UNIVERSITY OF TWENTE.

Deep Inversion, Autoencoders for Learned Regularization of Inverse Problems

Yoeri Boink*†, Srirang Manohar†, Leonie Zeune*‡, Leon Terstappen‡, Stephan van Gils*, *Christoph Brune**

Biomedical Photonic Imaging†

Medical Cell Biophysics‡

Applied Mathematics*

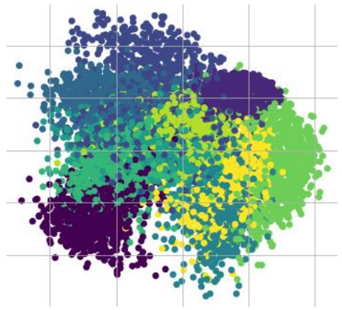
OUTLINE

1. **Robustness** of learned primal-dual (L-PD) reconstruction in photoacoustic tomography (PAT)

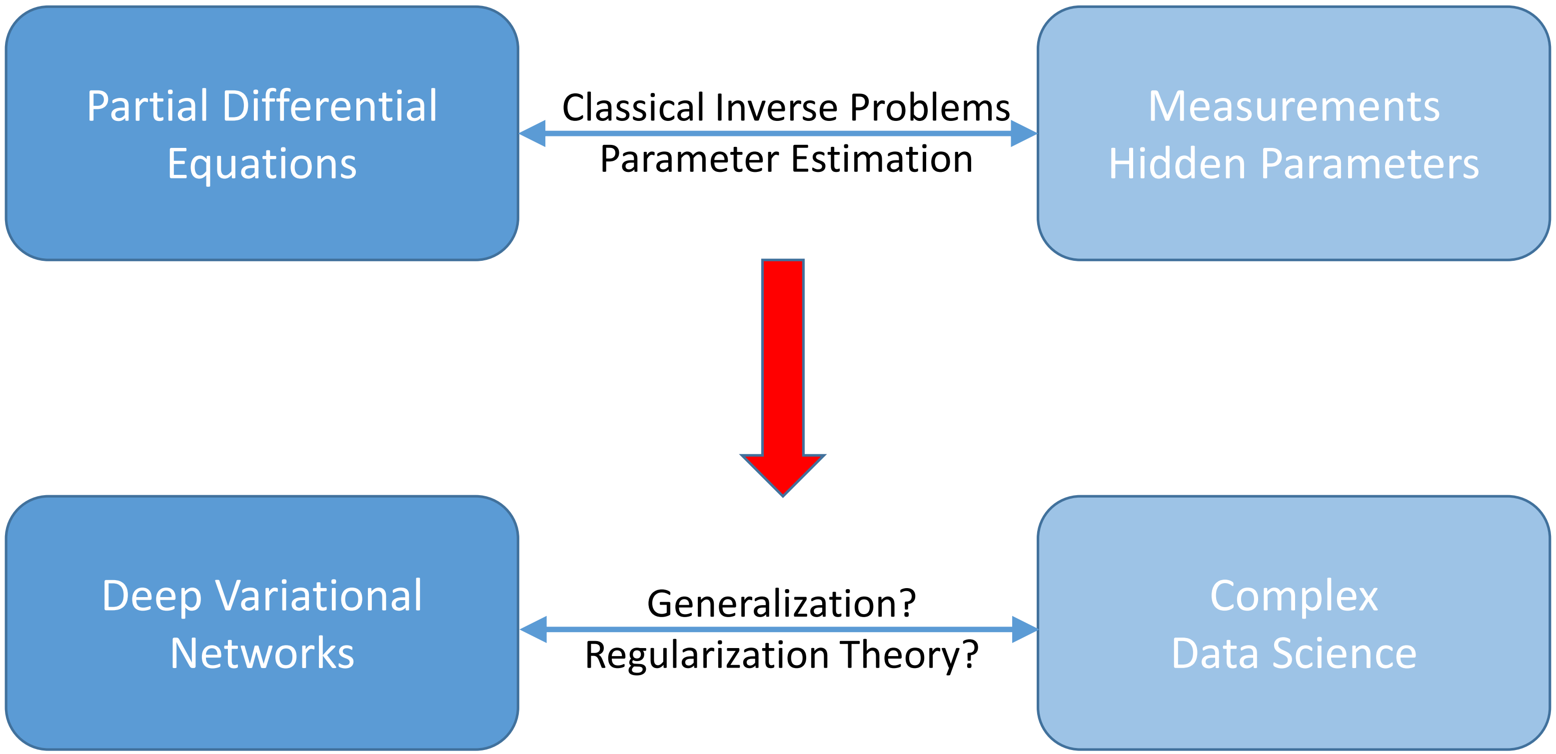
2. **Functional learning** with an unrolled gradient descent (GD) scheme
   → guaranteed convergence and stability!

3. **Learn latent representation** of data space and image space via variational autoencoder (VAE)

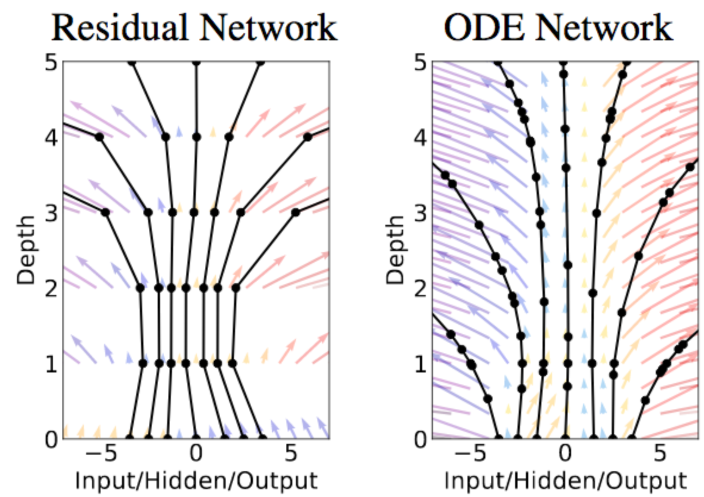
Density fitting $g_\theta(\{z_i\}_i) \approx \{x_i\}_i$ ⟷ Optimal transport map d_ξ ⟷ Auto-encoders $g_\theta(d_\xi(x_i)) \approx x_i$



