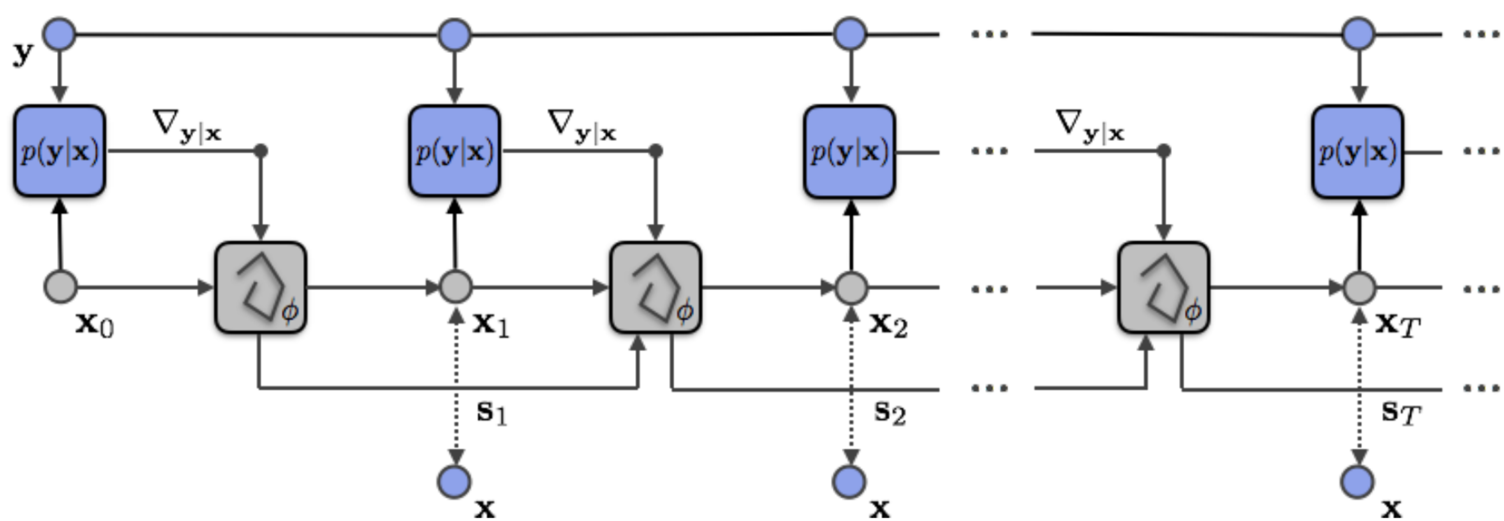
Learn to optimize using a RNN.

$$X_{t+1} = F_\theta(X_t, \nabla_x \log P_A(Y|X_t), S_t)$$
$$S_{t+1} = G_\nu(X_t, \nabla_x \log P_A(Y|X_t), S_t)$$

