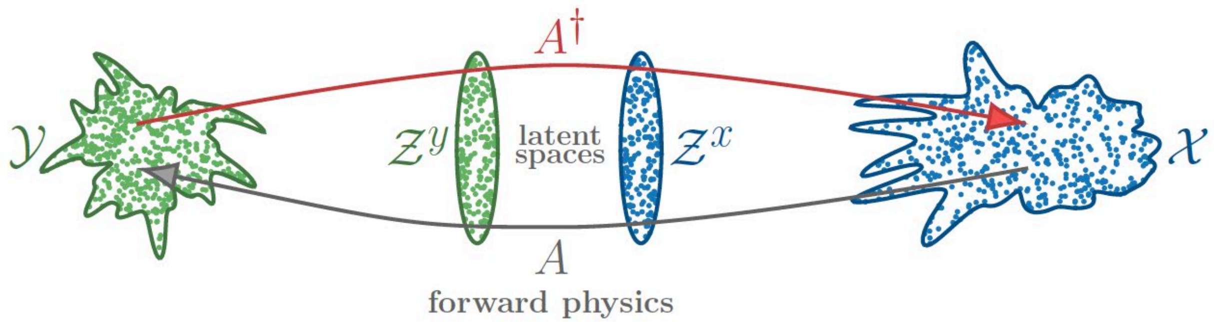
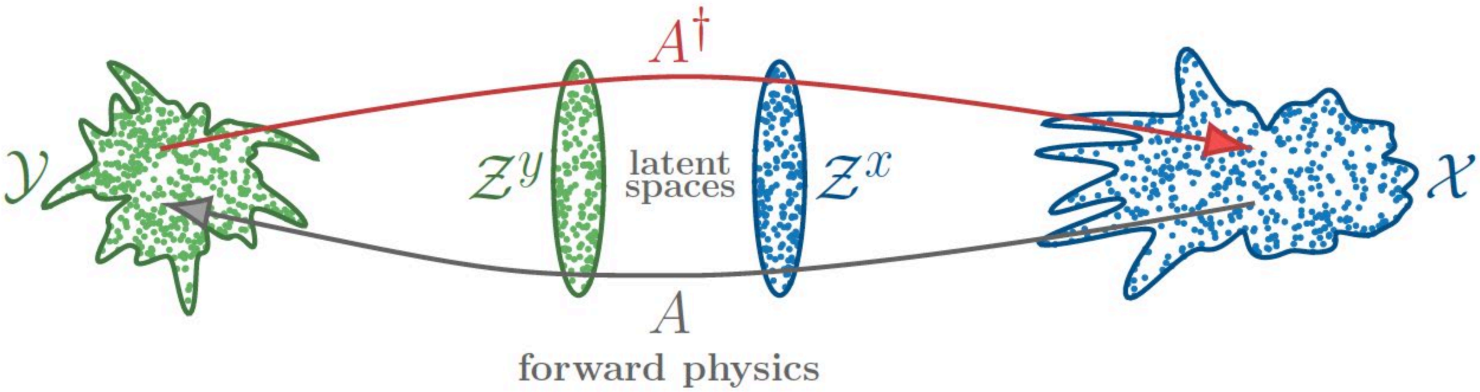
Learned SVD
Deep Learning Decomposition for Inverse Problems

Yoeri Boink, Srirang Manohar, Stephan van Gils, Christoph Brune

Applied Analysis, SACS cluster
Applied Mathematics



WHAT IS A MEANINGFUL MAP?



- Well-posed (existence, uniqueness, stability)?
- Robust inversion / parameter estimation?
- Structure preserving?
- Understandable decomposition?
- Expressive?
- Generalizable?

$$\varphi(x) := \tau(W_L \tau(W_{L-1} \dots \tau(W_1 x) \dots))$$

INVERSE PROBLEMS

- Given measurements of the form
$$y^\delta = Ax + \eta^\delta$$
with x, y in appropriate Banach spaces and A (linear) operator between these spaces.

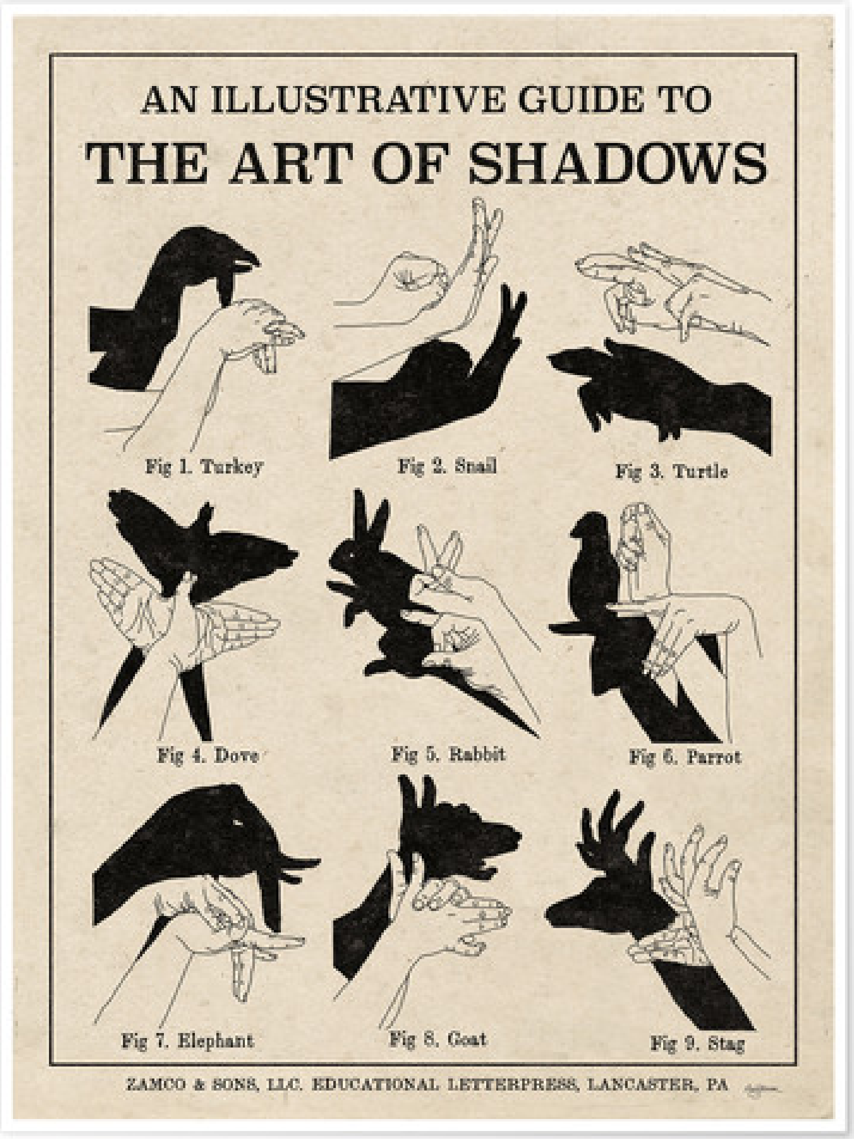
- Noise distribution η^δ known

- If pseudo-inverse $A^\dagger$ available, it is typically **ill-posed (following Hadamard)**

- Example: A Radon transform (CT)

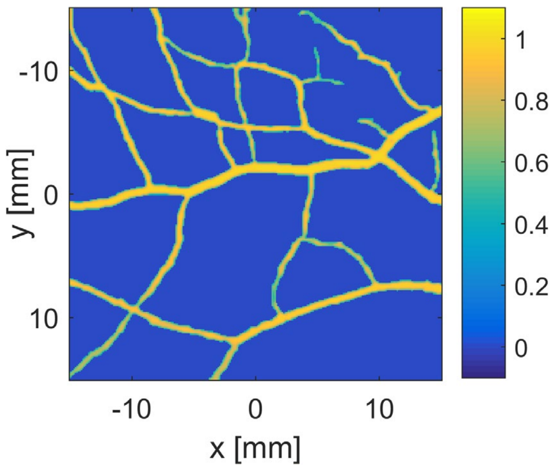
- **Classical solution**: Tikhonov regularization

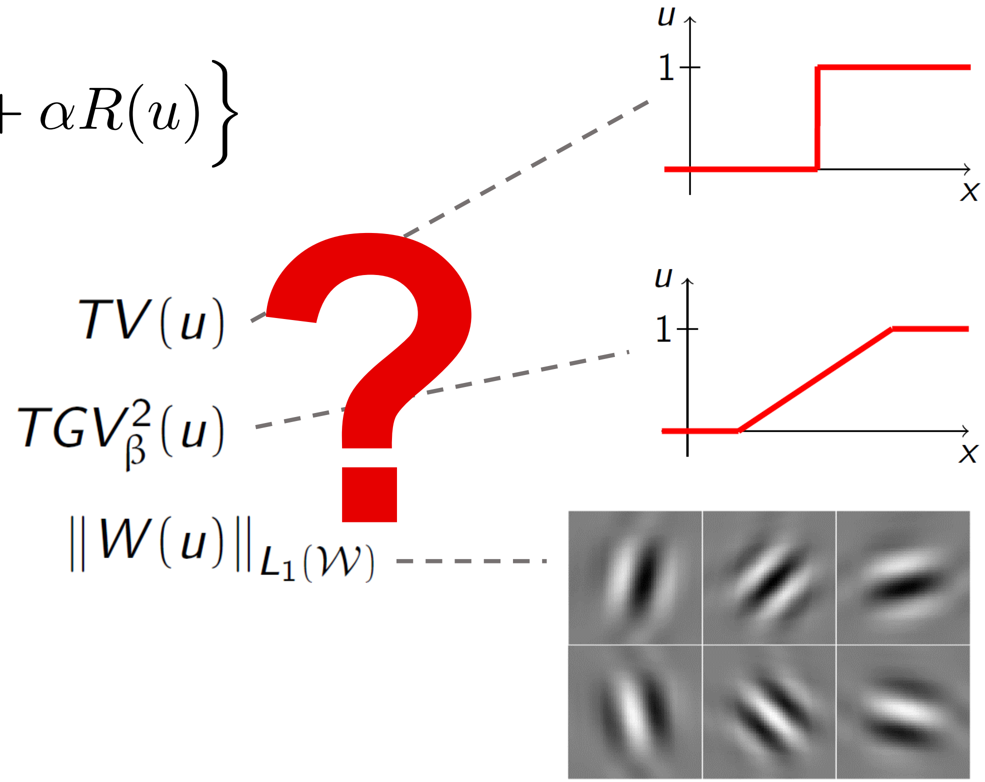
$$x_\alpha = (A^*A + \alpha \mathrm{Id})^{-1} A^* y^\delta = V \underbrace{\left(S^2 + \alpha \mathrm{Id}\right)^{-1} S}_{S_\alpha^{-1}} U^* y^\delta$$



VARIATIONAL REGULARIZATION - CHOOSING FUNCTION SPACES

$$\min_{u \in L^2(\Omega)} \left\{ ||Au - f||^2_{L^2(\Sigma)} + \alpha R(u) \right\}$$





Engl, Hanke, Neubauer (1996)
Natterer, Wübbeling (2001)
Candes, Romberg, Tao (2006)
Kaltenbacher, Neubauer, Scherzer (2008)
Schuster et al (2008)
Osher, Burger, Goldfarb, Xu, Yin (2005)
Rudin, Osher, Fatemi - Nonlinear total variation based noise removal algorithms (1992)
Bredies, Kunisch, Pock - Total Generalised Variation (2010)

VARIATIONAL METHODS AND DEEP LEARNING

$$\min_{u \in L^2(\Omega)} \left\{ ||Au - f||^2_{L^2(\Sigma)} + \alpha R(u) \right\}$$