VARIATIONAL METHODS AND DEEP LEARNING



VARIATIONAL METHODS AND DEEP LEARNING

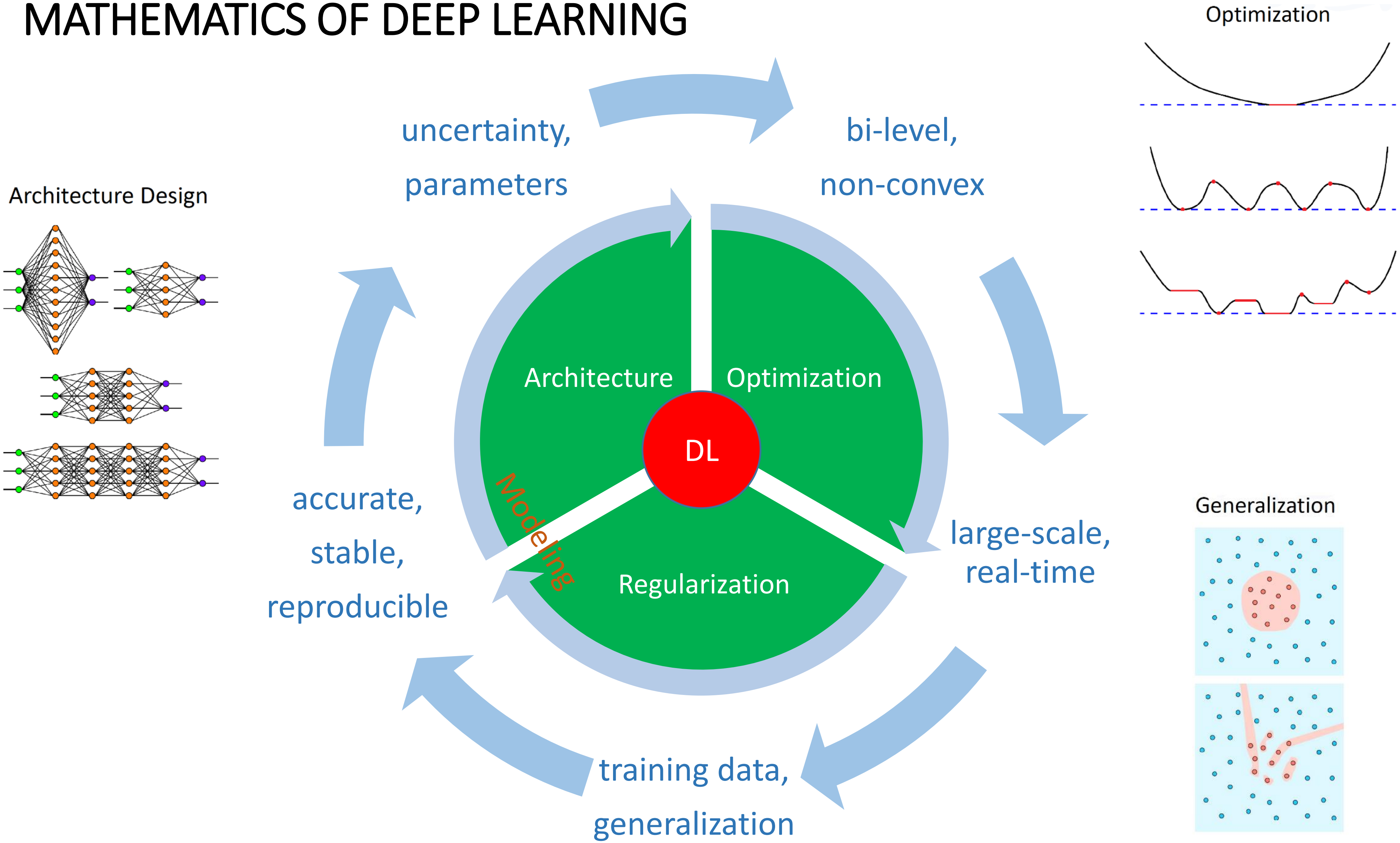


Deep Residual Neural Networks
are connected to
Partial Differential Equations

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	GANs VAEs?	Multiple populations?	Residual? Skip connections?

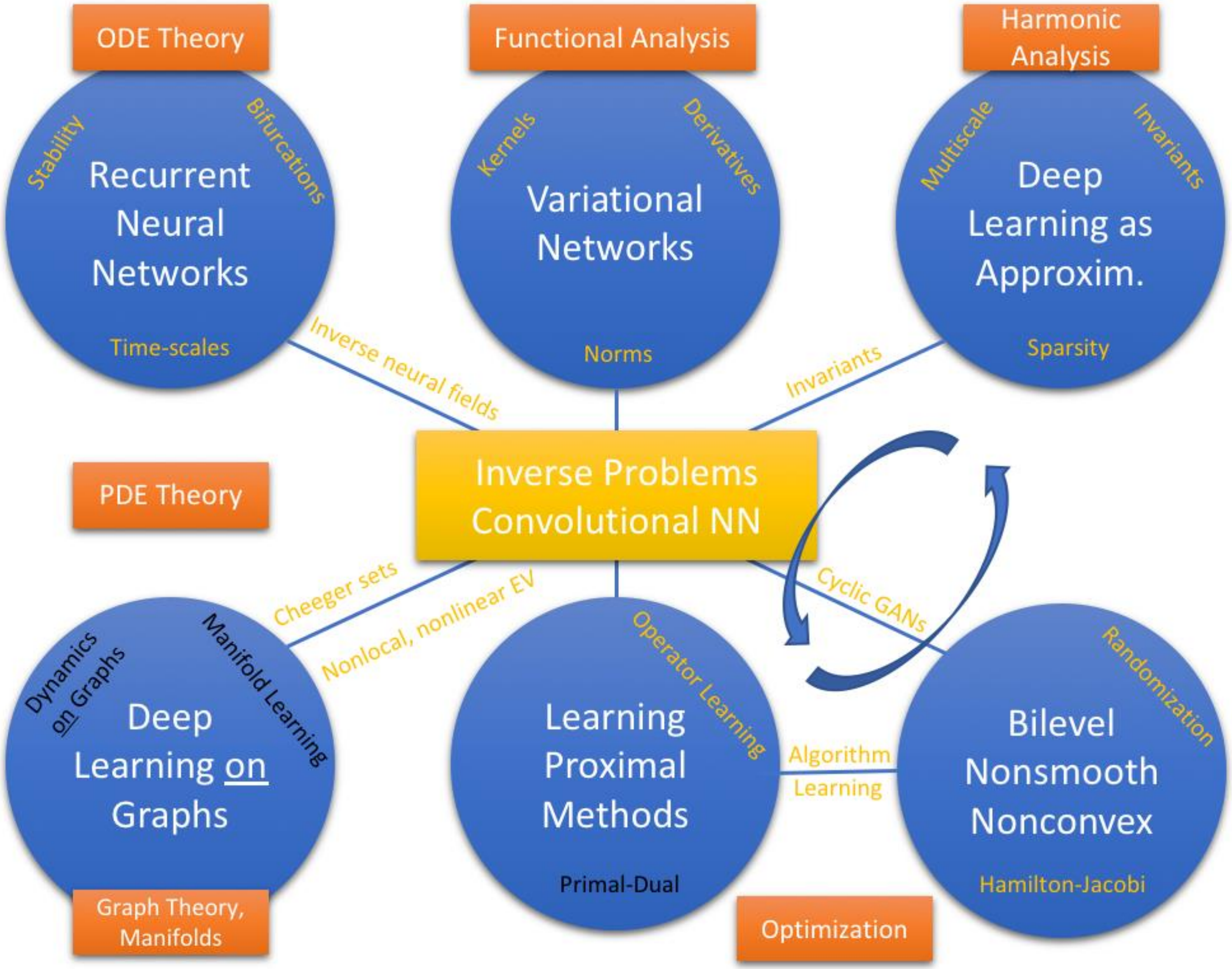
- Chen, Pock - Trainable Nonlinear Reaction Diffusion, 2016
- Mallat - Understanding deep convolutional networks, 2016
- Chen et al - Neural Ordinary Differential Equations, 2018
- Ciccone et al - Stable Deep Networks from Non-Autonomous DEs, 2018
- Haber, Ruthotto - Stable architectures for deep neural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

MATHEMATICS OF DEEP LEARNING

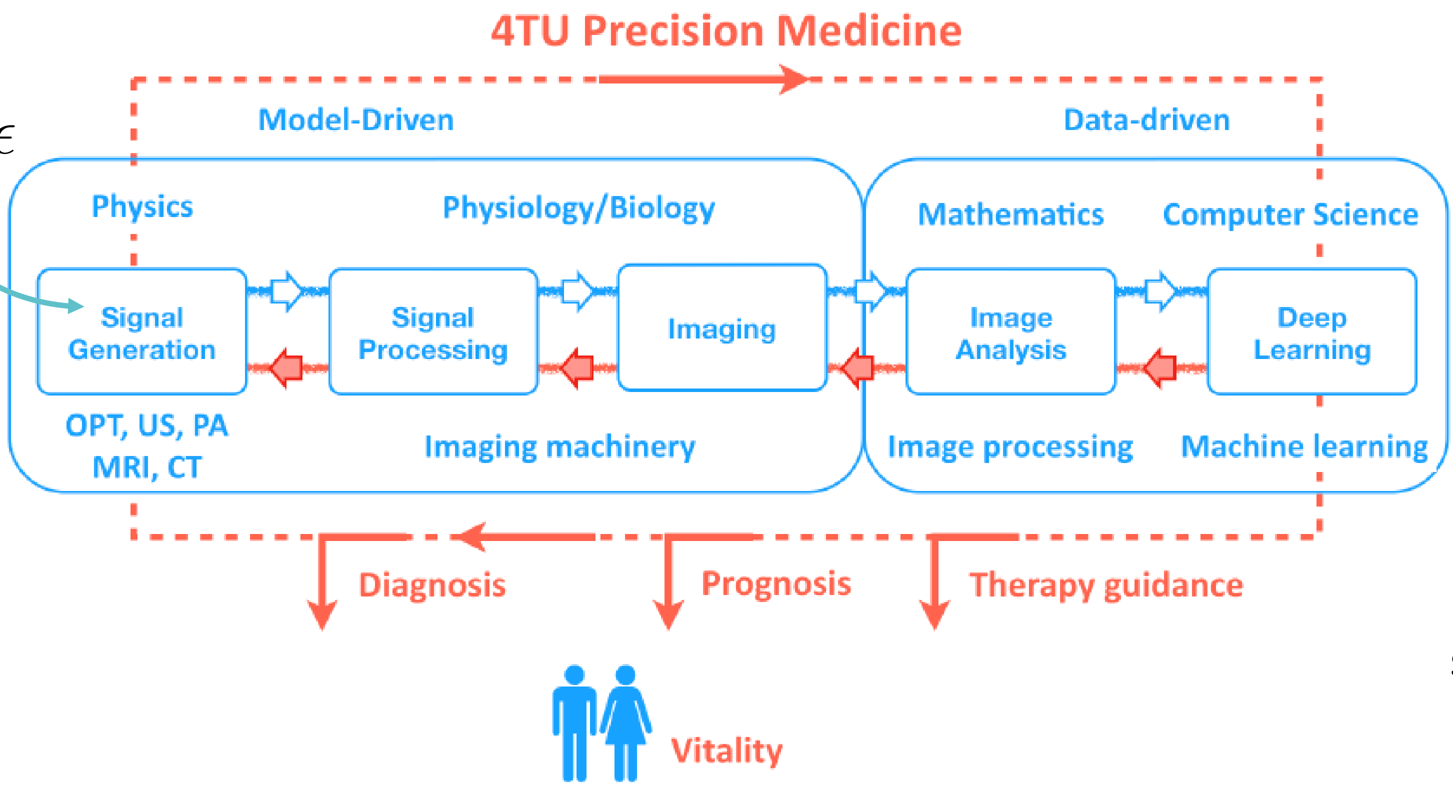


Vidal et al. Mathematics of Deep Learning, 2018

DEEPER INSIGHTS INTO DEEP INVERSION



$$f = Au^* + \epsilon$$



natural
robust

performance
simple

Challenges: C1: sparse/big data; C2: multi-dimensionality, heterogeneity
C3: non-linearity; C4: super-resolution