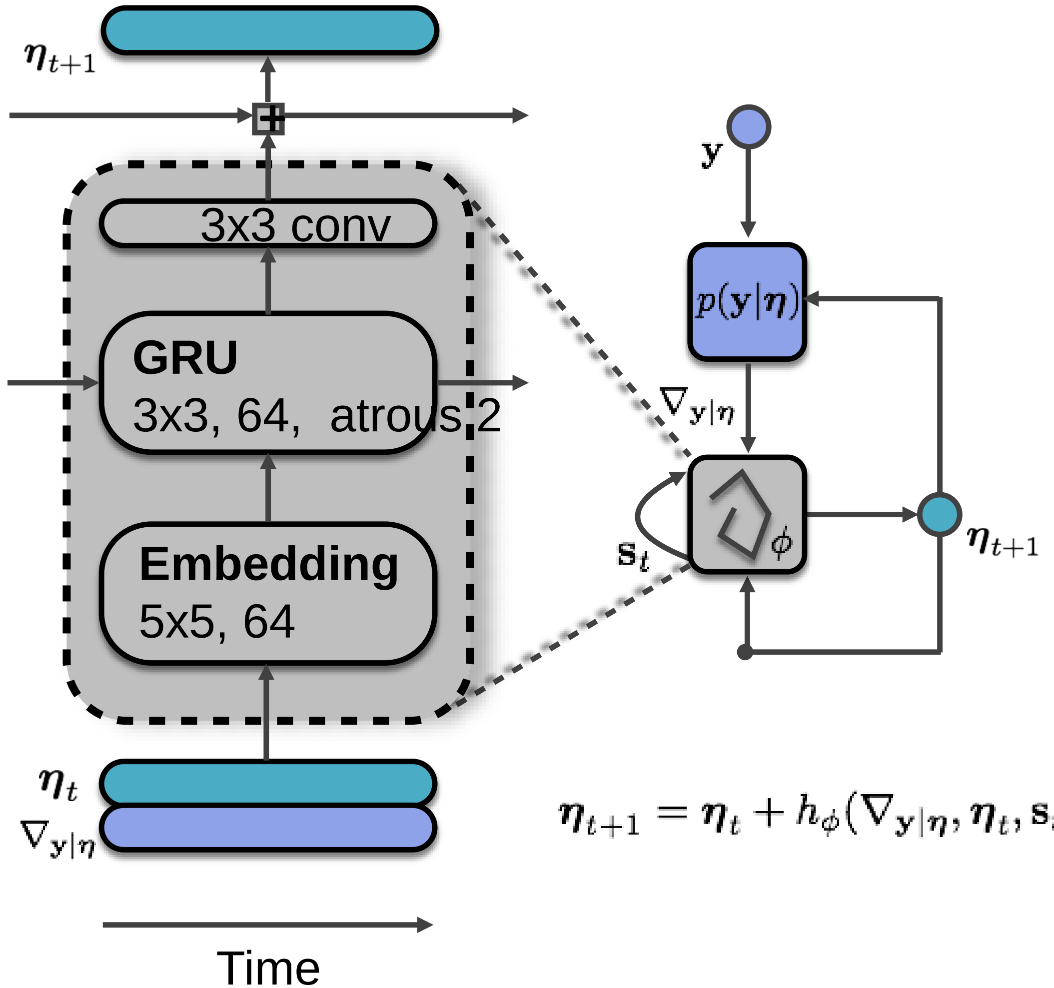
CNN/RNN

external information

memory state

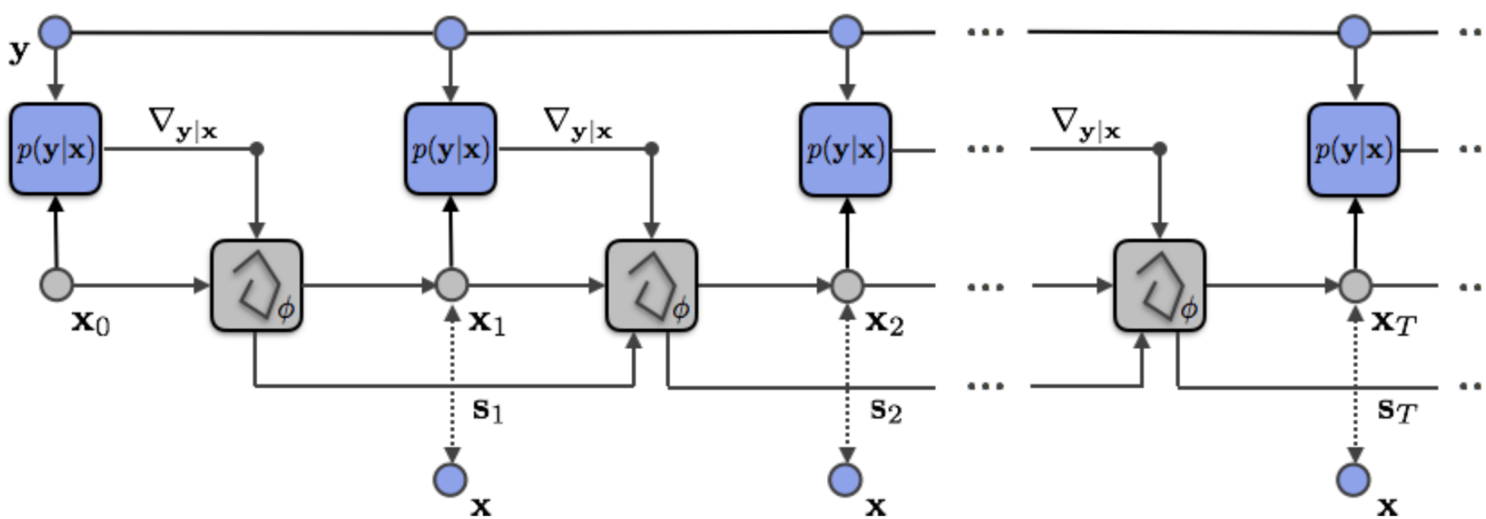


Recurrent Inference Machine



$$\eta_{t+1} = \eta_t + h_\phi(\nabla_{\mathbf{y}|\eta}, \eta_t, \mathbf{s}_t)$$

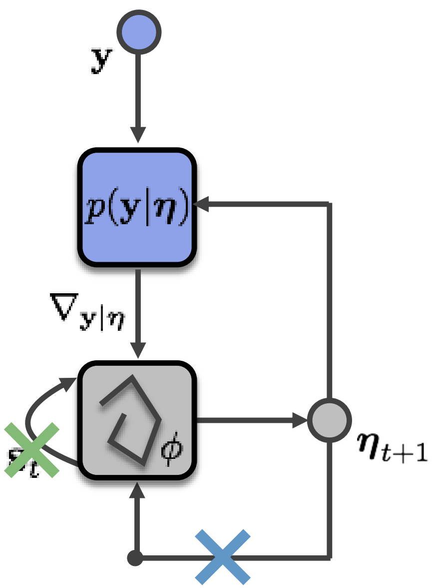
Recurrent Inference Machines in Time

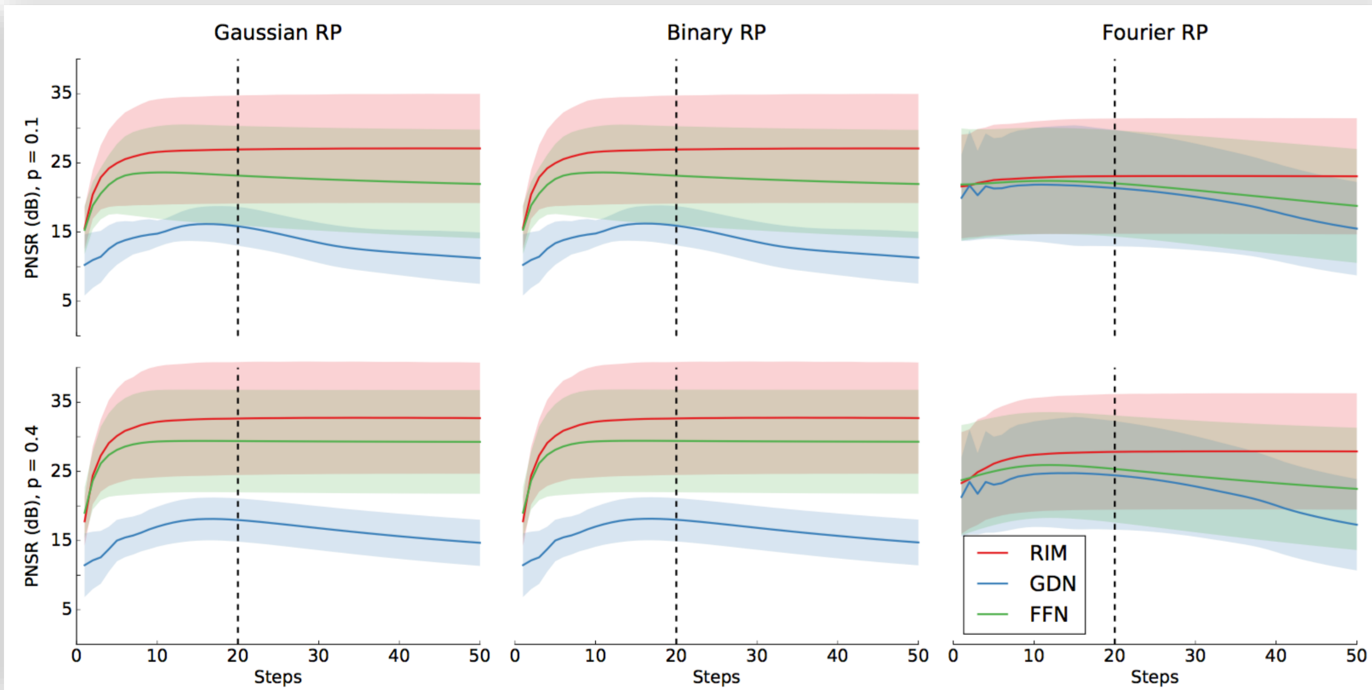


Objective $g(\phi) = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}^{(i)} - \hat{\mathbf{x}}_t^{(i)})$