Partial Differential Equations ⟷ Measurements Hidden Parameters

Classical Inverse Problems
Parameter Estimation

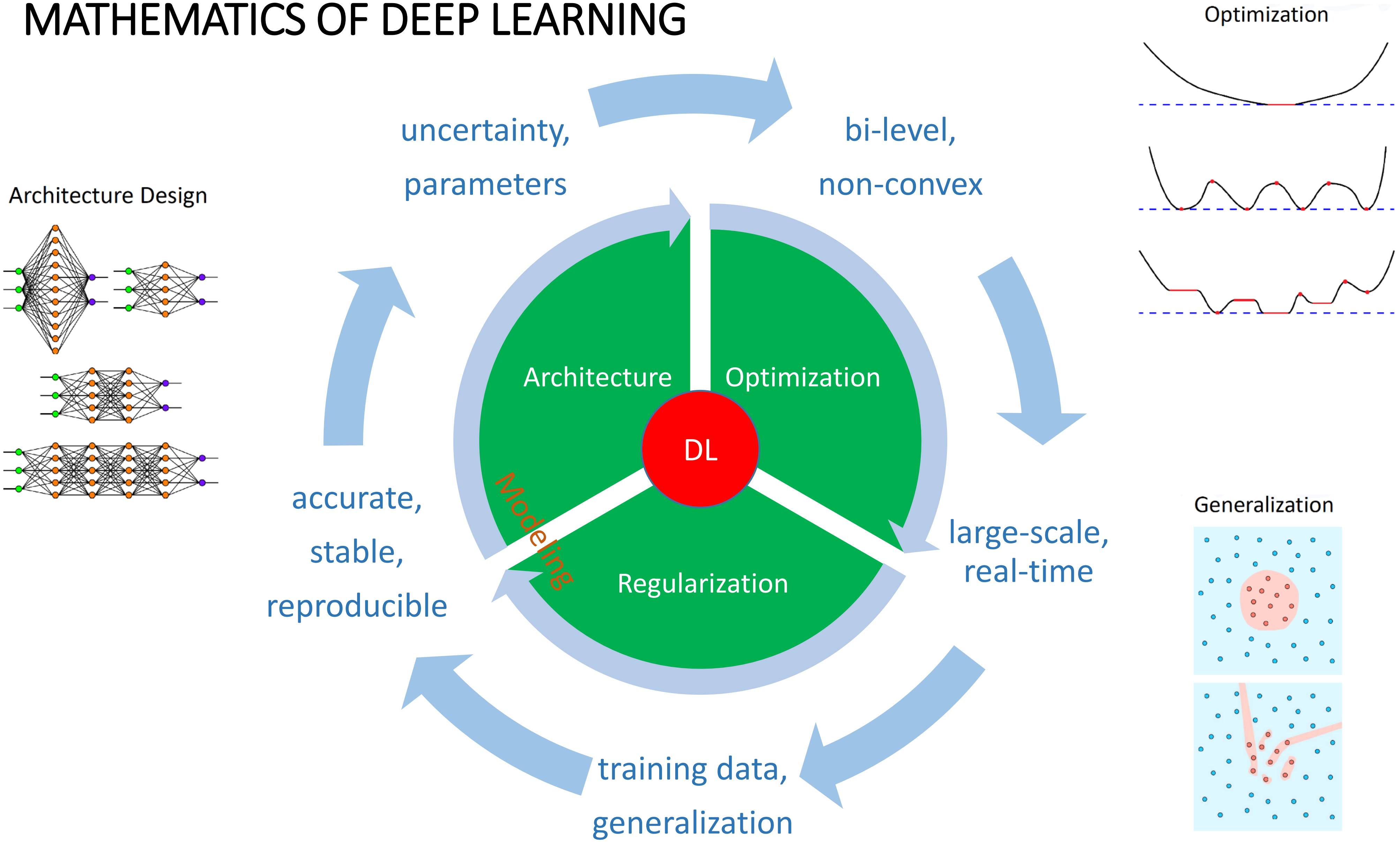
↓

Deep Variational Networks ⟷ Complex Data Science

Generalization?
Regularization Theory?

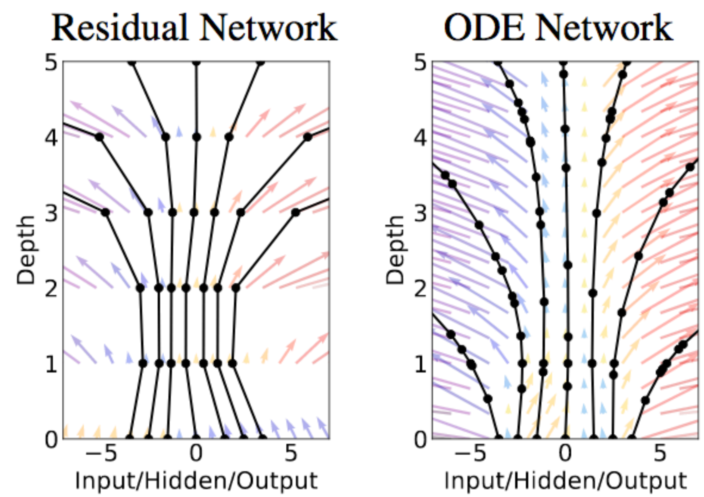
Learned generalized inverse?

MATHEMATICS OF DEEP LEARNING



Vidal et al. Mathematics of Deep Learning, 2018

VARIATIONAL METHODS AND DEEP LEARNING



Deep Residual Neural Networks
are connected to
Partial Differential Equations

Variational methods	**Norms nonconvex**	**Differential operators**	Scale-space, Harmonic analysis	Regularization theory	Inverse problems	**Vector fields, multimodality**	**Time dependent modeling**
Deep networks	Activation functions ReLU, sigmoid	Convolutions functions per layer	Scattering networks	Generalization properties	AEs GANs	Multiple populations?	Residual? Skip connections?

- Chen, Pock - Trainable Nonlinear Reaction Diffusion, 2016
- Mallat - Understanding deep convolutional networks, 2016
- Chen et al - Neural Ordinary Differential Equations, 2018
- Ciccone et al - Stable Deep Networks from Non-Autonomous DEs, 2018
- Haber, Ruthotto - Stable architectures for deep neural networks, 2017

DEEP LEARNING FOR INVERSE PROBLEMS

- Fully Learned Models
  $$x = \varphi_\theta(y)$$
- Post Processing
  $$x = \varphi_\theta(A^\dagger(y))$$
- Iterative Schemes
  $$x_0 = A^\dagger(y)$$
  $$x_{n+1} = \varphi_\theta(x_n, \nabla_{x_n} \|A(x_n) - y\|)$$
- Learning a Regularizer
- Review

Zhu et al – Image reconstruction by domain-transform manifold learning (Nature, 2018)

Jin, McCann, Froustey, Unser – Deep Convolutional Neural Network for Inverse Problems in Imaging (2018)

Putzky, Welling - Recurrent Inference Machines for Solving Inverse Problems (2017)

Meinhardt, Möller, Hazirbas, Cremers - Learning Proximal Operators (2017)

Banert, Ringh, Adler, Karlsson, Öktem – Data-driven nonsmooth optimization (2018)

Li, Schwab, Antholzer, Haltmeier - NETT Solving Inverse Problems with Deep Neural Networks (2018)

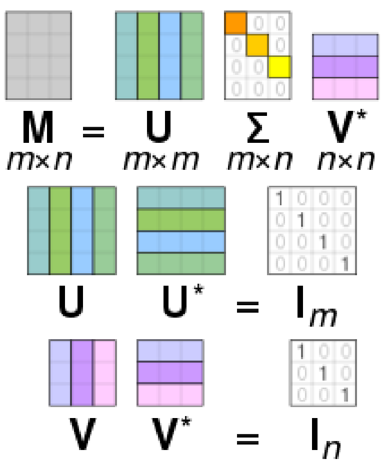
Lunz, Öktem, Schönlieb – Adversarial Regularizers in Inverse Problems (2018)

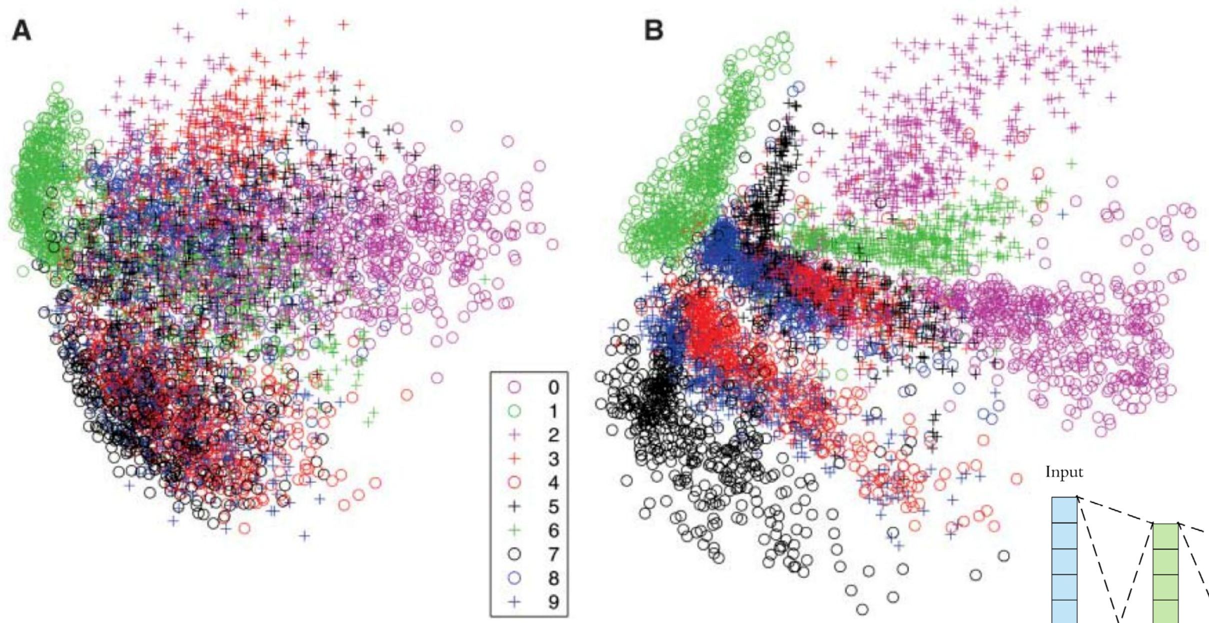
Arridge et al – Solving inverse problems using data-driven models (Acta Numerica, 2019)

DATA DIMENSIONALITY REDUCTION

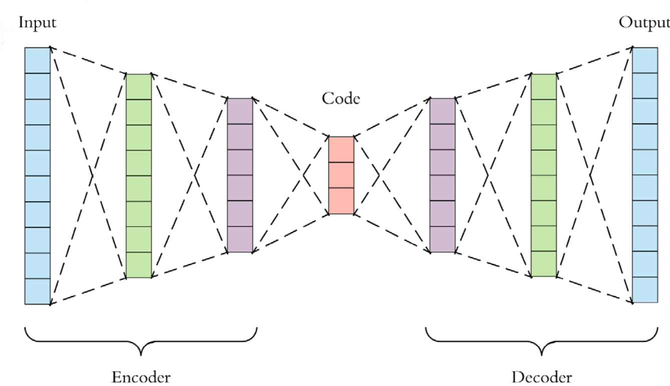
STRUCTURE, SCALE AND INTERPOLATION OF DATA