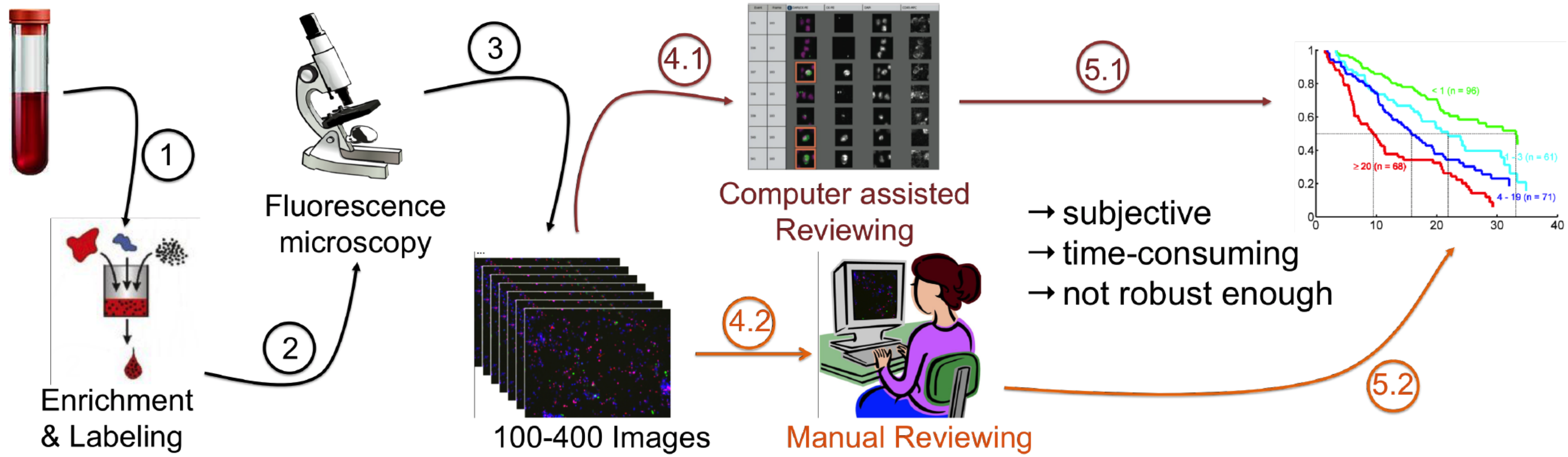
Deep learning for model-driven imaging

- "classical" physics-based reconstruction enriched by deep learning → latent operator parameters, robustness
- "black-box" deep learning enriched by physical constraints → more natural, latent data structures, robustness

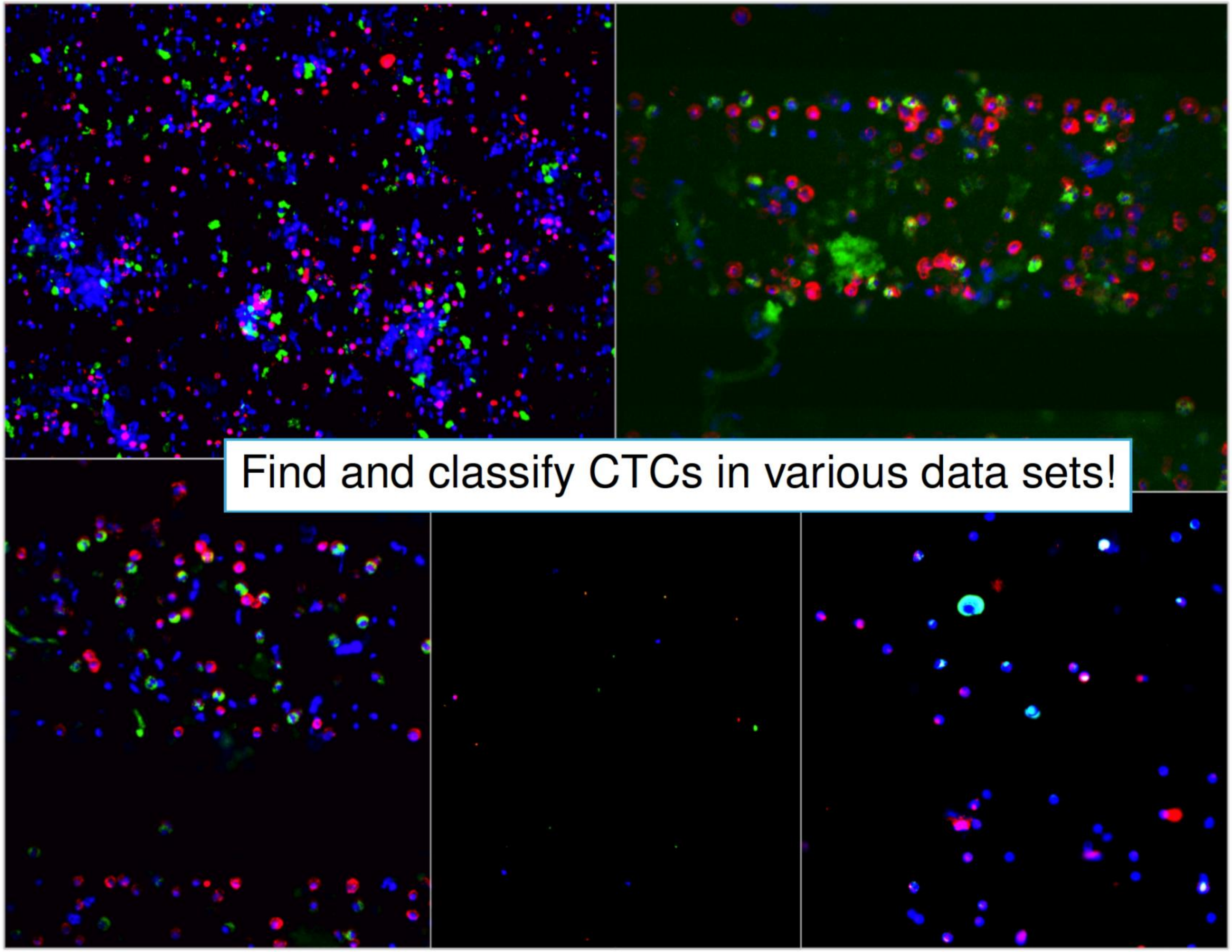
FINDING NEEDLES IN A HAYSTACK

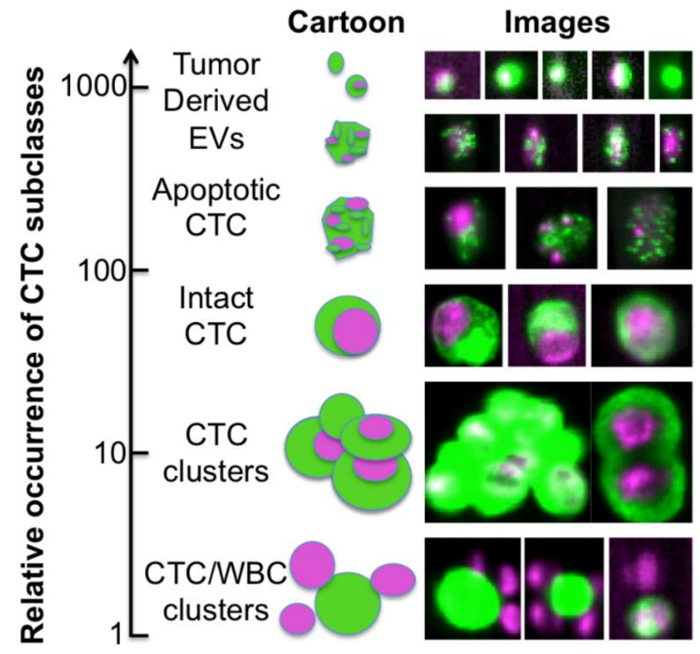
Cancer-ID aims to validate blood-based biomarkers for cancer