Simple Super-Resolution

Reconstruction from Random Projections

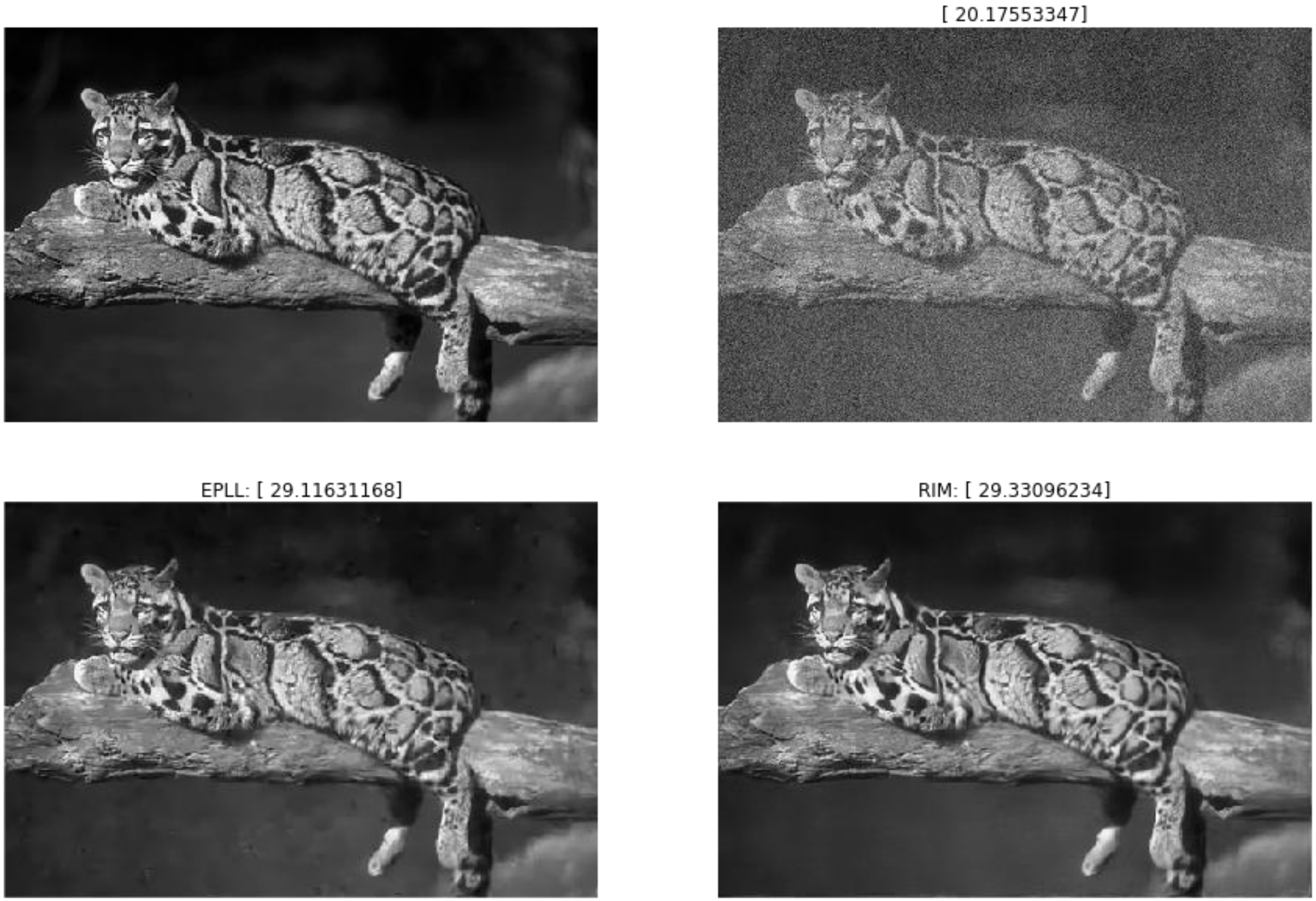




32 x 32 pixel image patches

Fast Convergence on all tasks

Image Denoising



Denoising trained on small image patches, generalises to full-sized images

Super-resolution

LR | HR | Bicubic Interpolation | RIM

Super-resolution

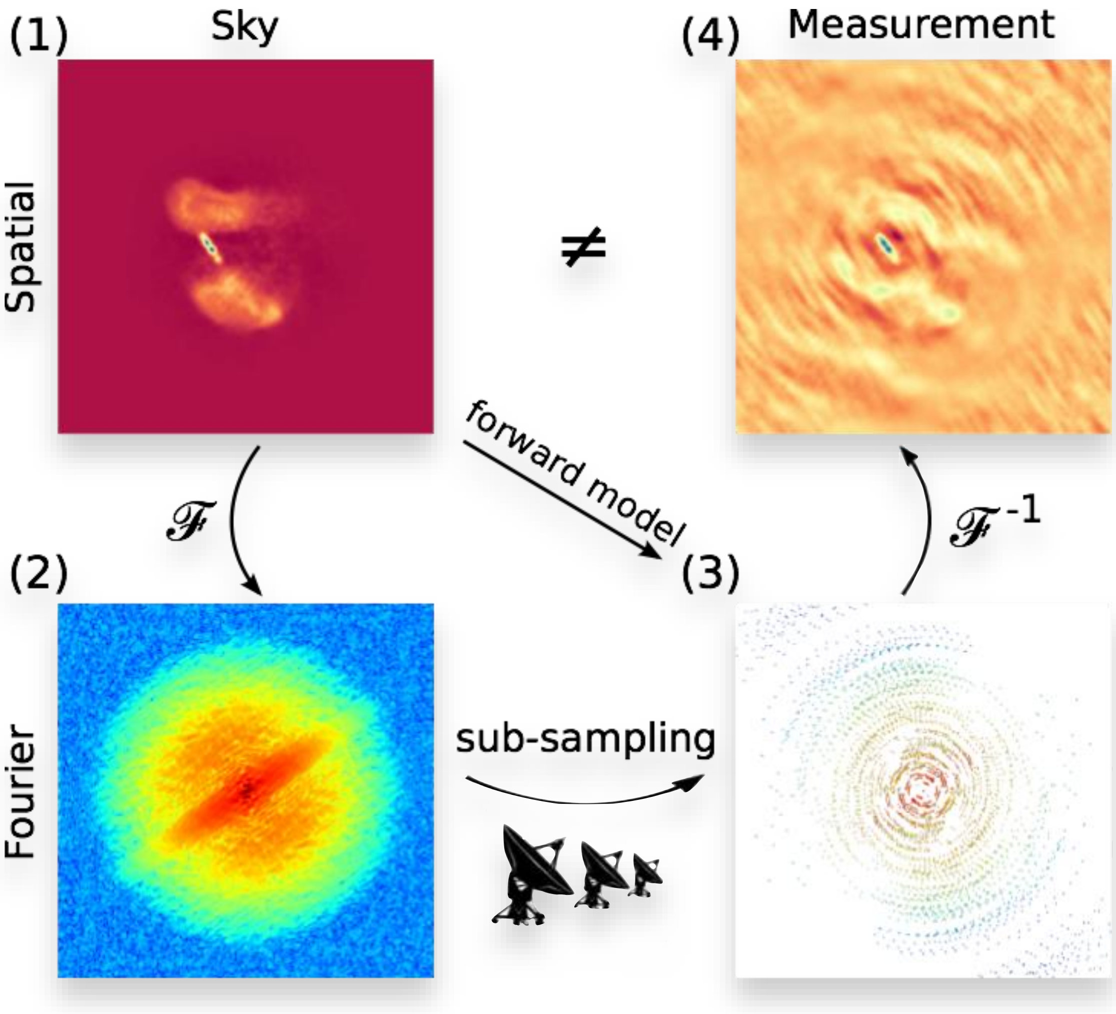
(a) Original Image (b) Bicubic: 30.43/0.8326 (c) SRCNN: 31.34/0.8660

(d) A+: 31.43/0.8676 (e) SelfExSR: 31.18/0.8656 (f) RIM: **31.59**/**0.8712**

Metric	Scale	Bicubic	SRCNN	A+	SelfExSR	**RIM (Ours)**
PSNR	2x	29.55 ± 0.35	31.11 ± 0.39	31.22 ± 0.40	31.18 ± 0.39	$\mathbf{31.39} \pm 0.39$
	3x	27.20 ± 0.33	28.20 ± 0.36	28.30 ± 0.37	28.30 ± 0.37	$\mathbf{28.51} \pm 0.37$
	4x	25.96 ± 0.33	26.70 ± 0.34	26.82 ± 0.35	26.85 ± 0.36	$\mathbf{27.01} \pm 0.35$
SSIM	2x	0.8425 ± 0.0078	0.8835 ± 0.0062	0.8862 ± 0.0063	0.8855 ± 0.0064	$\mathbf{0.8885} \pm 0.0062$
	3x	0.7382 ± 0.0114	0.7794 ± 0.0102	0.7836 ± 0.0104	0.7843 ± 0.0104	$\mathbf{0.7888} \pm 0.0101$
	4x	0.6672 ± 0.0131	0.7018 ± 0.0125	0.7089 ± 0.0125	0.7108 ± 0.0124	$\mathbf{0.7156} \pm 0.0125$