SVD
linear





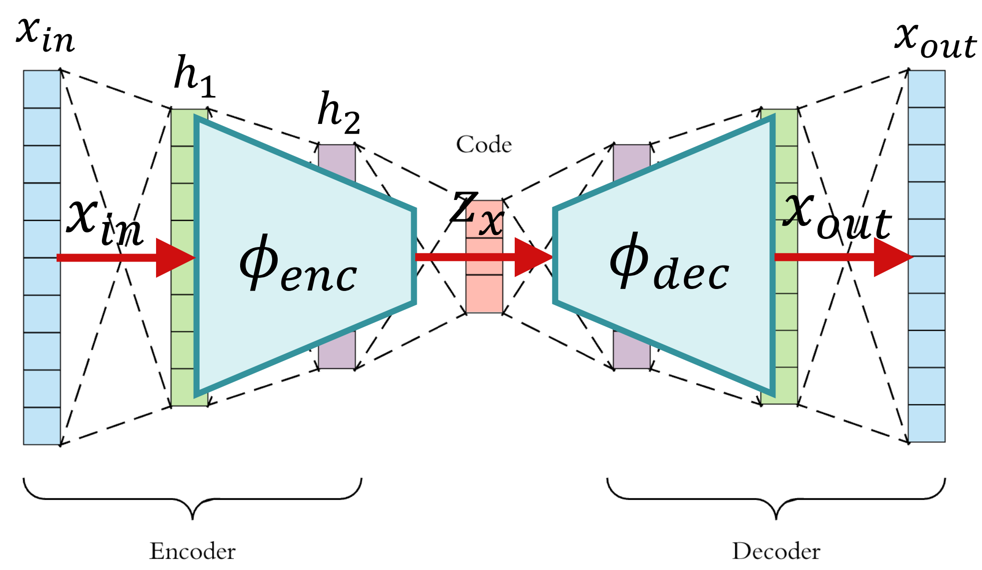
Autoencoder
nonlinear



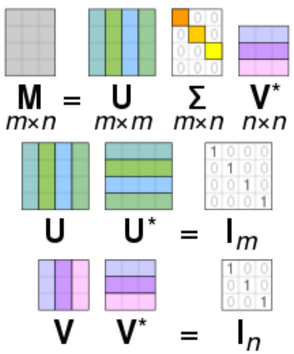
Hinton - Reducing the Dimensionality of Data with Neural Networks (Science, 2006)

AUTOENCODERS (AE)

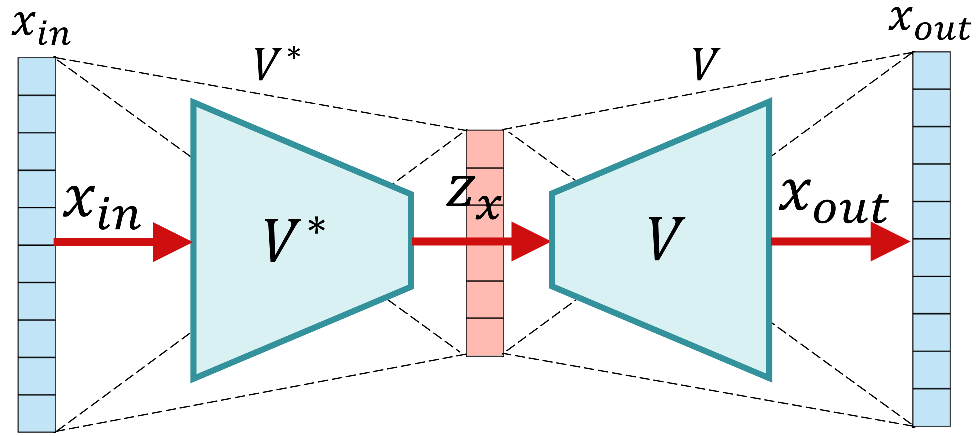
$$\left.\begin{aligned} h_1 &= \sigma(W_1 x_{in} + b_1) \\ h_2 &= \sigma(W_2 h_1 + b_2) \\ &\vdots \end{aligned}\right\} \quad x_{out} = \phi_{dec}(\phi_{enc}(x_{in}))$$



SVD: 2 connected AE



- 1 layer
- no nonlinearities σ
- no bias b



AUTOENCODERS AND GRADIENT FLOWS

- Assume we have a trained AE with tied weights, i.e. $f(x) = W^\intercal \phi(Wx)$
- *x* and *f(x)* live in the same vector space, i.e.

$$G(x) = f(x) - x$$

  is a vector field pointing from *x* to reconstructed *f(x)*
- Under assumptions, *G(x)* is the gradient field of an energy *E(x)*

$$G(x) = f(x) - x \sim -\partial_x E$$

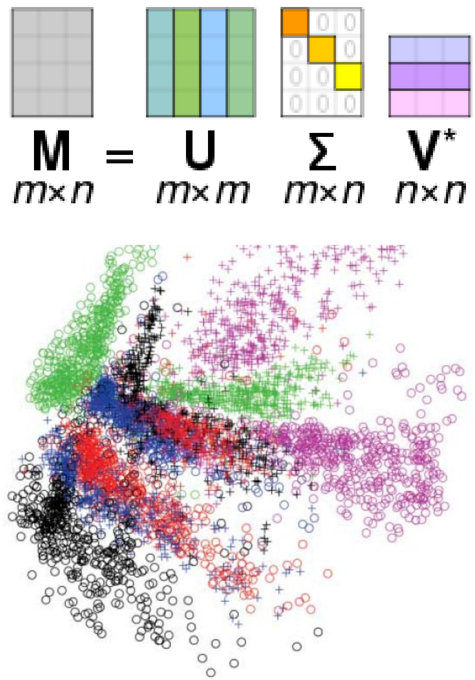
$$\Rightarrow \exists \text{ energy } E(x;W) \text{ with } f(x) = W^\intercal \phi(Wx) = x - \partial_x E(x;W)$$

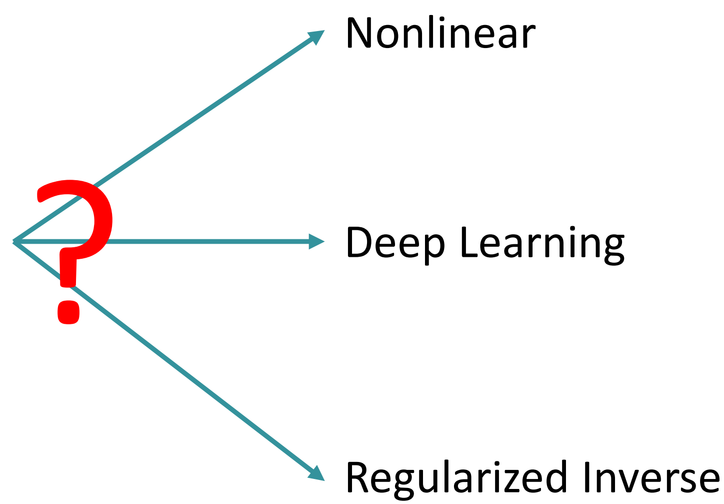
$$E(x) = \sum_{i=1}^{N^2} \Phi((Wx)_i) + \frac{1}{2}\|x\|_2^2 \quad \text{with} \quad \phi(x) = -\Phi'(x)$$

NEW FRAMEWORK

LEARNED SVD

DEEP LEARNING DECOMPOSITION FOR INVERSE PROBLEMS



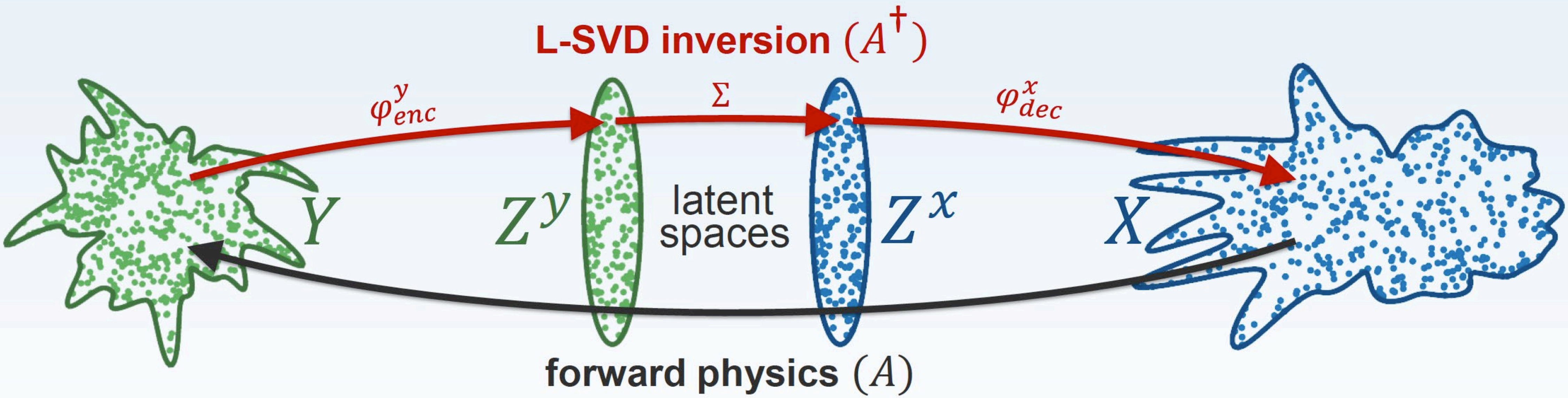


DEEP LEARNING MODEL

STRUCTURE, INTERPOLATION AND SCALE?

LEARNED SINGULAR VALUE DECOMPOSITION (L-SVD)

- Truncated SVD for linear inverse problems $y^\delta = Ax + \eta^\delta, \quad A = USV^* \;\Rightarrow\; \tilde{x} = V_K S_K^{-1} U_K^* y^\delta$
- Propose nonlinear, learned SVD $\hat{x}^\Sigma = \varphi_{dec}^x\left(\Sigma\, \varphi_{enc}^y(y^\delta)\right)$



- From model-driven to data-driven
- From linear to nonlinear encoding
- Regularization by hybrid autoencoding

LEARNED SINGULAR VALUE DECOMPOSITION (L-SVD)

We define the nonlinear functions

$$\varphi^y_{enc} : Y \mapsto Z^y, \quad \varphi^y_{dec} : Z^y \mapsto Y, \quad \varphi^x_{enc} : X \mapsto Z^x, \quad \varphi^x_{dec} : Z^x \mapsto X$$

and we define the square matrix $\Sigma \in \mathbb{R}^{k \times k}$. *Moreover we define the variables*

$$\begin{aligned} z_y &:= \varphi^y_{enc}(y^\delta), & z^{AE}_x &:= \varphi^x_{enc}(x), & z^\Sigma_x &:= \Sigma z_y, \\ \hat{y}^{AE} &:= \varphi^y_{dec}(z_y), & \hat{x}^{AE} &:= \varphi^x_{dec}(z^{AE}_x), & \hat{x}^\Sigma &= \varphi^x_{dec}(z^\Sigma_x). \end{aligned}$$

$$\min_{\text{par}_{\text{NN}}} \left\{ \sum_{i=1}^{\#\text{train}} \underbrace{D_1\left(\hat{x}^\Sigma_{(i)}, x_{(i)}\right)}_{\text{reconstruction}} + \alpha_y \underbrace{D_2\left(\hat{y}^{\text{AE}}_{(i)}, y_{(i)}\right)}_{\text{autoencoder}} + \alpha_x \underbrace{D_3\left(\hat{x}^{\text{AE}}_{(i)}, x_{(i)}\right)}_{\text{autoencoder}} \right\}$$

LEARNED SINGULAR VALUE DECOMPOSITION (L-SVD)