- cells dissociate from primary tumor and invade blood circulation
- rare cell events, challenging to detect
- **circulating tumor cell** (CTC) count has prognostic value for survival outcome
- no overall CTC definition exists yet



FINDING NEEDLES IN A HAYSTACK





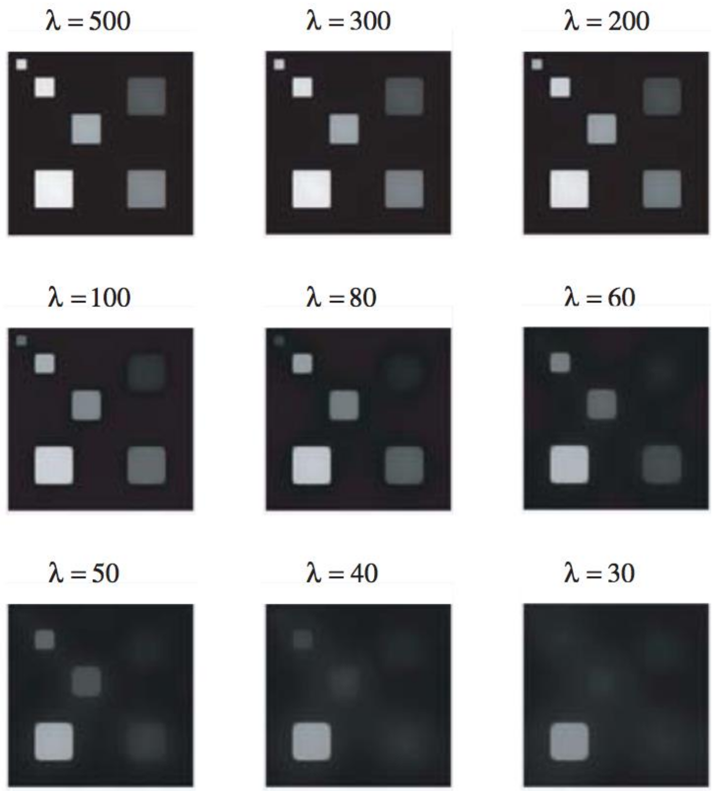
Denoising and Scale Analysis by Local Diffusion

- given noisy input $u^\delta = u + \delta$
- denoise image by minimizing the Total Variation (TV) flow

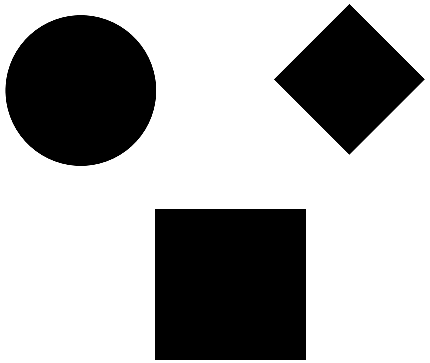
$$\boxed{\begin{aligned} u_t &= -p \text{ for } p \in \partial TV(u) \\ u(t=0) &= u^\delta \end{aligned}}$$

→ Time-discrete case: solving in every step the ROF [Rudin,92] problem:

$$\frac{1}{2}||u - u_n||_2^2 + \frac{1}{\lambda} TV(u) \rightarrow \min_u$$



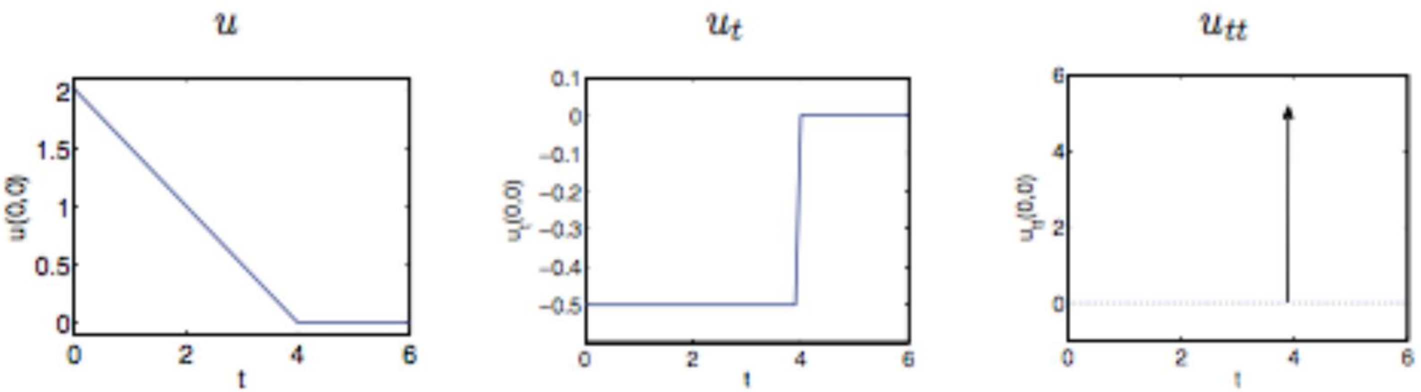
Denoising and Scale Analysis by Local Diffusion



Idea: Solution of nonlinear eigenvalue problem

$$\lambda u \in \partial TV(u)$$

transformed to a peak in the spectral domain

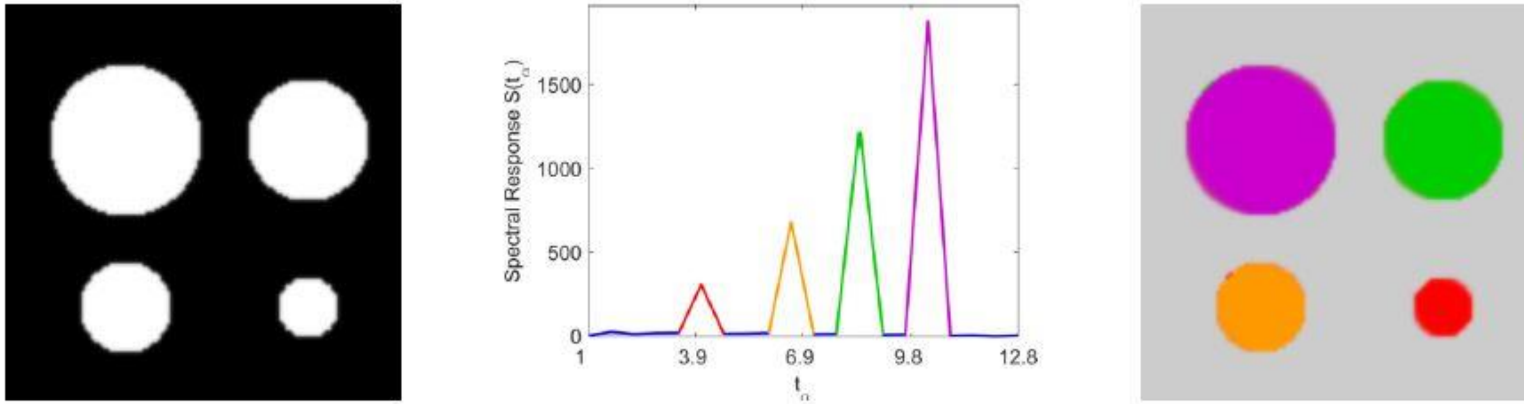


→ time-point *t* where the eigenfunctions are completely removed depends on size and height of the disc, thus scale indicator

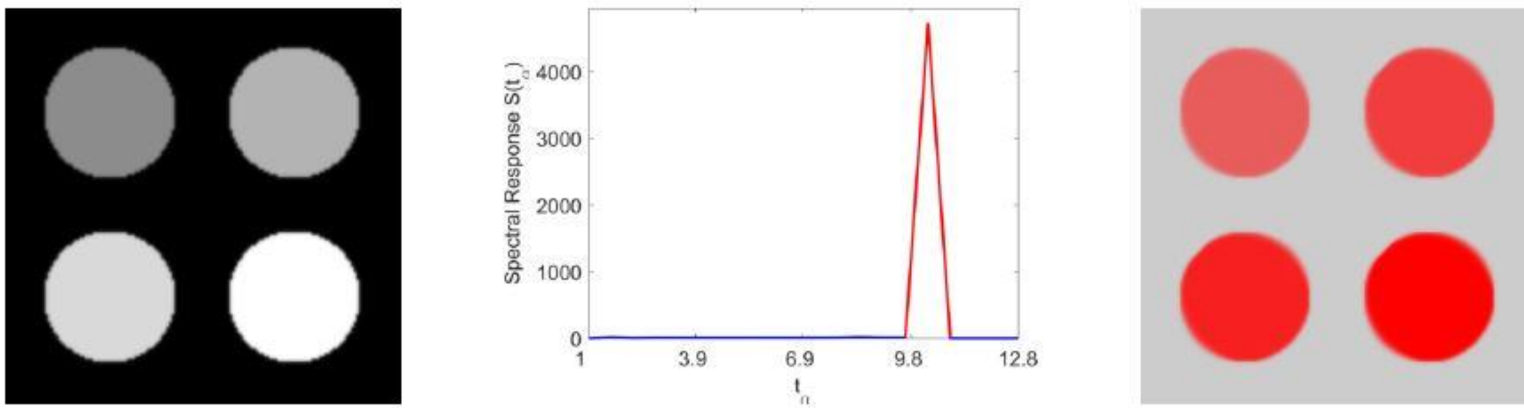
[Gilboa 2013,2014],[Horesh, Gilboa 2015],[Burger et al., 2015,2016]
[Aujol, Gilboa, Papadakis, 2015, 2017]