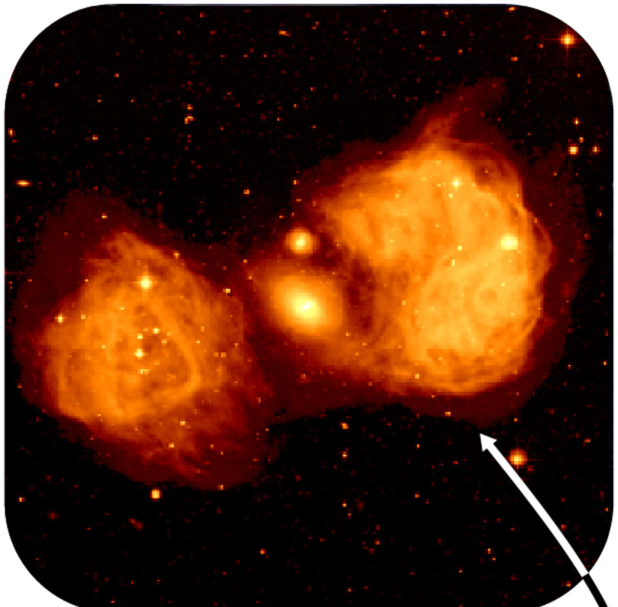
Square Kilometer Array

Jorn Peters

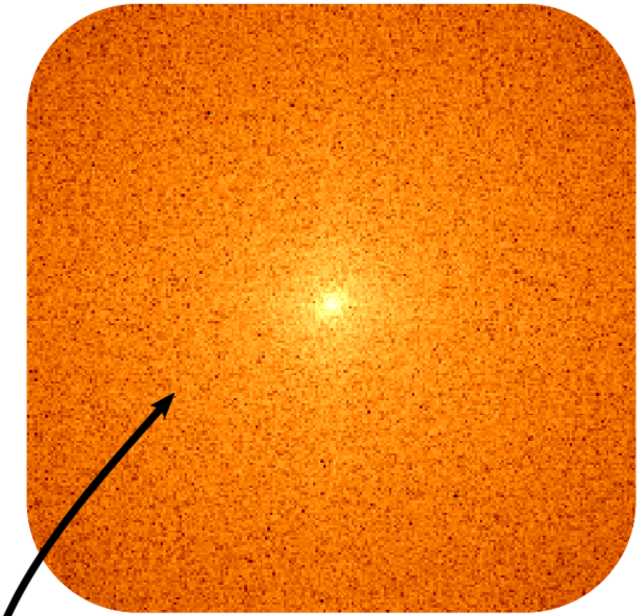


Up to 14.4 Gigapixels
With thousands of Channels

Radio Astronomy: Observations



$\overset{\mathcal{F}}{\rightleftharpoons}$



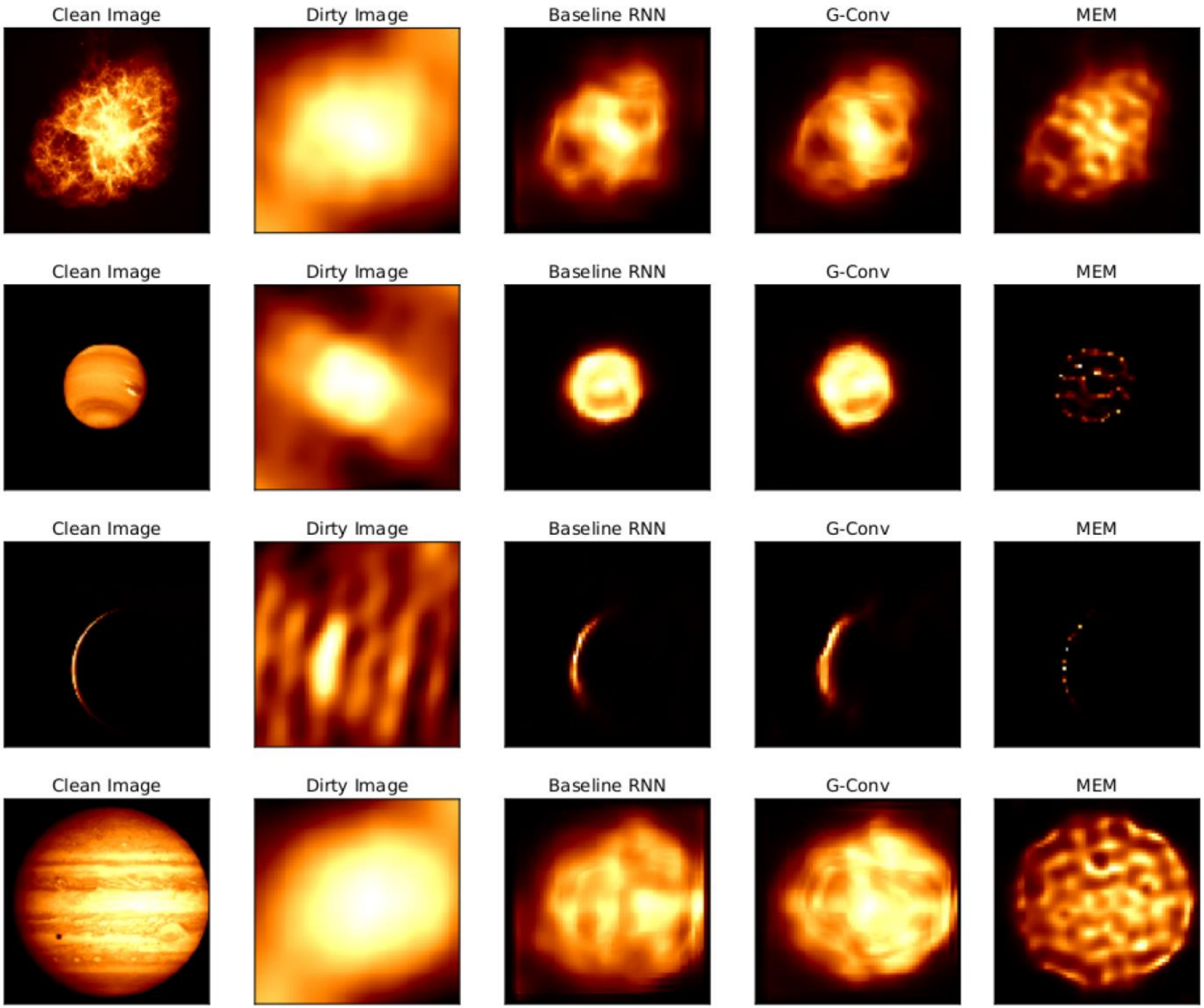
Intensity Map

UV-plane