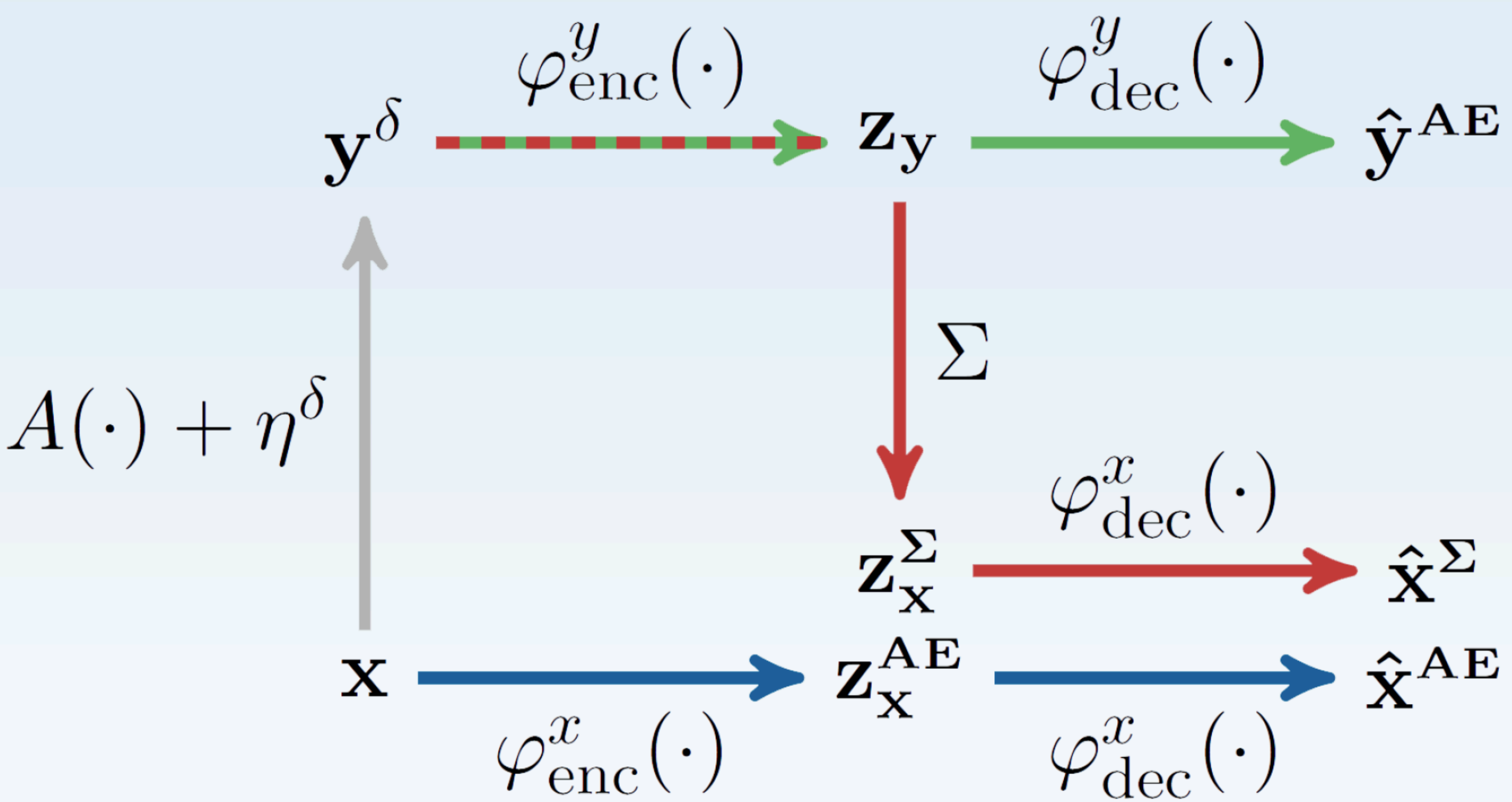
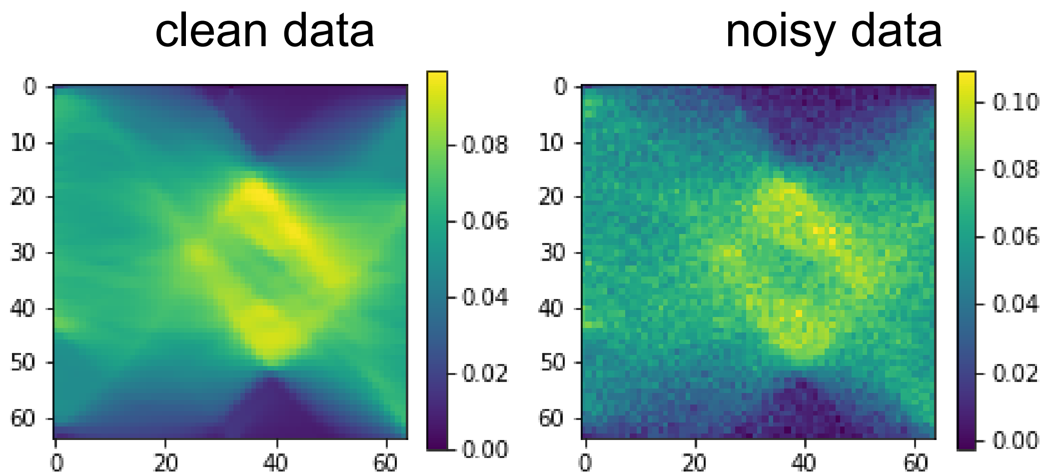
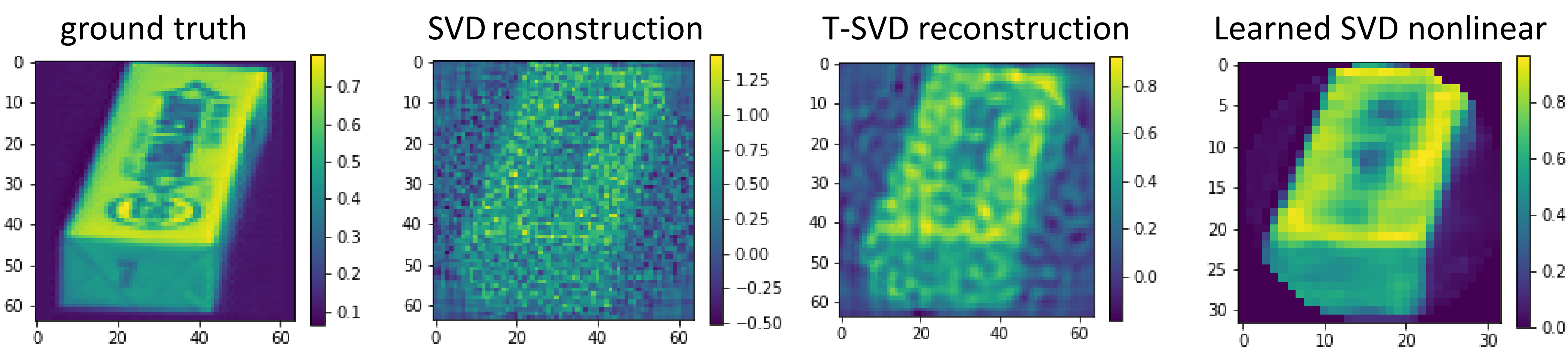


***Figure 1:** in **green** the autoencoder for data y^δ; in **blue** the autoencoder for signal x; in **red** the reconstruction procedure.*

SOLVING THE INVERSE PROBLEM



$$\hat{x}^{\Sigma} = \varphi_{dec}^{x}\left(\Sigma\, \varphi_{enc}^{y}(y^{\delta})\right)$$



SIMULATION EXPERIMENTS

- Simulation experiments for CT-reconstruction (A is Radon transform).
- Trained on rescaled MNIST dataset (64x64 pixels).
- Two scenarios were tested:
  1) linear L-SVD, 'full-angle' (64 angles), no bottleneck latent space ($Y = Z^y = Z^x = X = \mathbb{R}^{4096}$);
  2) nonlinear L-SVD, limited-angle (8 angles), bottleneck latent space ($Y = \mathbb{R}^{256}, \quad Z^y = Z^x = \mathbb{R}^{64}, \quad X = \mathbb{R}^{4096}$).
- Gaussian noise on data with standard deviation $\delta = 0.05$.

- Fully-connected networks, 4 layers in $AE_{X,}$ 1 layer in AE_Y .
- No biases.
- Leaky ReLU with parameter 0.1 in autoencoders.
- Adam optimiser with learning rate 0.001.

MODEL-DRIVEN TO DATA-DRIVEN

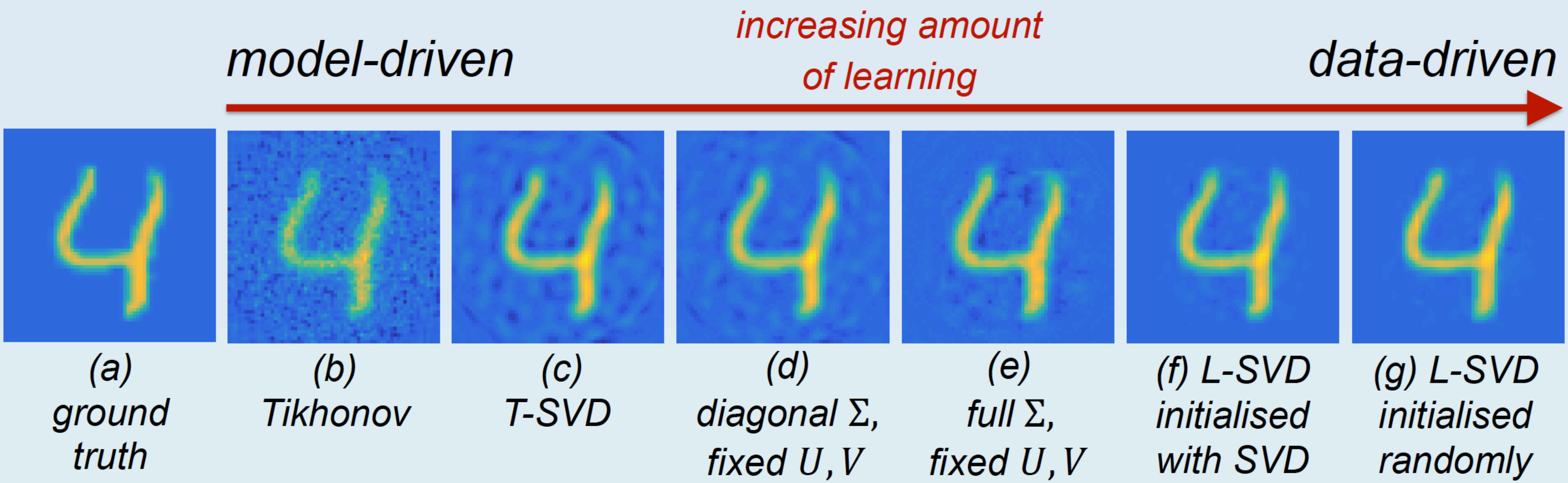


Figure 2: *linear reconstruction methods with increasing amount of learning. Linear L-SVD provides better results than methods that are less data-driven.*