Contrast Invariance of L1-TV Denoising

Size Scales

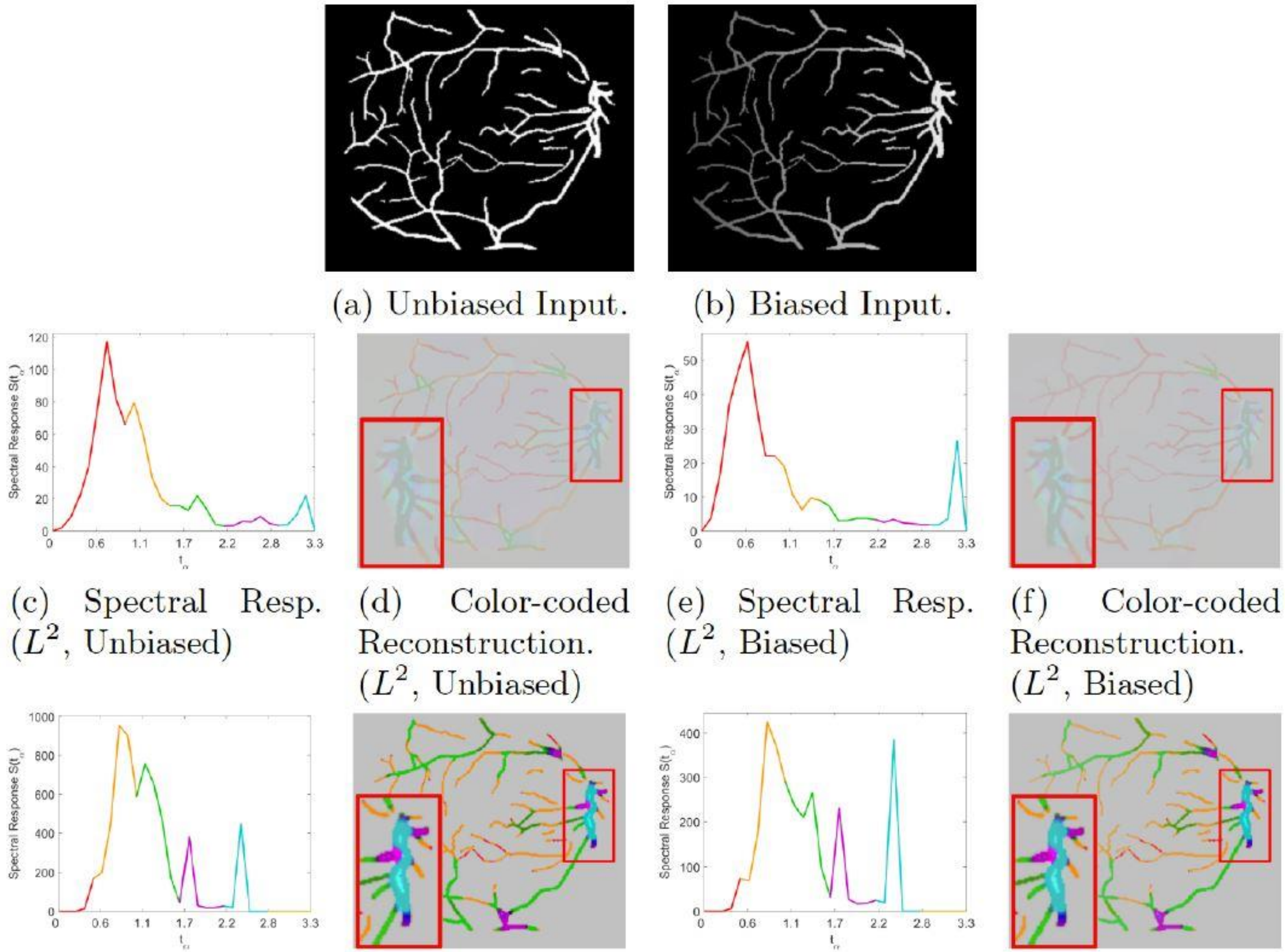


Intensity Scales



L2-norm

L1-norm



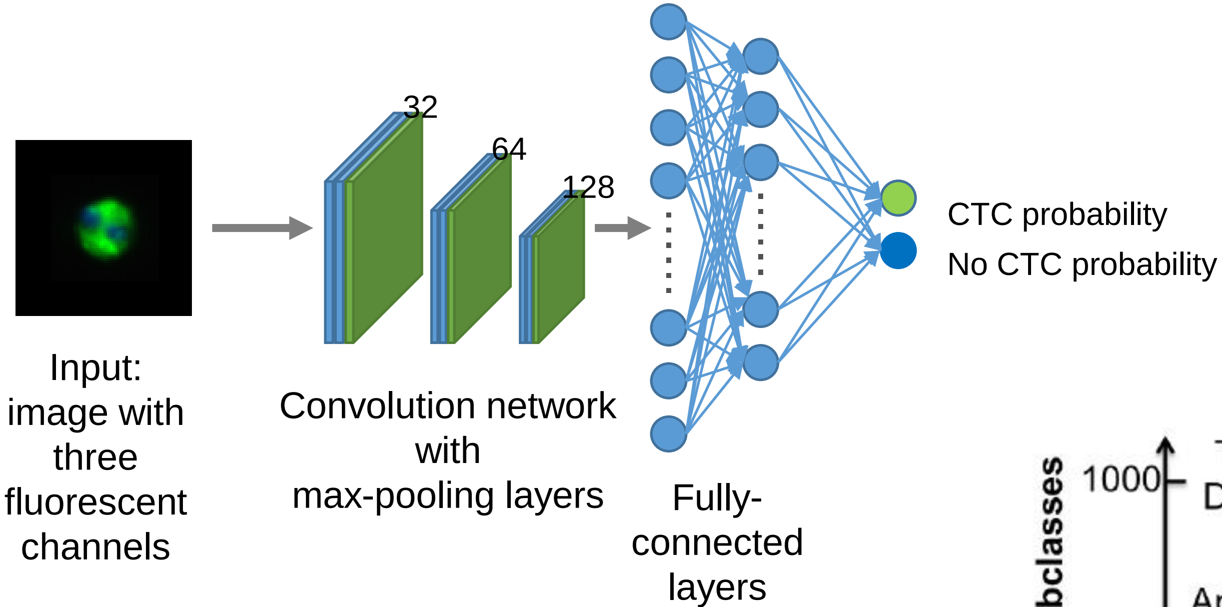
(a) Unbiased Input. (b) Biased Input.

(c) Spectral Resp. (L^2, Unbiased) (d) Color-coded Reconstruction. (L^2, Unbiased) (e) Spectral Resp. (L^2, Biased) (f) Color-coded Reconstruction. (L^2, Biased)

(g) Spectral Resp. (L^1, Unbiased) (h) Color-coded Reconstruction. (L^1, Unbiased) (i) Spectral Resp. (L^1, Biased) (j) Color-coded Reconstruction. (L^1, Biased)

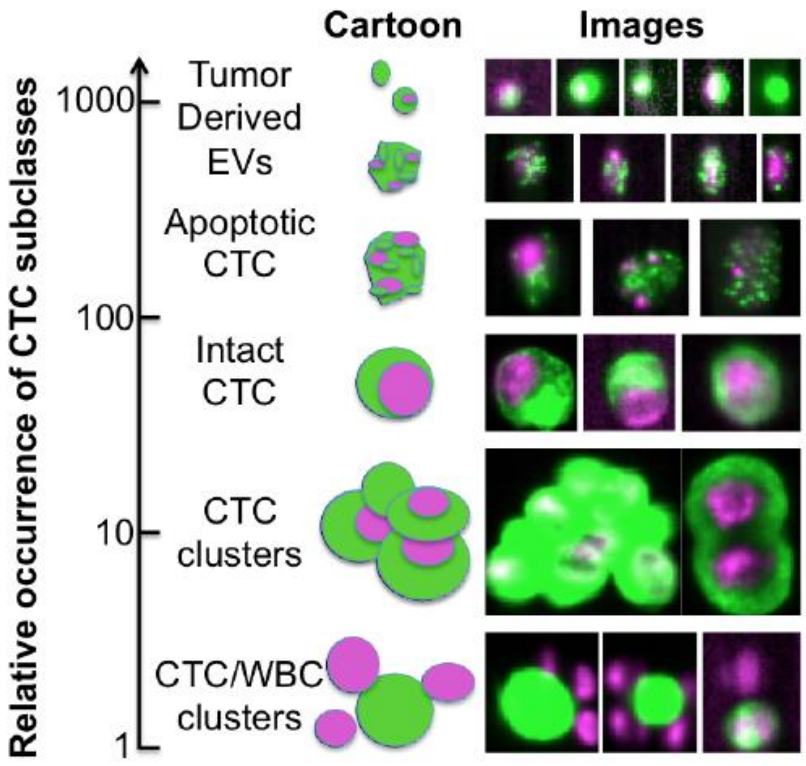
Zeune et al - Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis, 2017