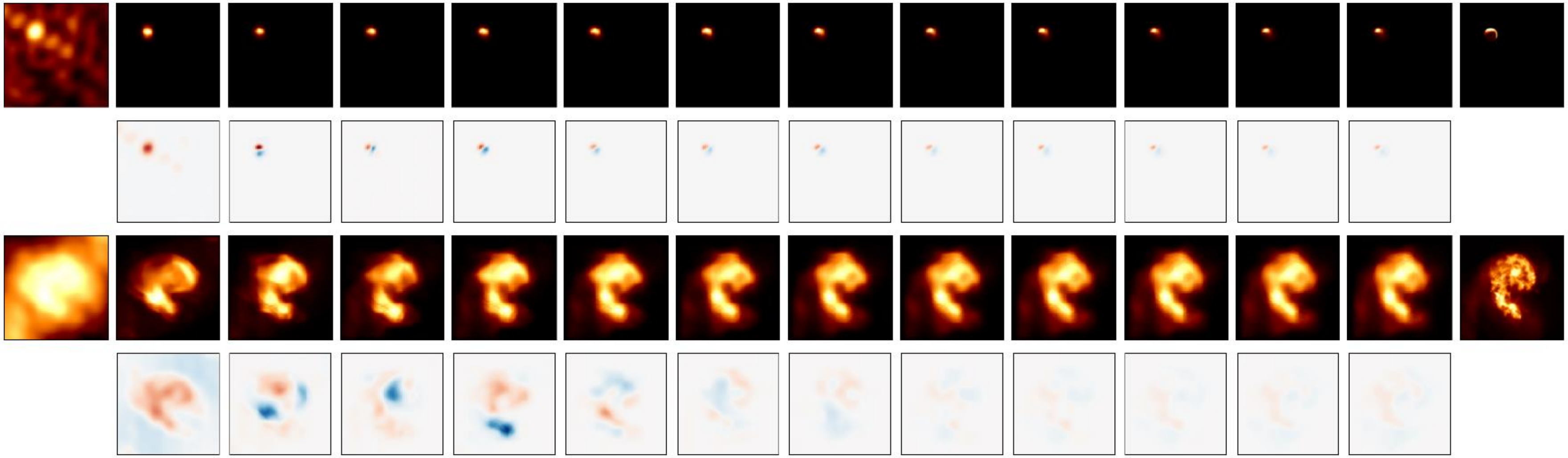
$\mathcal{V}(u,v)$

$I(l,m)$

Experiments: VLBI Inverse Models

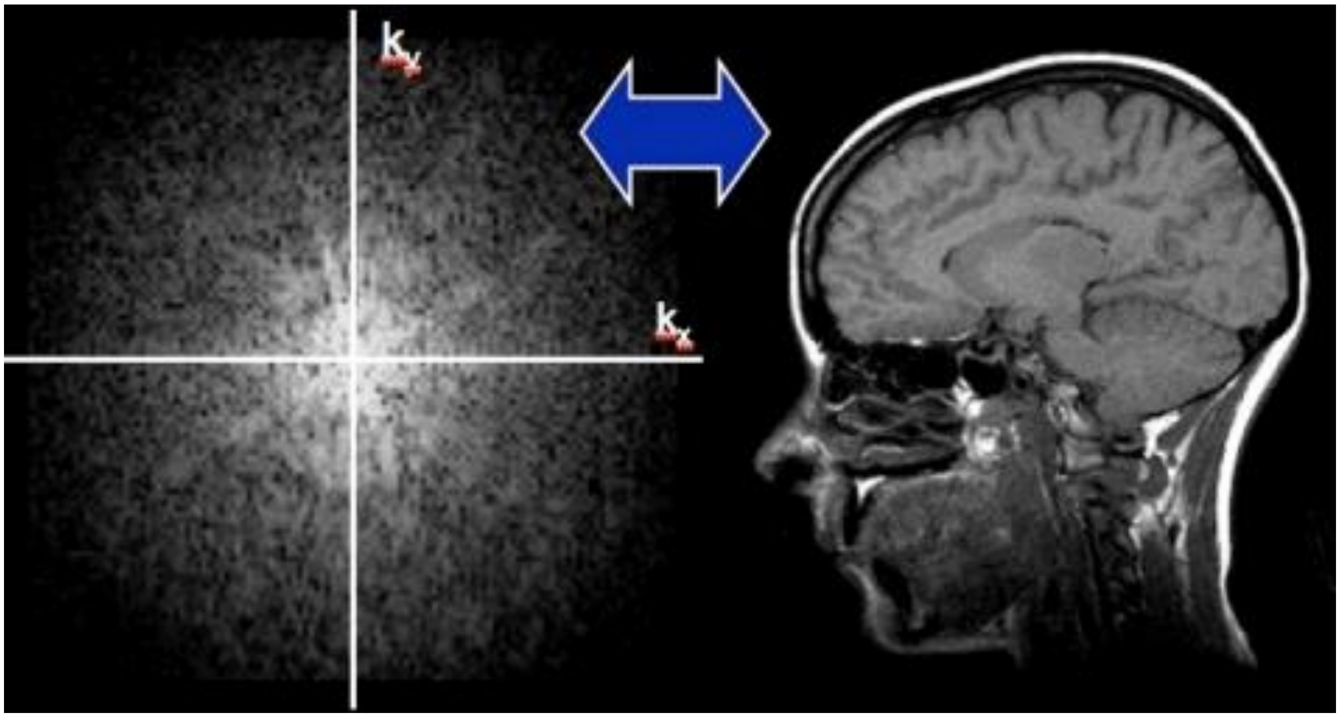


Direct Fourier Transform Forward Model



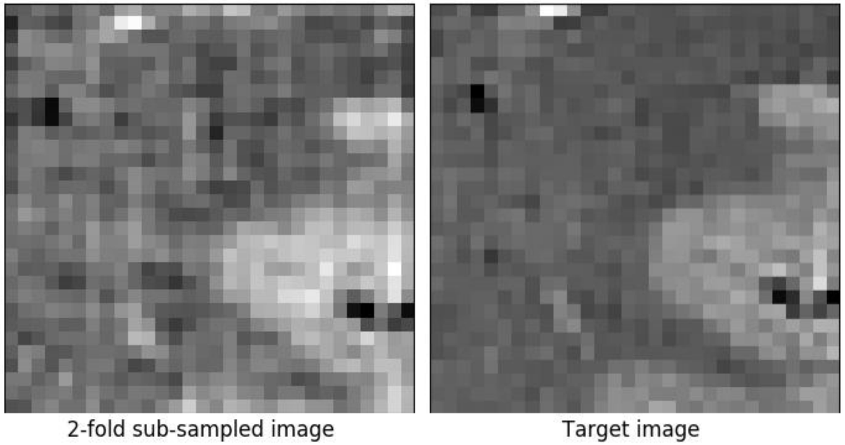
Deep Learning for Inverse Problems

w/ Kai Lonning & Matthan Caan

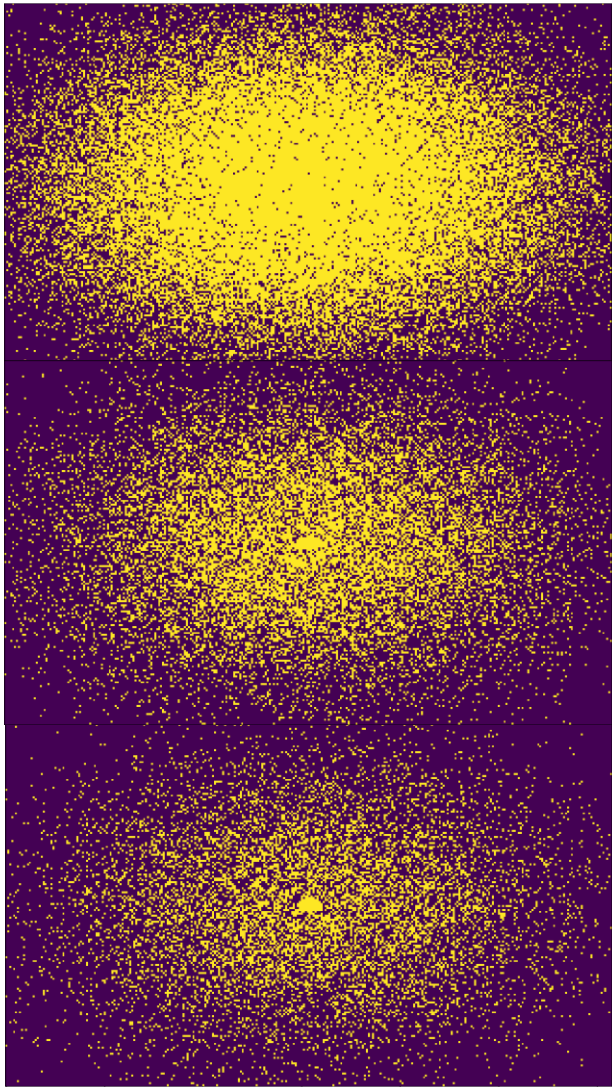