HYBRID DICTIONARY LEARNING

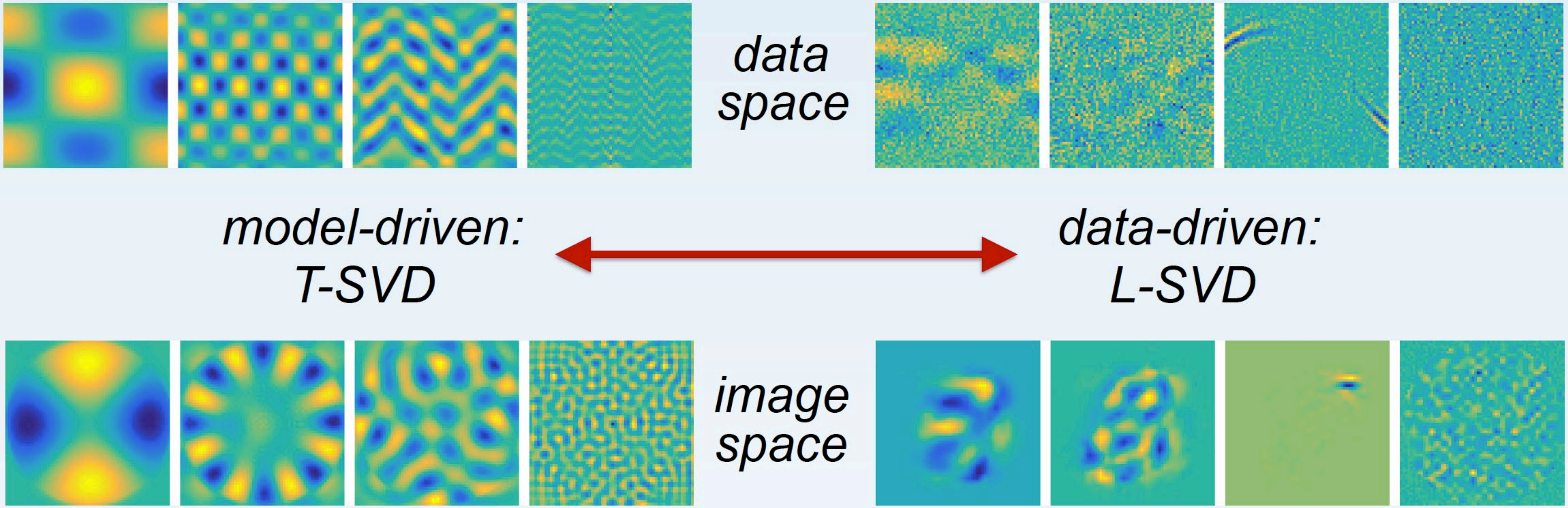
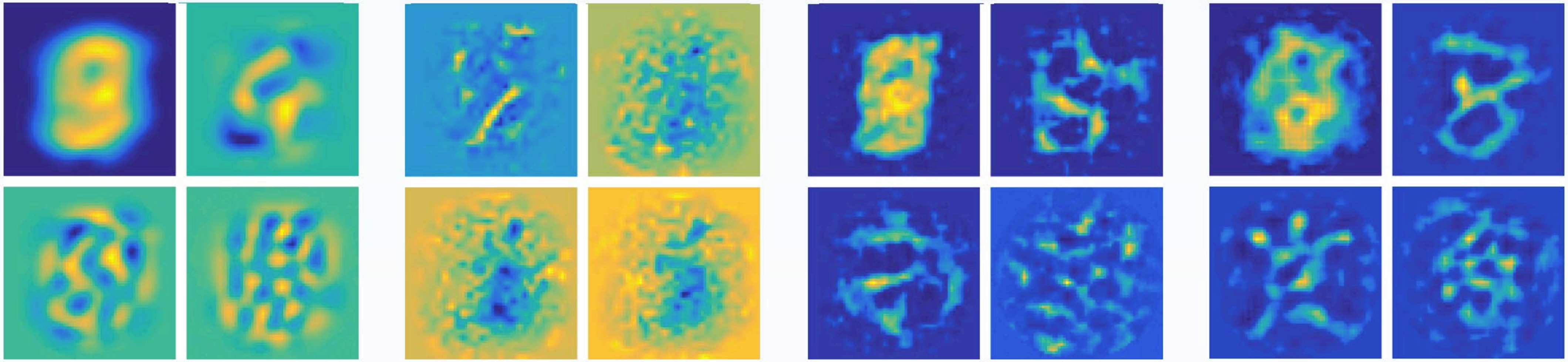


***Figure 3:** selected 'dictionary elements' from T-SVD and linear L-SVD. L-SVD shows similar structures (harmonics) as T-SVD, but with a stronger focus on the specific data.*

HYBRID LATENT SPACE REPRESENTATION

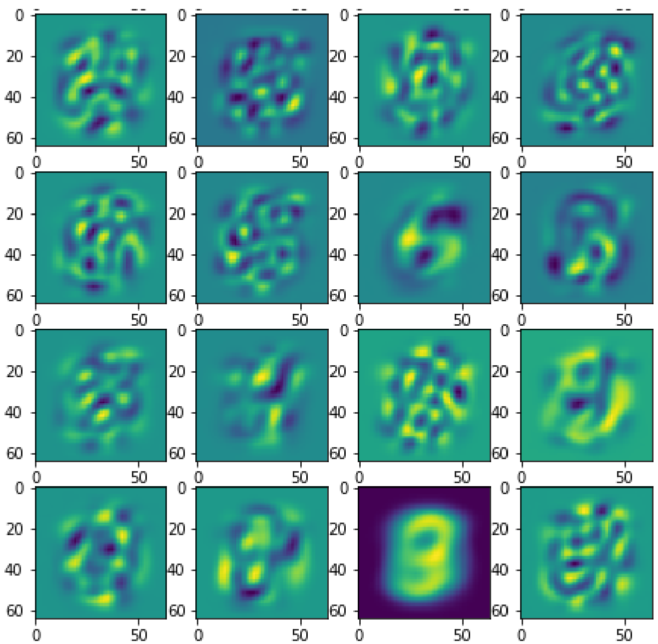


(a) linear AE/L-SVD (b) nonlinear AE (c) L-SVD without AE branches (d) L-SVD

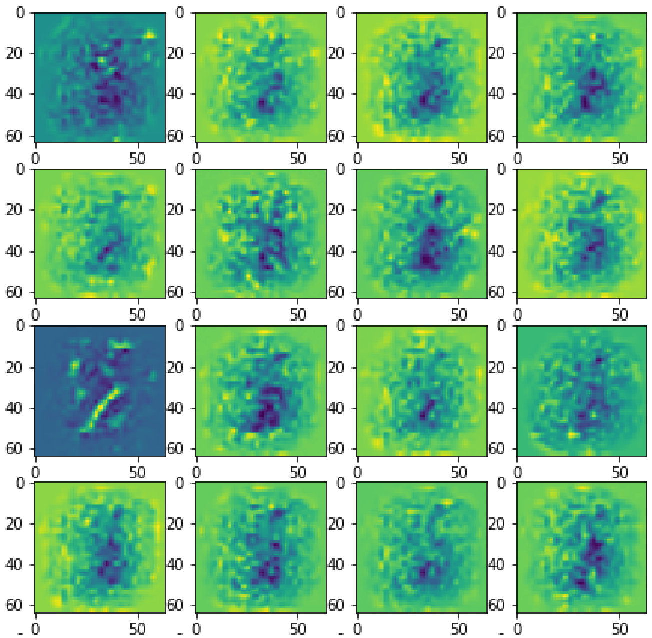
- Selected unit vectors from latent space decoded to image space
- Nonlinear L-SVD learns more interpretable representation
- Combines features from data space, image space and operator
- Autoencoder provides ‘smoother’ and better connected structures