CNN CLASSIFICATION FOR CANCER-ID



Manual \ DL	CTC	No CTC
CTC	0.9722	0.0278
No CTC	0.010	0.990

→ **Accuracy of 98,62%**



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 some (fulfilled) assumptions, *G(x)* is the gradient field of an energy *E(x)*

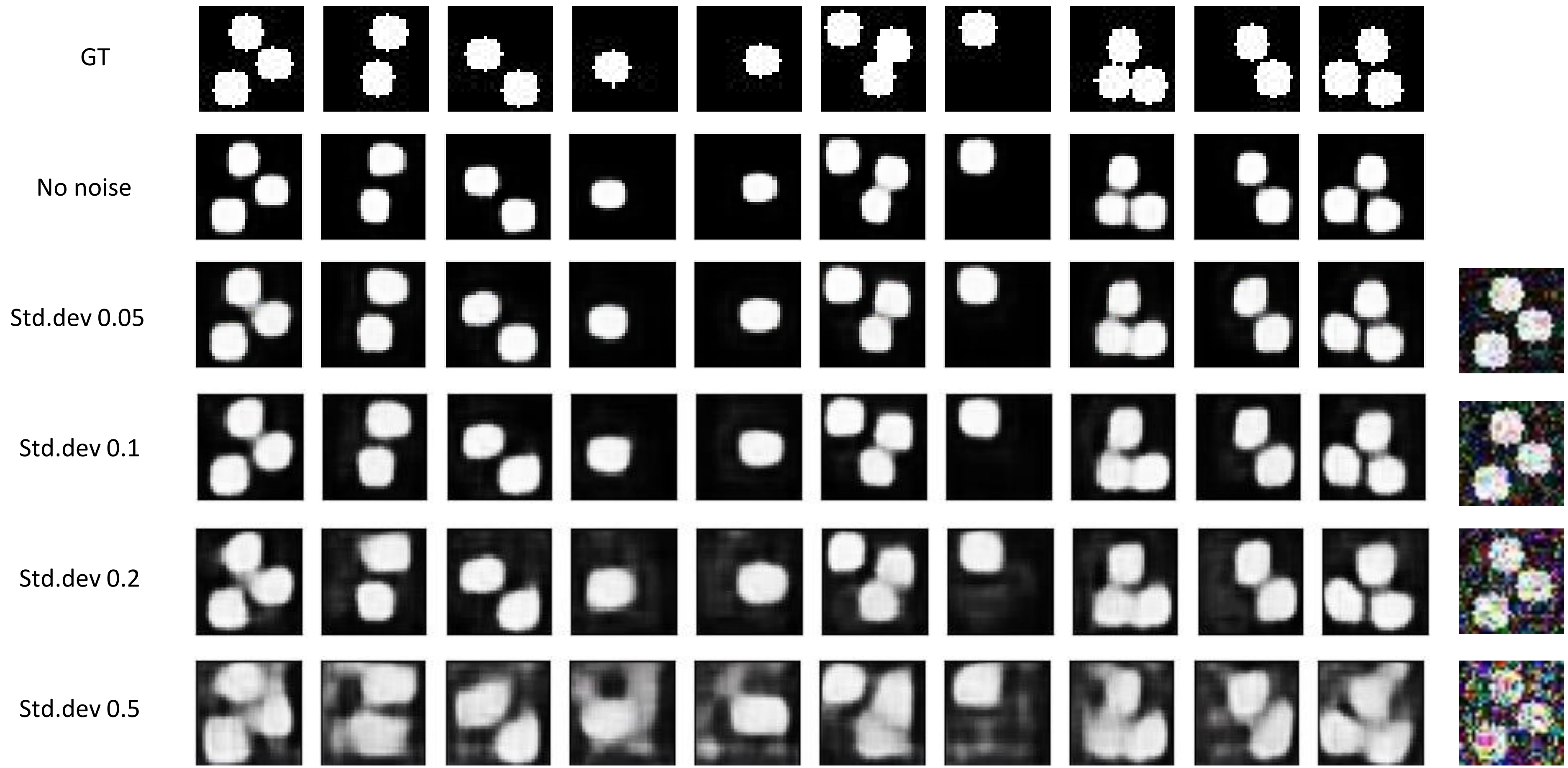
$$G(x) = f(x) - x \sim -\partial_x E$$

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

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

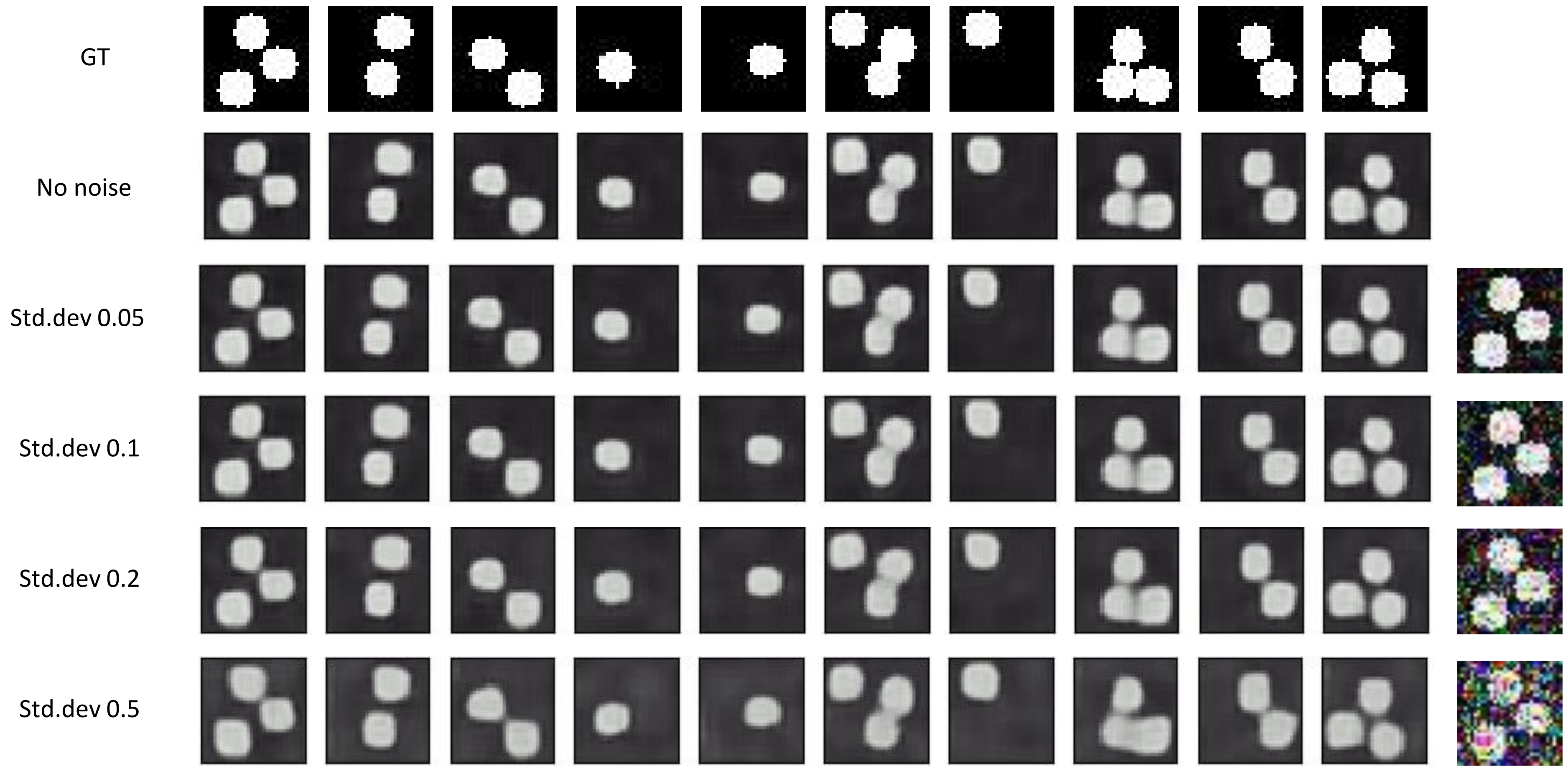
Convolutional AE

Trained on no noise data, 6000 training sets, 1000 test sets
3 Convolution (32 filters in total) + Pooling blocks for Encoder + Decoder → 4963 parameters