E.g. MRI Image Reconstruction

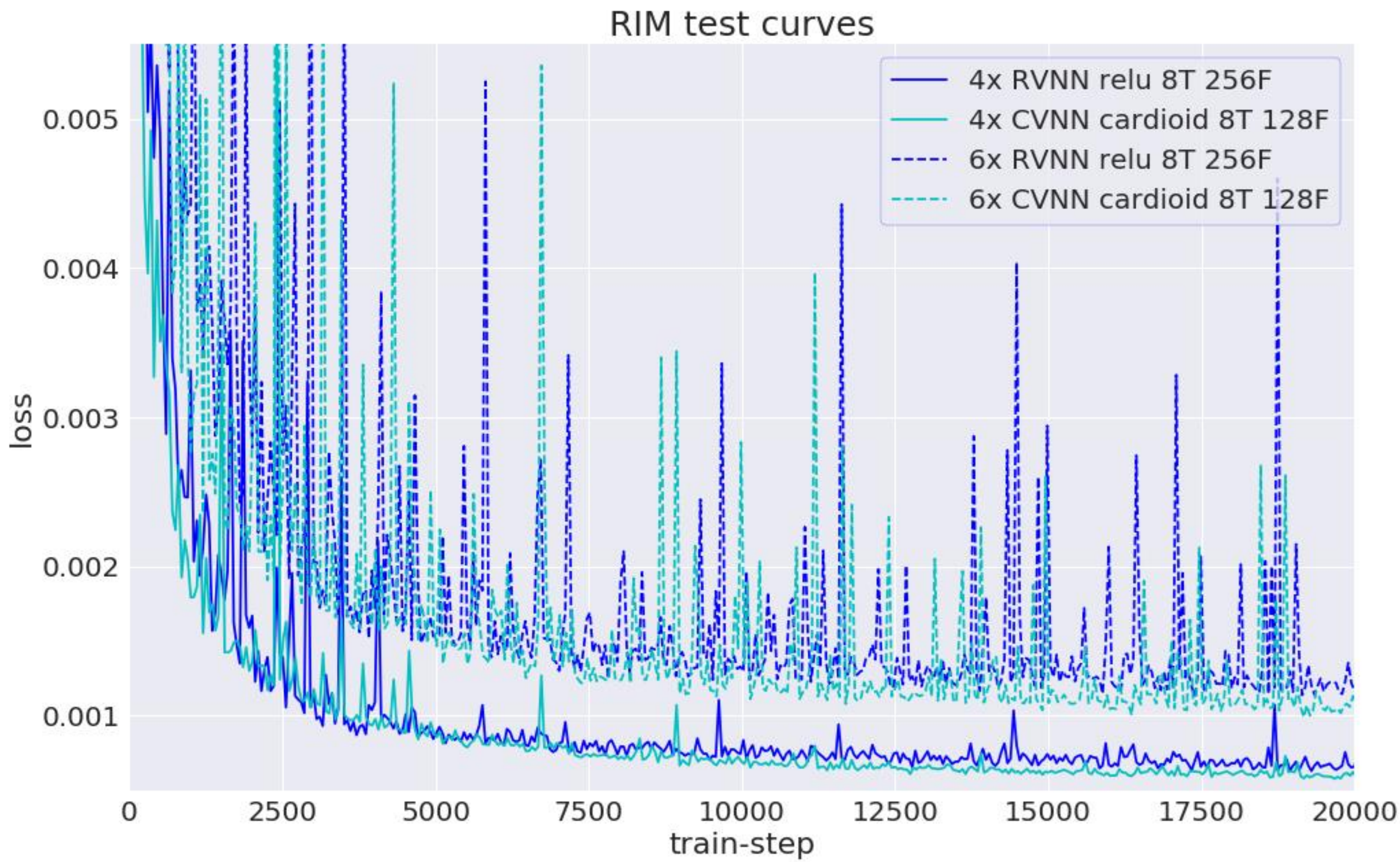
http://sbt.science.uva.nl/mri/about/ (Slides and website made by Kai Lonning)



Example of training data point, 30x30 image patch

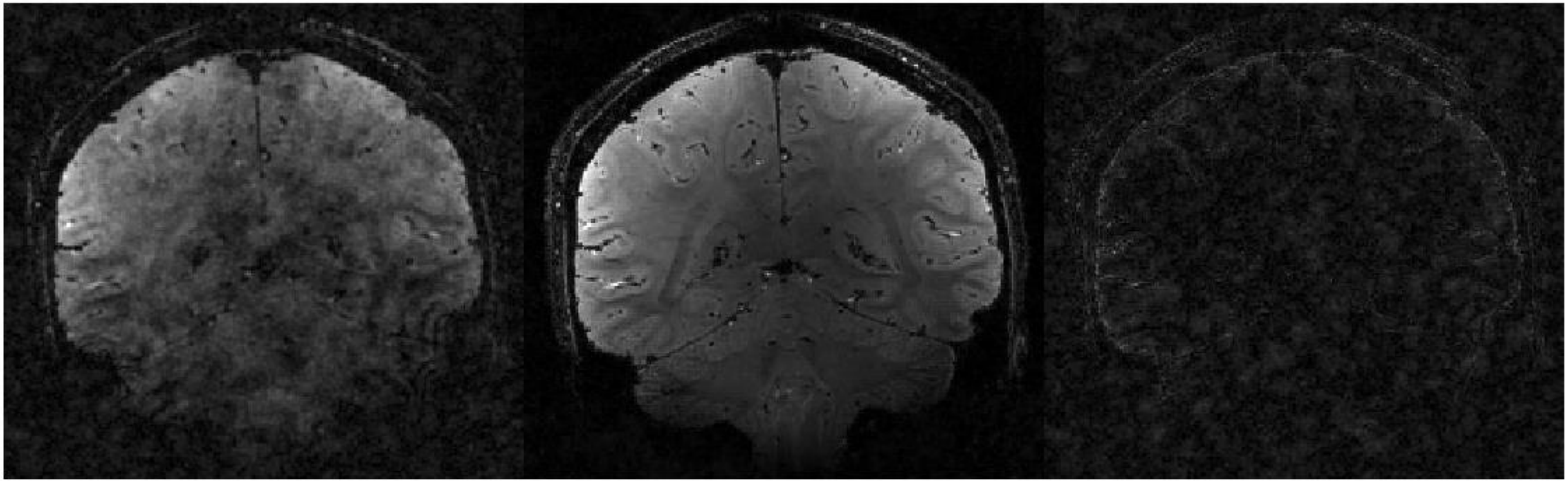


Testing done on full images, sub-sampling masks shown for 6x, 4x and 2x acceleration