STRUCTURE: LATENT REPRESENTATIONS

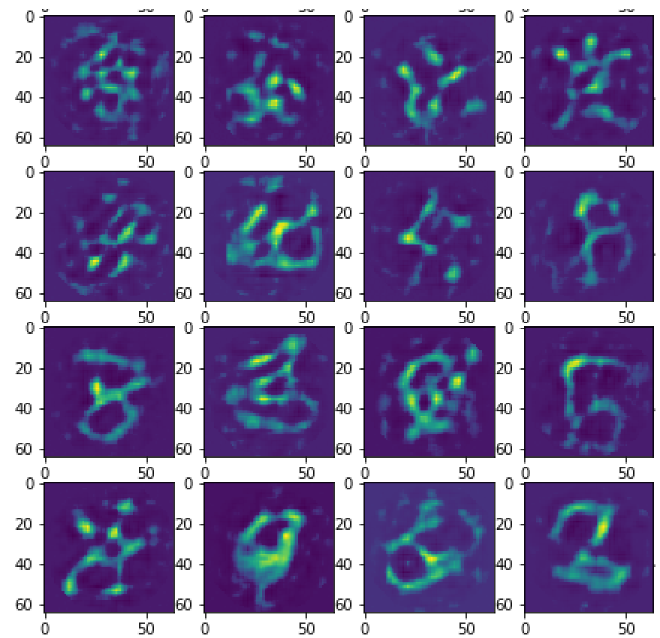
linear AE/L-SVD



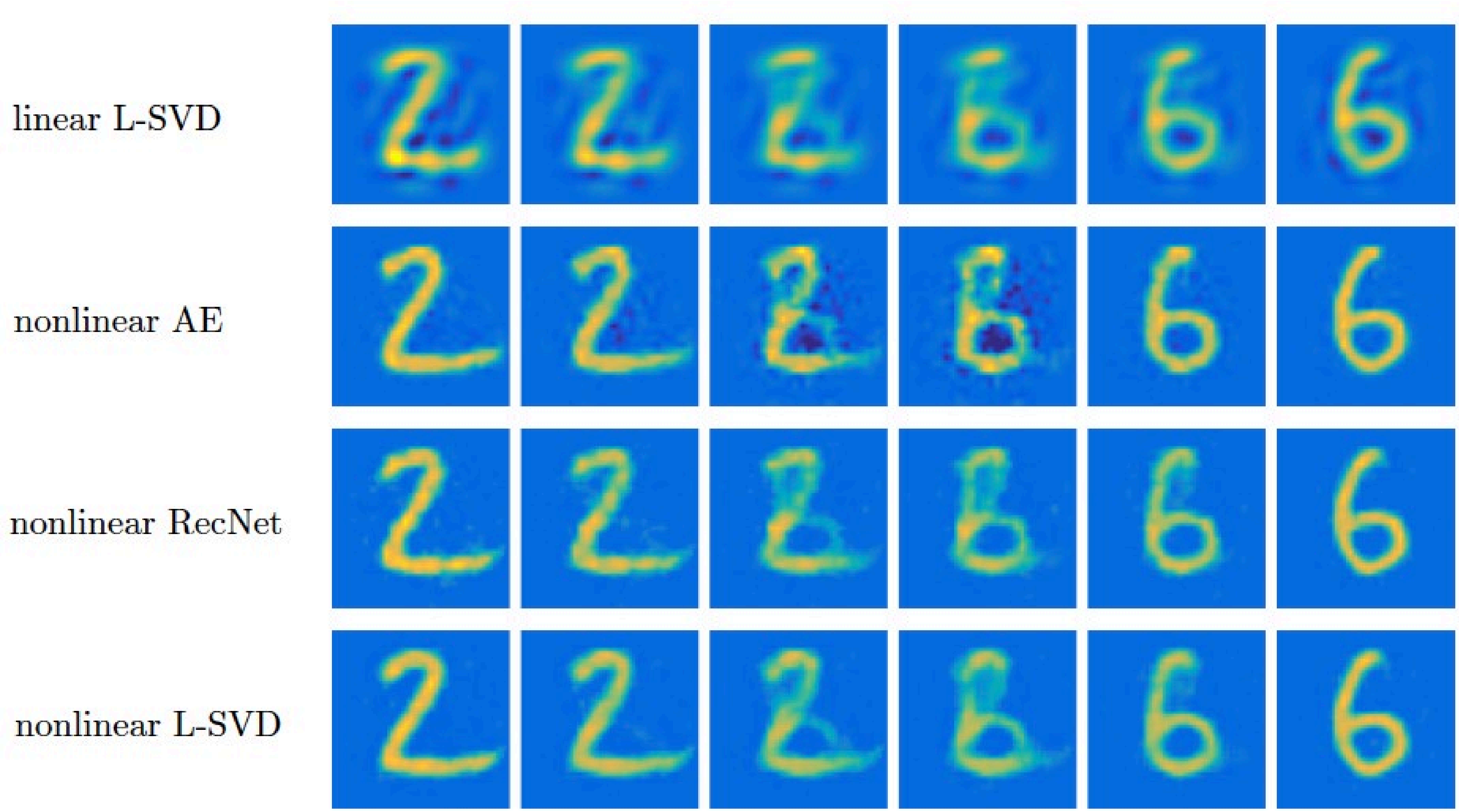
nonlinear AE



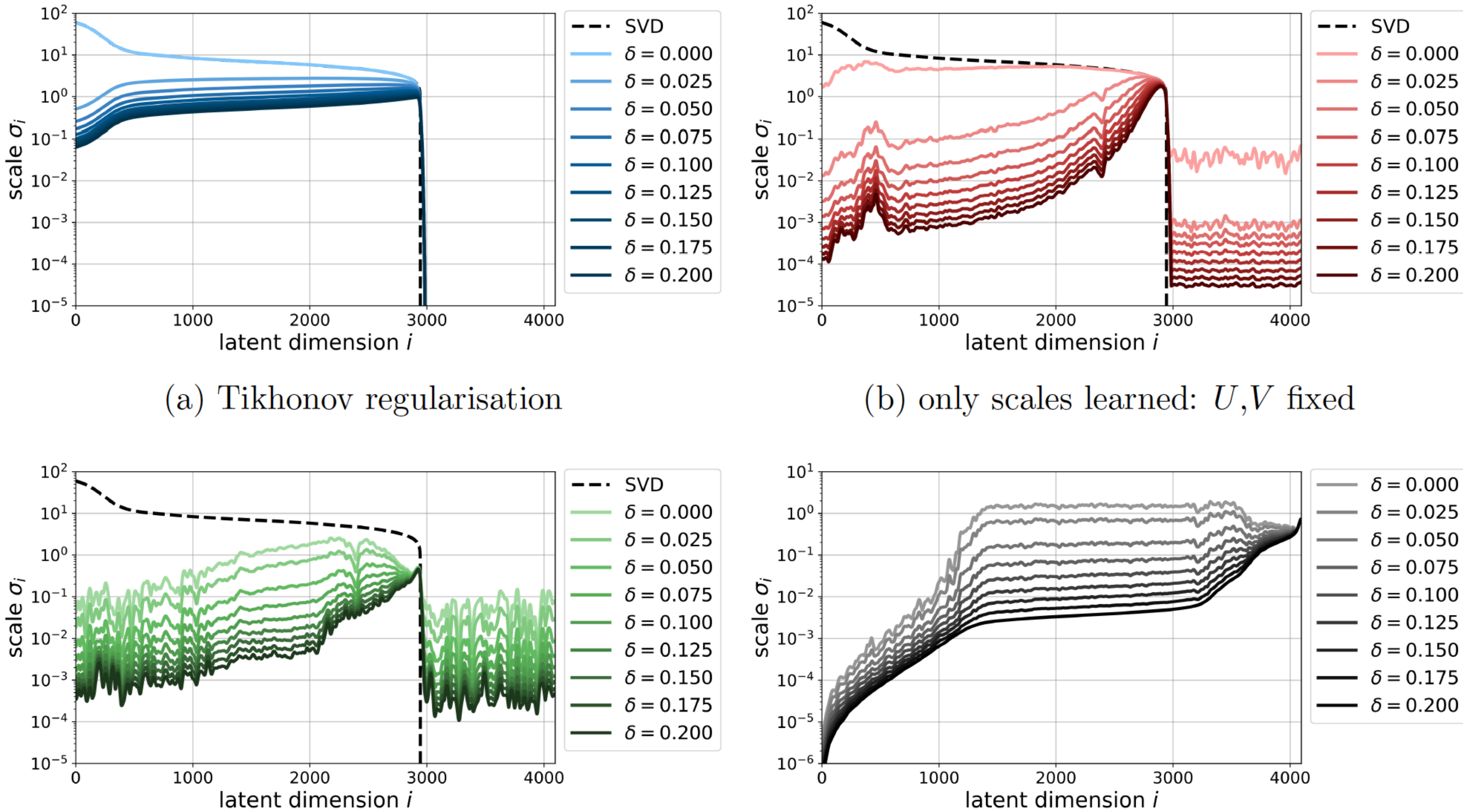
L-SVD



INTERPOLATION



LEARNED REGULARIZATION (SCALES FROM NOISE)



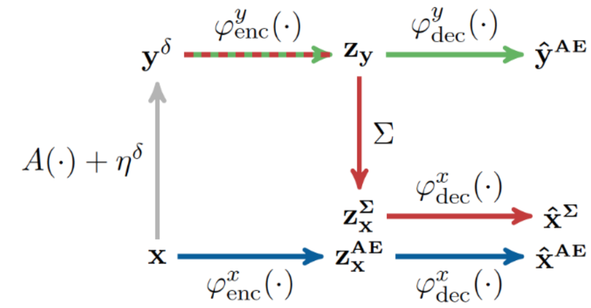
(a) Tikhonov regularisation

(b) only scales learned: U,V fixed

(c) L-SVD with SVD initialisation

(d) L-SVD with random initialisation

STABILITY ESTIMATE



$$
\begin{aligned}
\|\hat{x}^\Sigma_{(1)} - \hat{x}^\Sigma_{(2)}\|_{\ell^2} &\leq \|\varphi^x_{\text{dec}}\|_{\text{op}} \; \|\hat{z}_{x,(1)} - \hat{z}_{x,(2)}\|_{\ell^2} \\
&\leq |\sigma_{\max}| \; \|\varphi^x_{\text{dec}}\|_{\text{op}} \; \|\hat{z}_{y,(1)} - \hat{z}_{y,(2)}\|_{\ell^2} \\
&\leq \|\varphi^y_{\text{enc}}\|_{\text{op}} \; |\sigma_{\max}| \; \|\varphi^x_{\text{dec}}\|_{\text{op}} \; \|y_{(1)} - y_{(2)}\|_{\ell^2}
\end{aligned}
$$

RECONSTRUCTION ERROR ESTIMATE

Theorem 3.3. *Consider the L-SVD network as defined in section 3.1. Assume that for some $0 < \varepsilon_z < 1$ we have that for all $\tilde{z}_x \in B_1$, $\|\tilde{z}_x - z^\Sigma_x\|_{\ell^2} < \varepsilon_z$. Then there exists a $C > 0$ s.t. for all $\tilde{z}_x \in B_1$, $\|\varphi^x_{dec}(\tilde{z}_x) - \varphi^x_{dec}(\Sigma\varphi^y_{enc}(A\varphi^x_{dec}(\tilde{z}_x)))\|_{\ell^2} < C$. Moreover, for all $x \in \mathbb{R}^n$ for which $\varphi^x_{enc}(x) \in B_{1-\varepsilon_z}$, we have the error estimate $\|x - \varphi^x_{dec}(\Sigma\varphi^y_{enc}(Ax))\|_{\ell^2} < C$.*

Y.E. Boink, C. Brune - Learned SVD: Solving Inverse Problems via Hybrid Autoencoding (arXiv preprint, Nov 2019)

RESEARCH CONTEXT

MANIFOLD LEARNING, INVERSE PROBLEMS

- Zhu et al – Image reconstruction by domain-transform manifold learning (Nature, 2018)
- Lunz et al – Adversarial Regularizers in Inverse Problems (Neurips, 2018)
- Angles, Mallat – Generative networks as inverse problems with scattering transforms (ICLR, 2018)
- Wong et al – Training Auto-encoder-based Optimizers for Terahertz Image Reconstruction (arXiv, 2019)

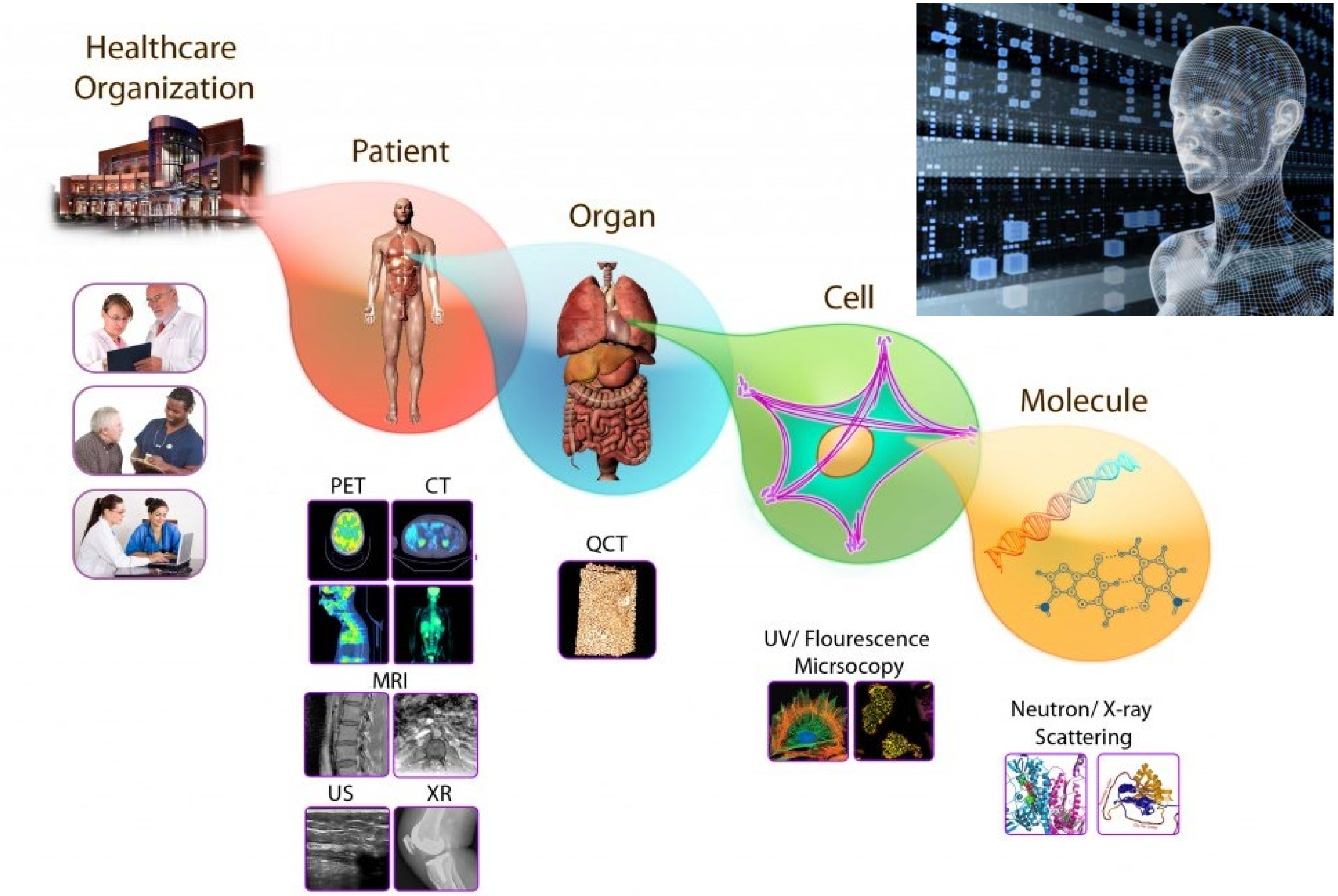
SEMI-SUPERVISED AUTOENCODERS

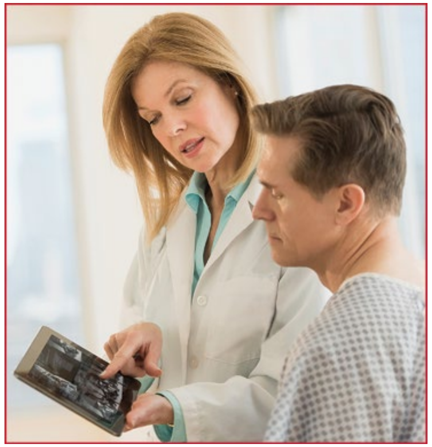
- Ye et al – Understanding Geometry of Encoder-Decoder CNNs (ICML, 2019)
- Le et al – Supervised autoencoders: Improving generalization performance with unsupervised regularizers (NeurIPS, 2018)
- Epstein et al – Generalization bounds for unsupervised and semi-supervised learning with AutoEncoders (arXiv, 2019)