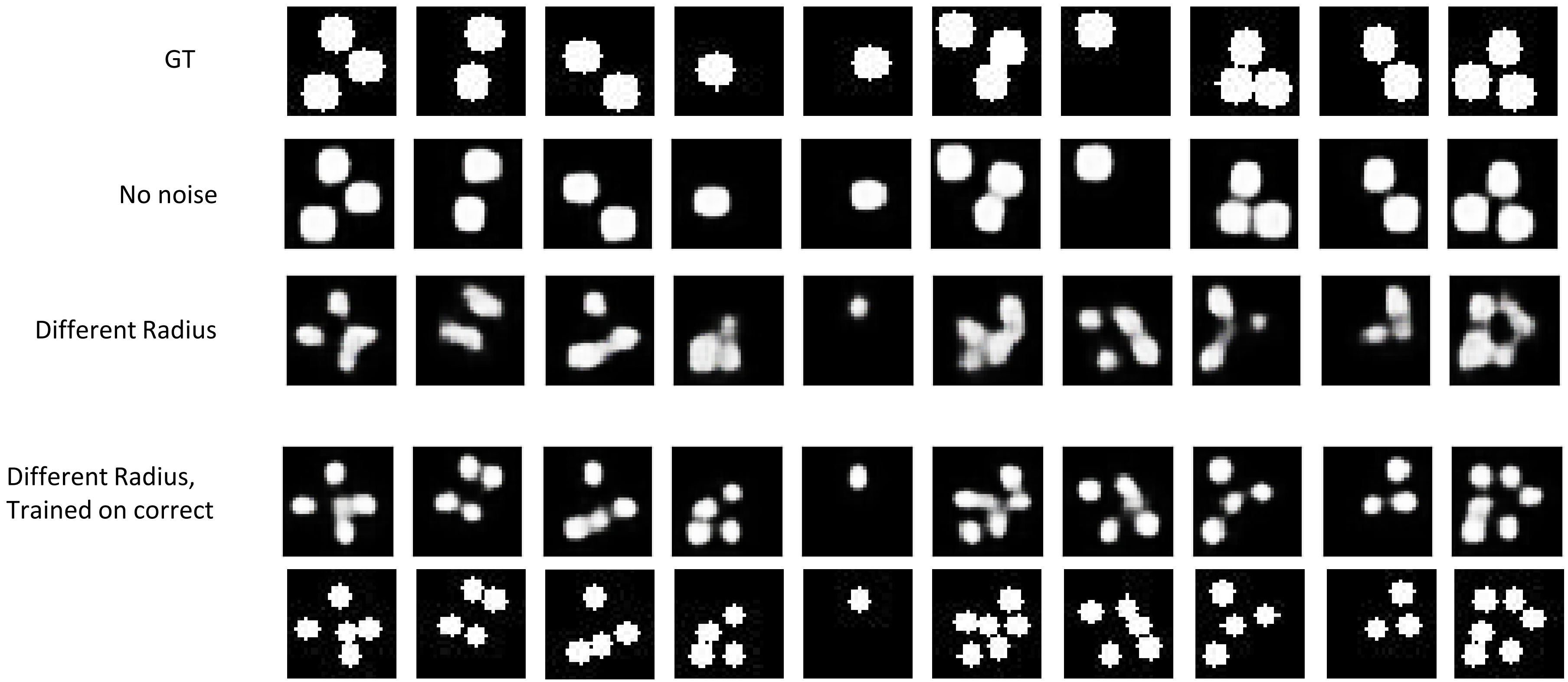
Convolutional AE

Trained on noisy data (std dev 0.2), 6000 training sets, 1000 test sets
3 Convolution (32 filters in total) + Pooling blocks for Encoder + Decoder → 4963 parameters

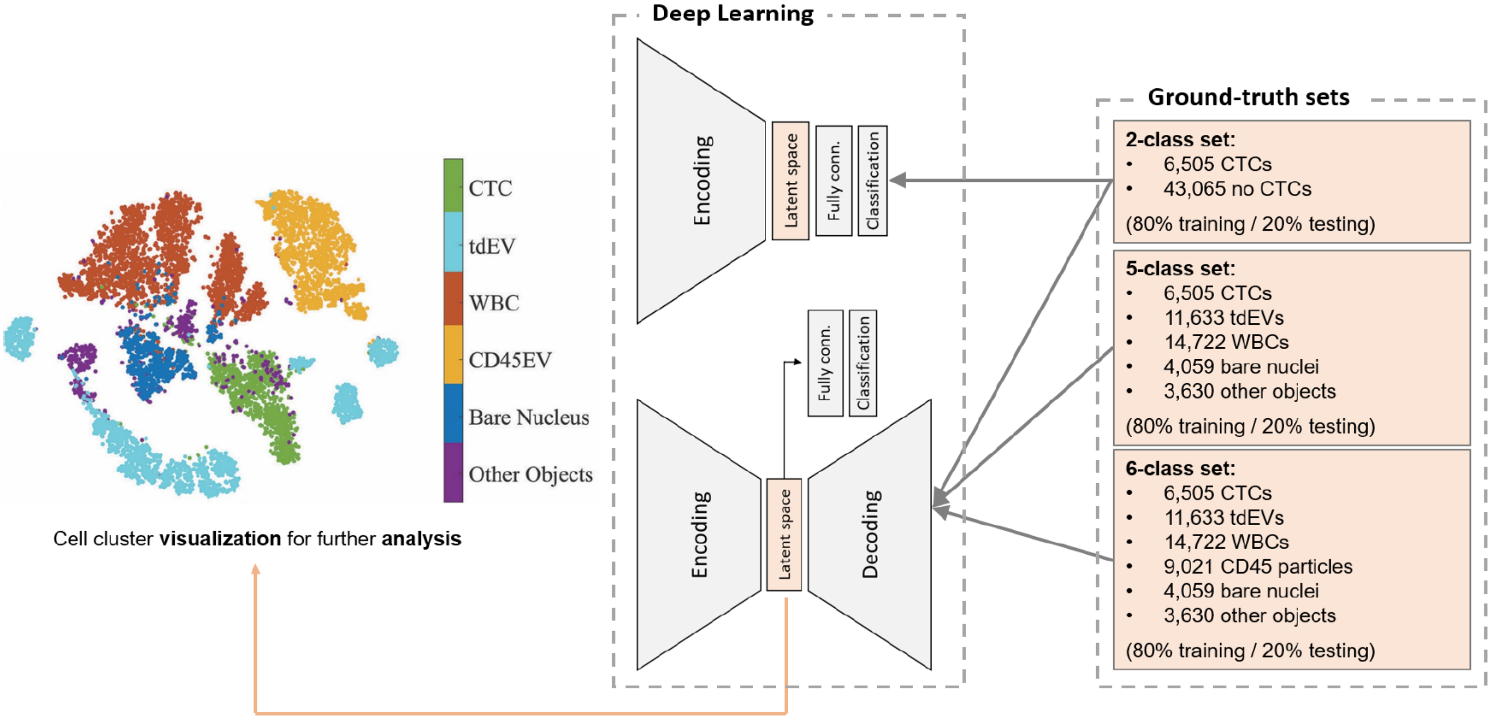


Convolutional AE

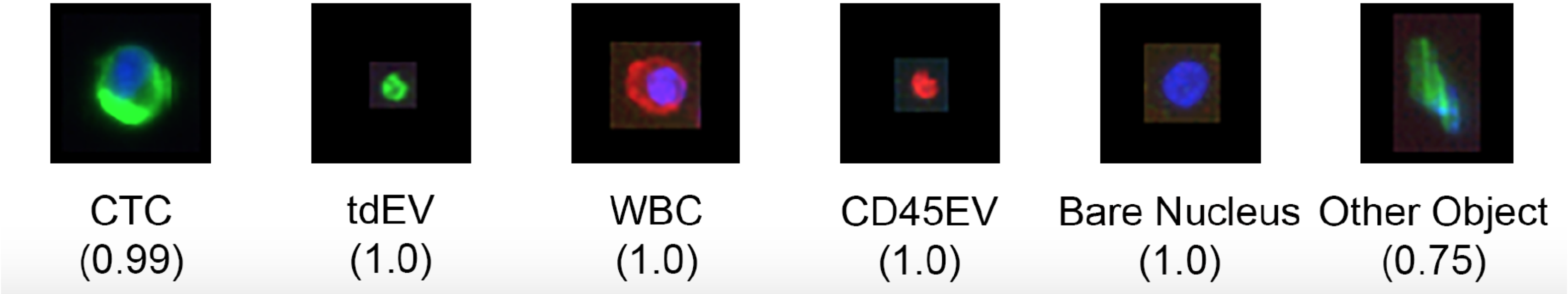
Trained on no noise data, 6000 training sets, 1000 test sets
3 Convolution (32 filters in total) + Pooling blocks for Encoder + Decoder → 4963 parameters