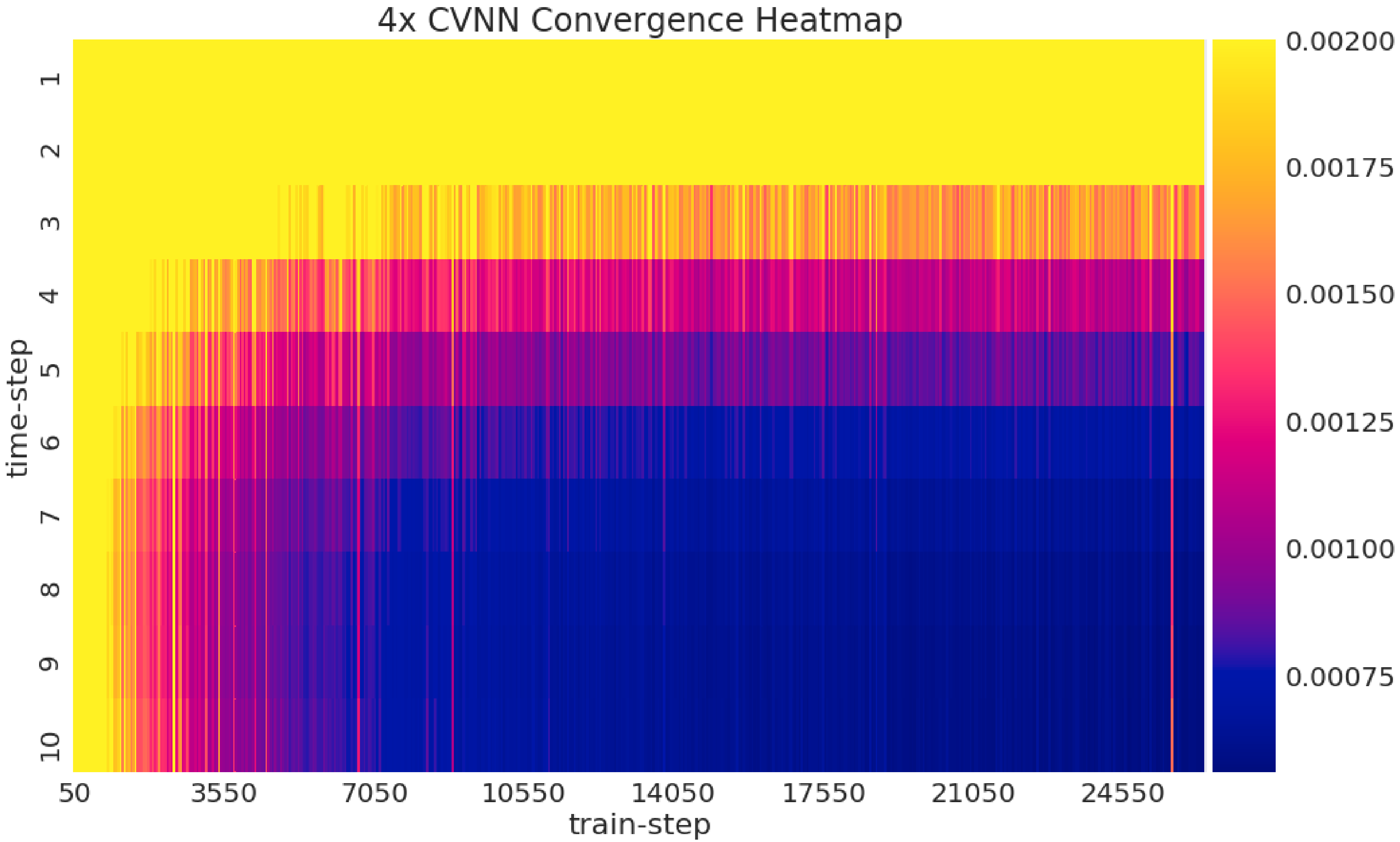


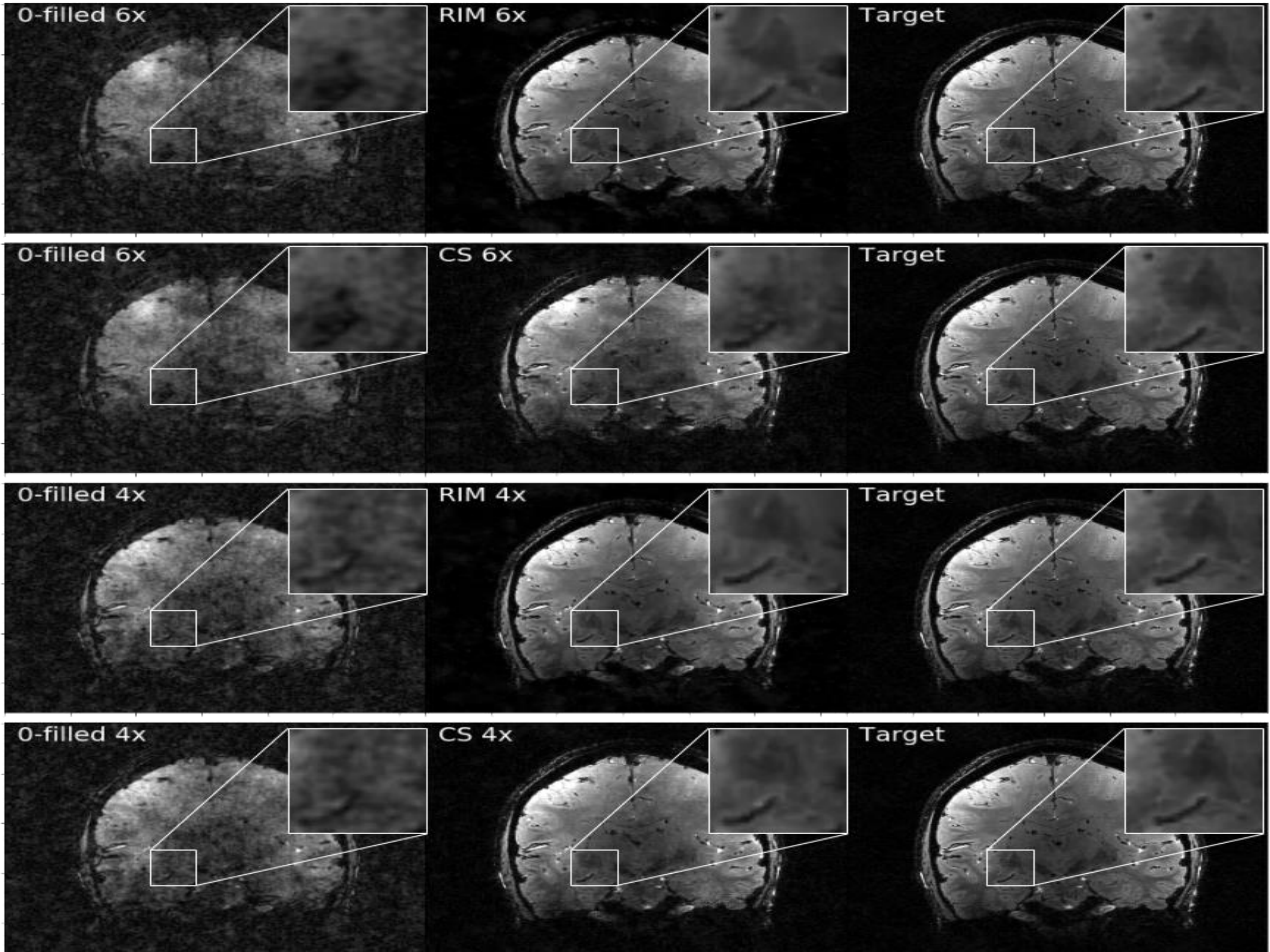
RIM test curves
loss
train-step
0.001
0.002
0.003
0.004
0.005
0
2500
5000
7500
10000
12500
15000
17500
20000
4x RVNN relu 8T 256F
4x CVNN cardioid 8T 128F
6x RVNN relu 8T 256F
6x CVNN cardioid 8T 128F

A full brain RIM reconstruction, starting from the 4 times sub-sampled corruption on the left, attempting to recover the target on the right.