MODEL REDUCTION AND LEARNING

BAYESIAN INVERSION AND SPARSITY

Deep Learning for Inverse Problems in Medical Imaging





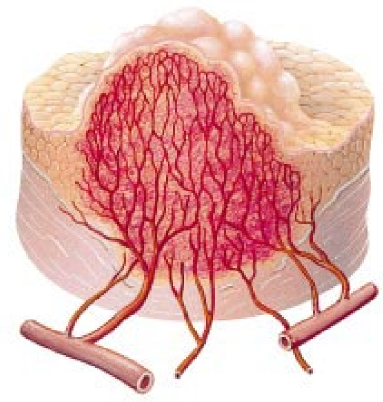
Multimodality, multiscale, multidimension, big data, mobile…

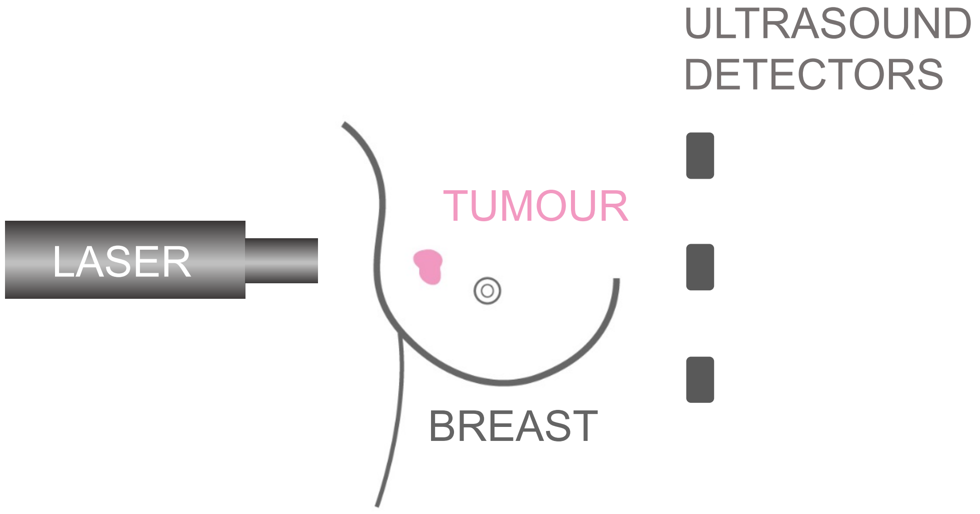
Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique, IEEE Trans Med Imag, 2016

PHOTOACOUSTIC BREAST IMAGING

(ILLUSTRATIONS MADE BY SJOUKJE SCHOUSTRA – BMPI GROUP, UNI TWENTE)

Folkman, 1996

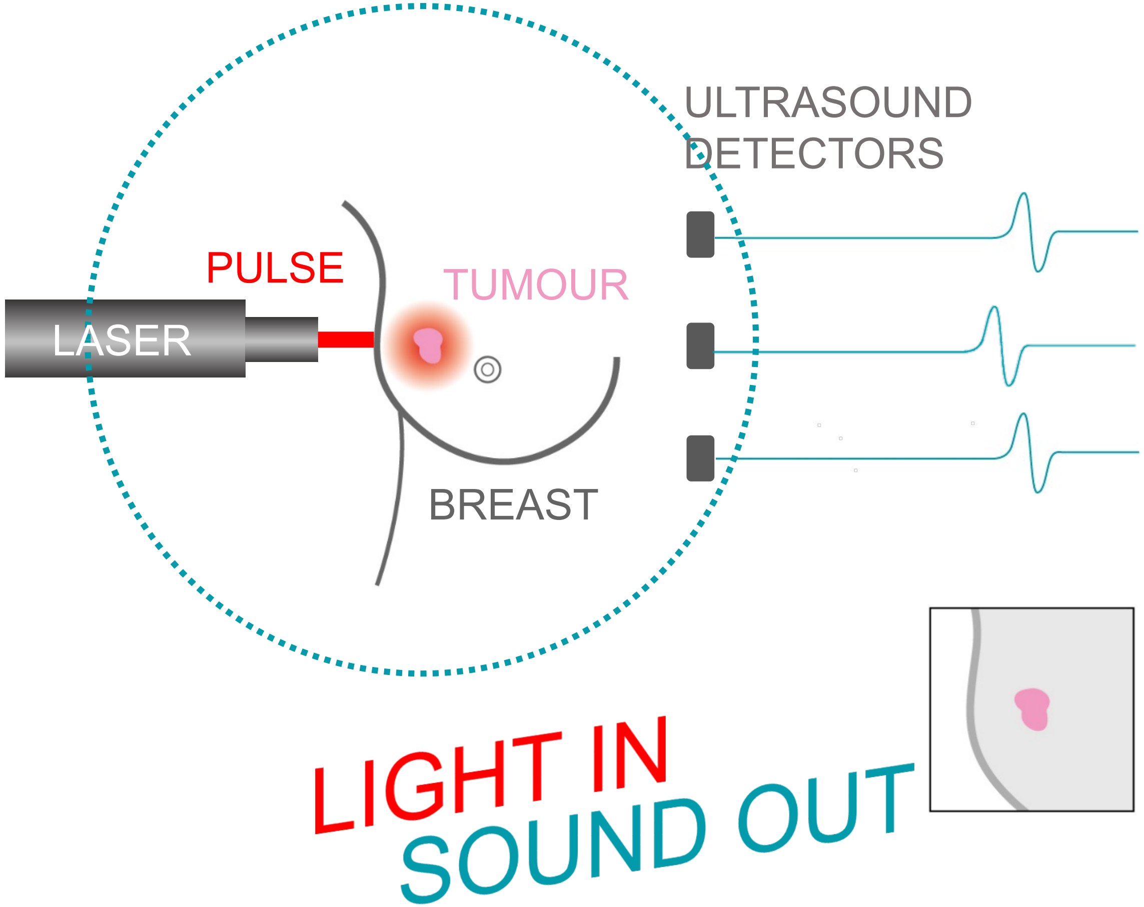




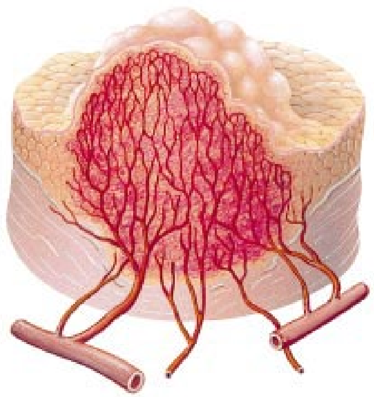
LIGHT IN
SOUND OUT

PHOTOACOUSTIC BREAST IMAGING

(ILLUSTRATIONS MADE BY SJOUKJE SCHOUSTRA – BMPI GROUP, UNI TWENTE)



Folkman, 1996



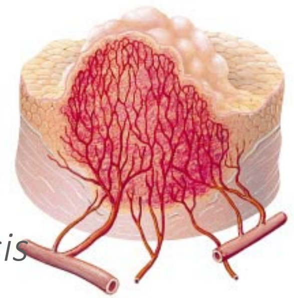
angiogenesis

PHOTOACOUSTIC EFFECT

- LIGHT ABSORPTION
- TEMPERATURE RISE
- EXPANSION
- PRESSURE RISE
- ULTRASOUND WAVE
- SIGNALS
- RECONSTRUCTION

PHOTOACOUSTIC TOMOGRAPHY

Folkman, 1996



angiogenesis

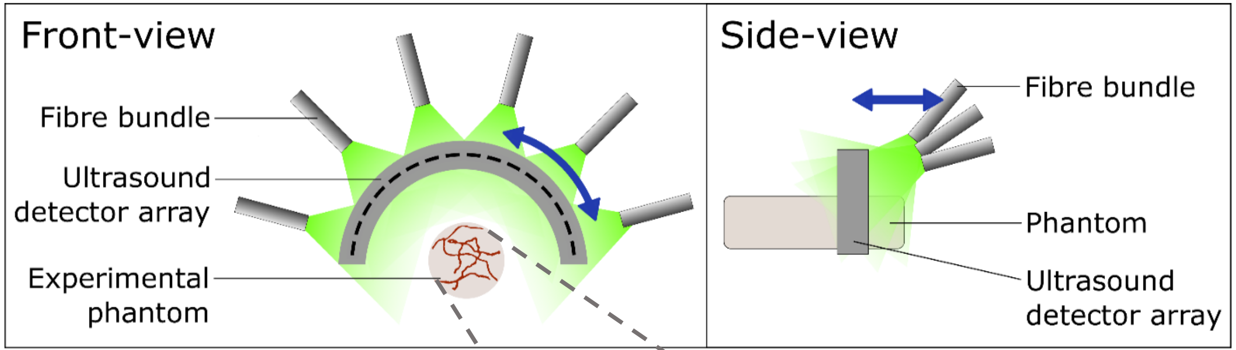


Figure: 2D slice-based imaging with rotating fibres and sensor array.[1]

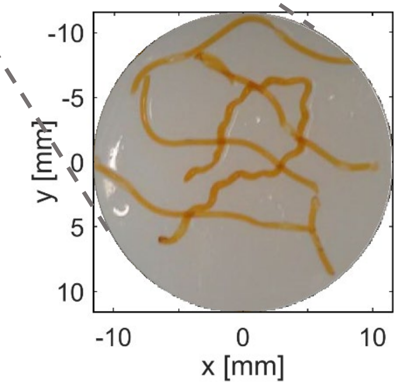
We make use of a projection model with calibration[2]:

$$\tilde{p}(x,t) = \left(\frac{1}{t} \iint_{|x-\tilde{x}|=ct} u(\tilde{x})\, \mathrm{d}\tilde{x} \right) *_t \frac{\partial I(t)}{\partial t} *_t h_{IR}(t),$$

$$\tilde{p}(x,t) = \left(\frac{1}{t} \iint_{|x-\tilde{x}|=ct} u(\tilde{x})\, \mathrm{d}\tilde{x} \right) *_t p_{\mathrm{cal}}(t),$$

$$f = Au := \iint_{|x-\tilde{x}|=ct} u(\tilde{x})\, \mathrm{d}\tilde{x}$$

PAT-operator



[1] Van Es, Vlieg, Biswas, Hondebrink, Van Hespen, Moens, Steenbergen, Manohar - Coregistered photoacoustic and ultrasound tomography of healthy and inflamed human interphalangeal joints (2015)