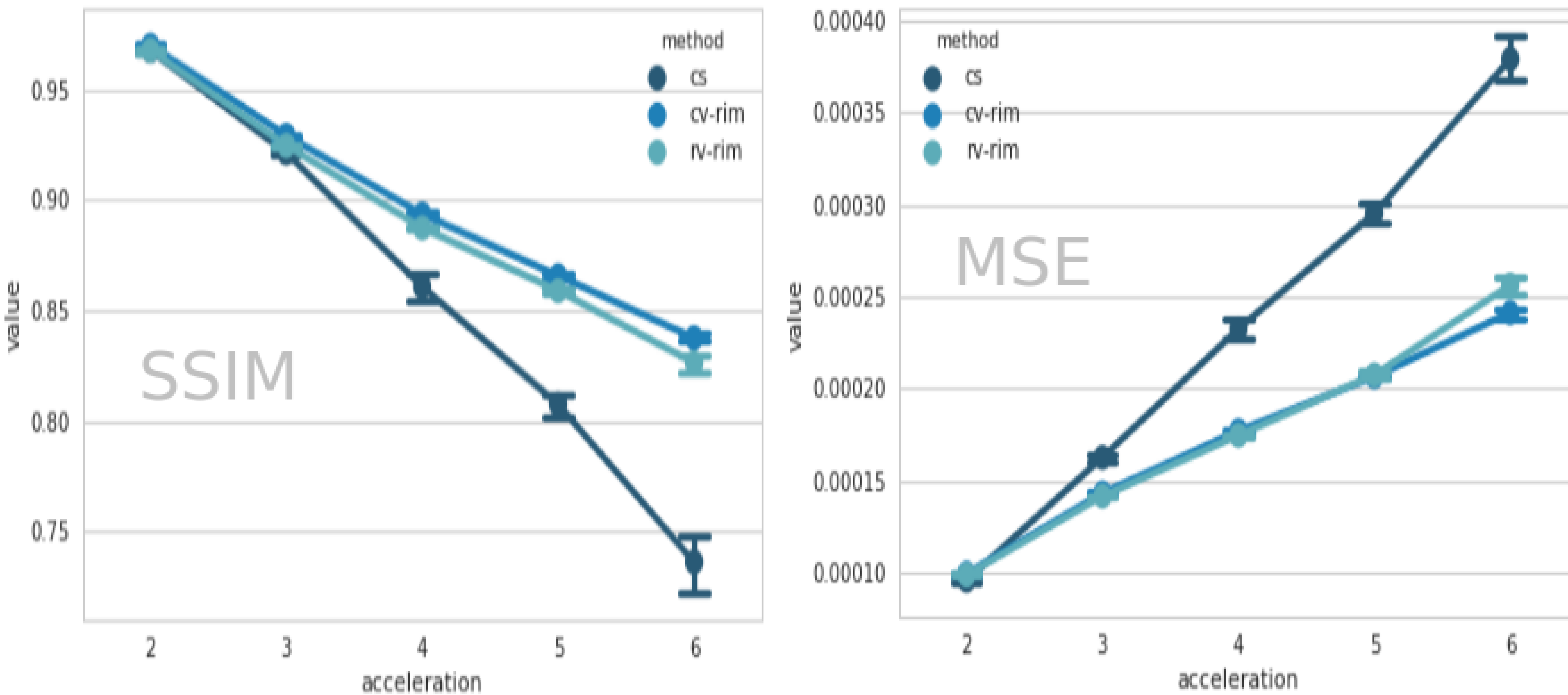


Each time-step in the Recurrent Inference Machine produces a new estimate, here shown to the left, from the 3x accelerated corruption until the 10th and final reconstruction. Target is in the middle, while the error (not to scale) is shown to the right.

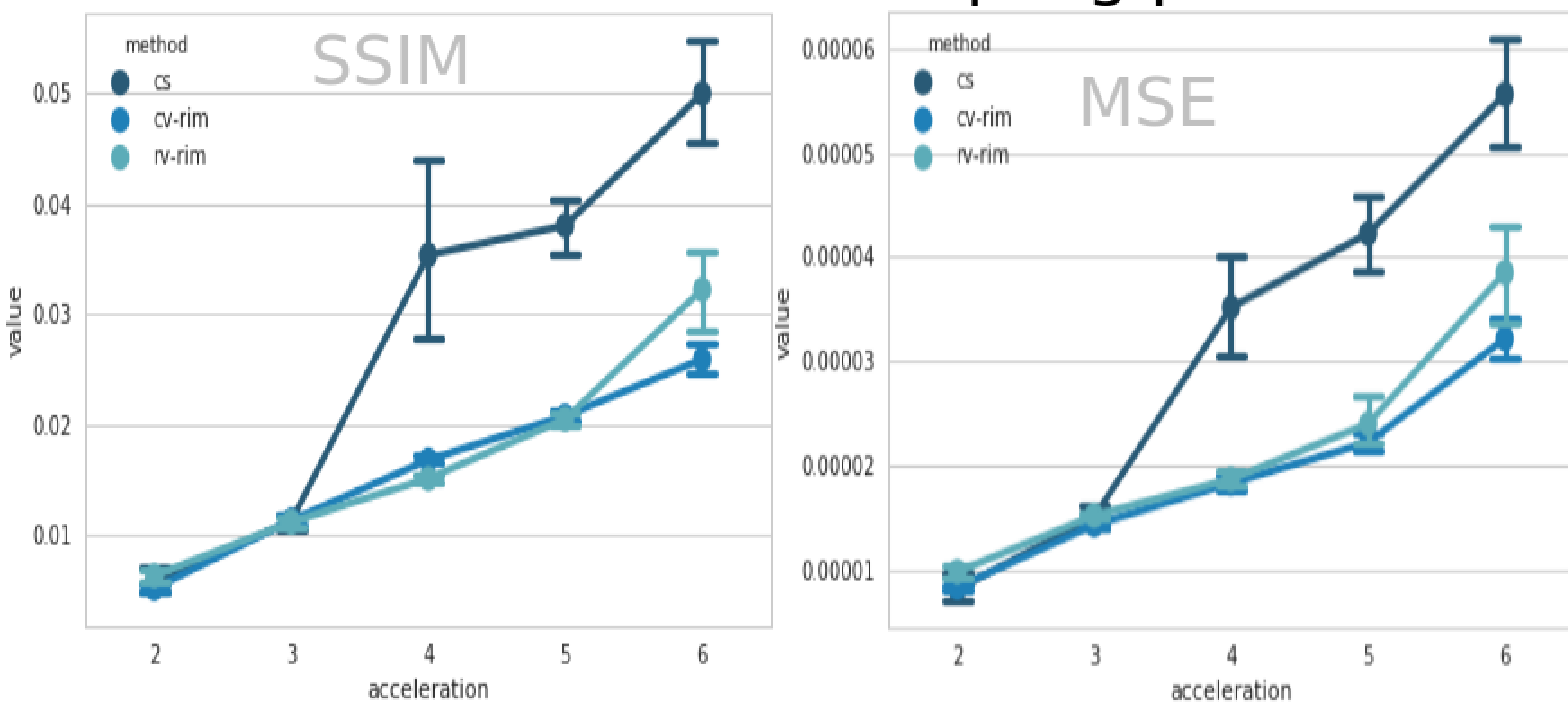

4x CVNN Convergence Heatmap
time-step
1
2
3
4
5
6
7
8
9
10
train-step
50
3550
7050
10550
14050
17550
21050
24550
0.00200
0.00175
0.00150
0.00125
0.00100
0.00075


0-filled 6x
RIM 6x
Target
0-filled 6x
CS 6x
Target
0-filled 4x
RIM 4x
Target
0-filled 4x
CS 4x
Target

Average across sub-sampling patterns



Std across sub-sampling patterns



Conclusions

- Meta learning is interesting new paradigm that can improve classical optimization and inference algorithms by exploiting patterns in classes of problems.

- RIM is a method that unrolls inference and learns to solve inverse problems.

- Great potential to improve & speed up radio-astronomy and MRI image reconstruction.

- Application to MRI-linac?

MRI-Guided Radiation Therapy

The promise of real-time visualization during radiotherapy treatment is pushing science and industry to develop exciting new advances in this cutting-edge technology

Elekta's MR-linac combines two technologies — an MRI scanner and a linear accelerator — in a single system. This allows physicians to precisely locate tumors, tailor the shape of X-ray beams and accurately deliver doses of radiation evento moving tumors.