[2] Wang, Xing, Zeng, Chen - Photoacoustic imaging with deconvolution (2004)

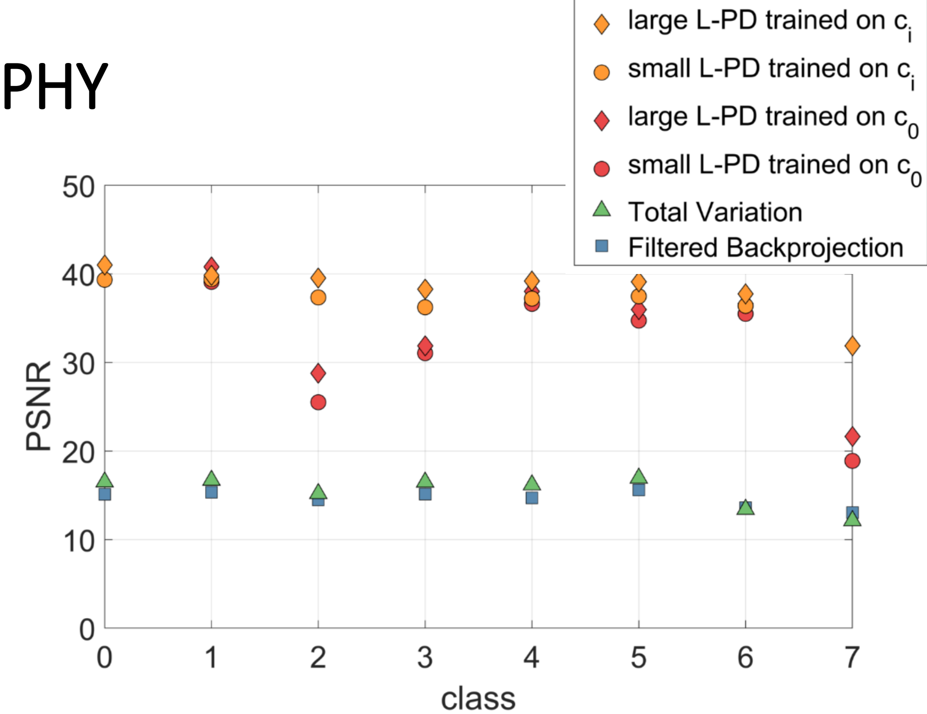
DEEP LEARNING PHOTOACOUSTIC TOMOGRAPHY

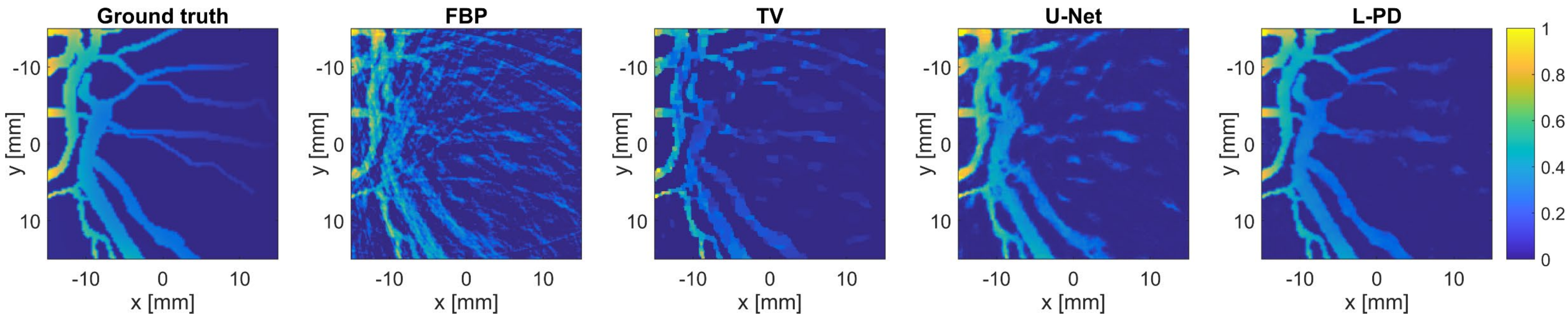
for $n \leftarrow 1$ to N **do**

$$q^{n+1}_{\{1,\dots,k\}} = q^{n}_{\{1,\dots,k\}} + \Gamma_{\Theta_n}\left(q^{n}_{\{1,\dots,k\}}, Au^{n}_{1}, f\right),$$

$$u^{n+1}_{\{1,\dots,k\}} = u^{n}_{\{1,\dots,k\}} + \Lambda_{\Theta_n}\left(u^{n}_{\{1,\dots,k\}}, A^{*}q^{n+1}_{1}\right).$$

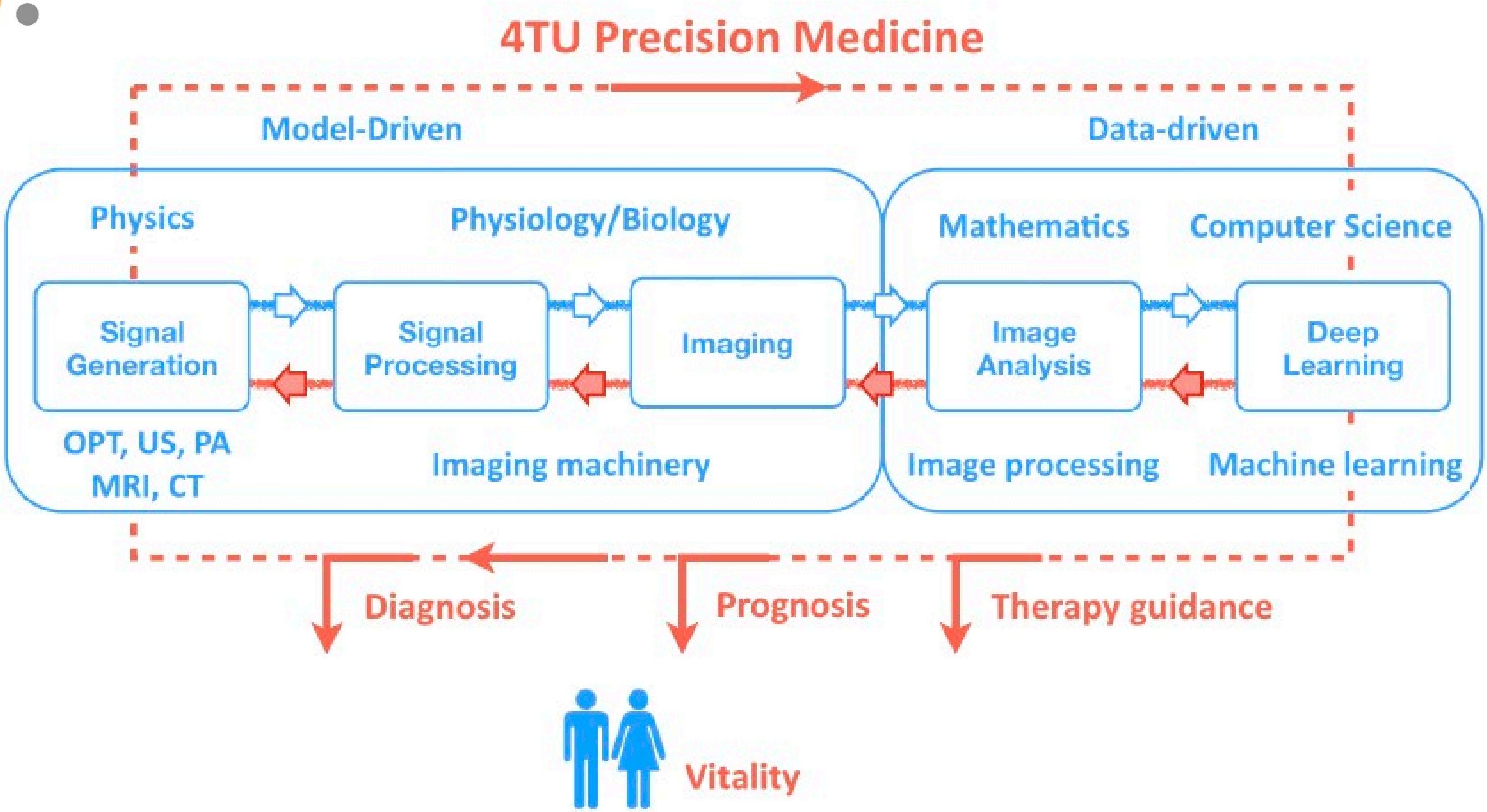
end for





Boink, van Gils, Manohar, Brune - Sensitivity of a partially learned model-based reconstruction algorithm (PAMM 2018)

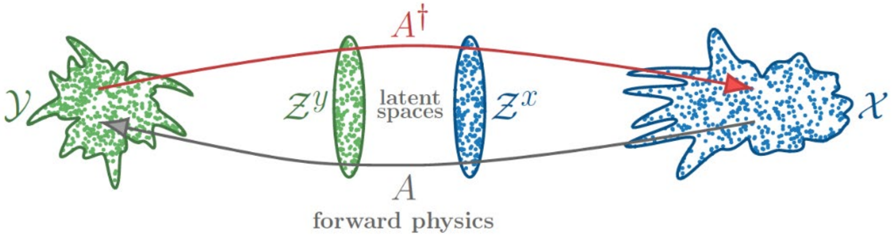
Boink, Manohar, Brune - A Partially Learned Algorithm for Joint Photoacoustic Reconstruction and Segmentation (IEEE 2019)



Deep learning for model-driven imaging

- “classical” physics-based reconstruction enriched by deep learning
- “black-box” deep learning enriched by physical constraints

CONCLUSIONS AND FUTURE WORK



Learned SVD framework for nonlinear decomposition and inversion

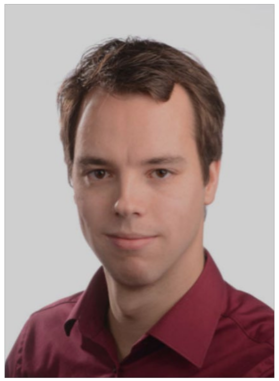
- L-SVD connects data and signal manifold in a straightforward manner with a linear layer
- Via nonlinear AE, L-SVD has superior performance with interpretable results
- Hybrid autoencoding provides regularization for reconstruction
- L-SVD has a lot of freedom in the exact architecture chosen. Future work will investigate more advanced autoencoders for larger-scale inverse problems. (GANs, VAEs)

Y.E. Boink, C. Brune - Learned SVD: Solving Inverse Problems via Hybrid Autoencoding (arXiv preprint, Nov 2019)

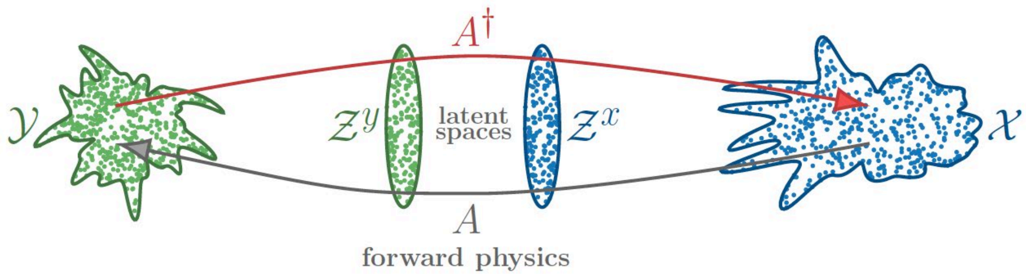
4TU.AMI

NDNS+

THANKS FOR YOUR ATTENTION



Yoeri Boink



Leonie Zeune

Srirang Manohar

Michel Versluis

Stephan van Gils

Raymond Veldhuis

Lu Liu