DEEP LEARNING OF CIRCULATING TUMOR CELLS



Leonie Zeune



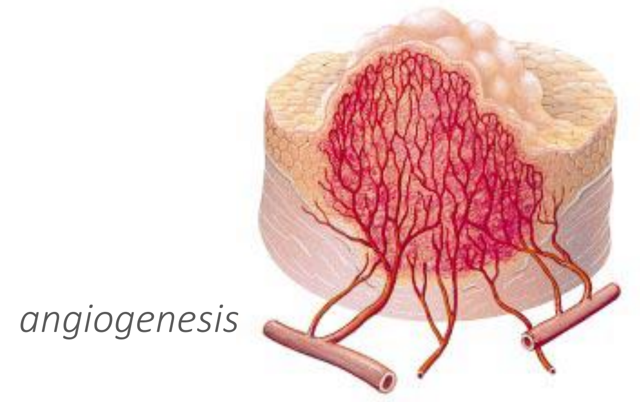
EU-IMI key opinion leaders: "*A notable highlight is the development of the open source image analysis program ACCEPT*

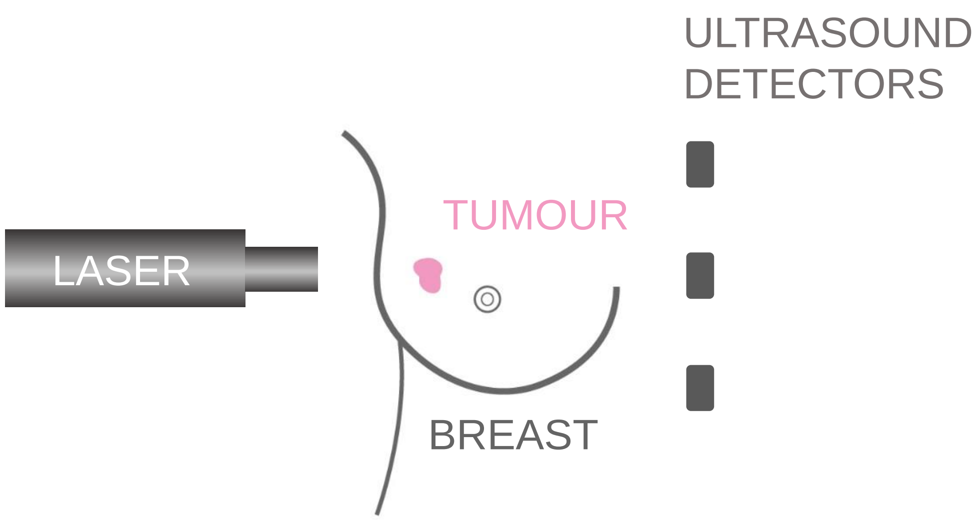
Zeune et al - Deep learning for tumor cell classification (2019)

PHOTOACOUSTIC BREAST IMAGING

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

Folkman, 1996



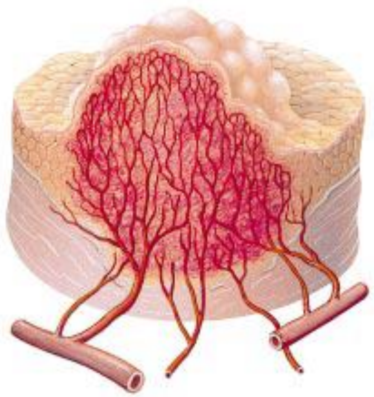


LIGHT IN
SOUND OUT

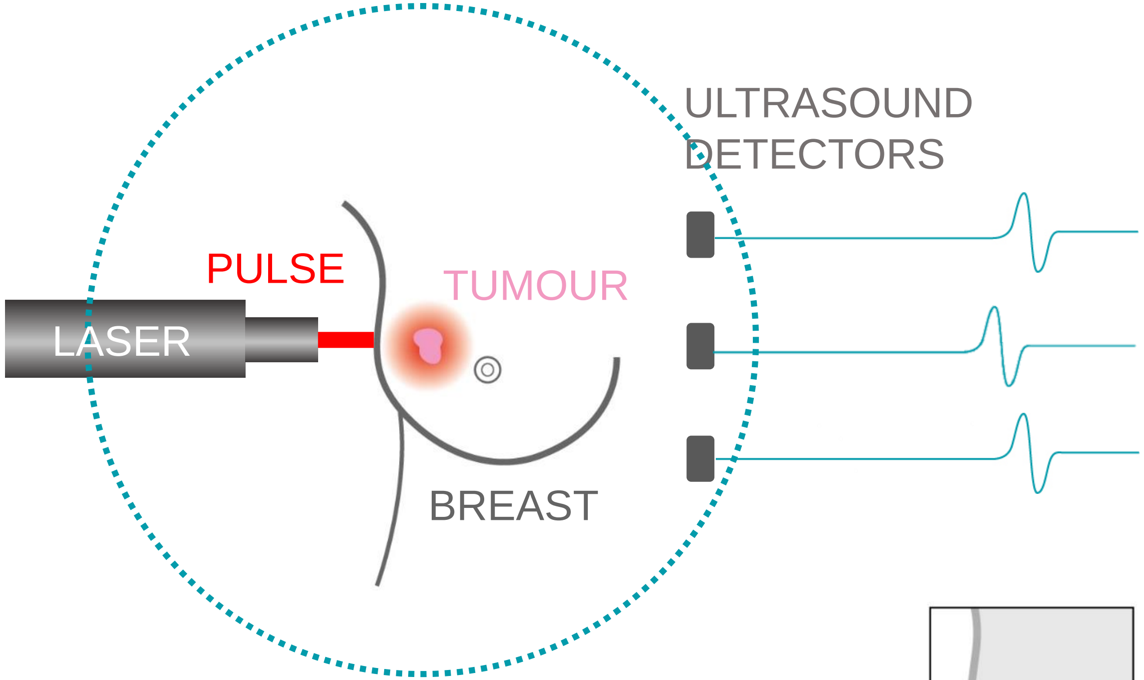
PHOTOACOUSTIC BREAST IMAGING

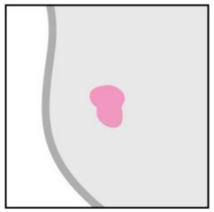
(ILLUSTRATIONS MADE BY SJOUKJE SCHOUSTRA – BMPI GROUP, UNI TWENTE)

Folkman, 1996



angiogenesis





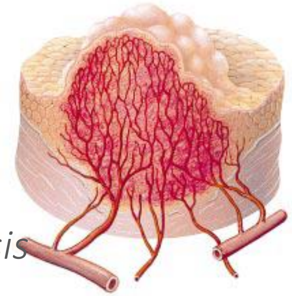
LIGHT IN
SOUND OUT

PHOTOACOUSTIC EFFECT

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

PHOTOACOUSTIC TOMOGRAPHY

Folkman, 1996



angiogenesis

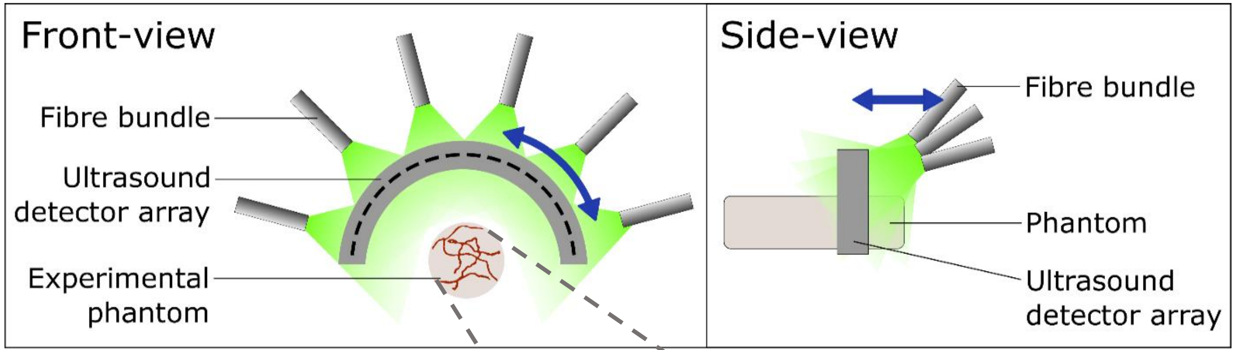


Figure: 2D slice-based imaging with rotating fibres and sensor array.[1]

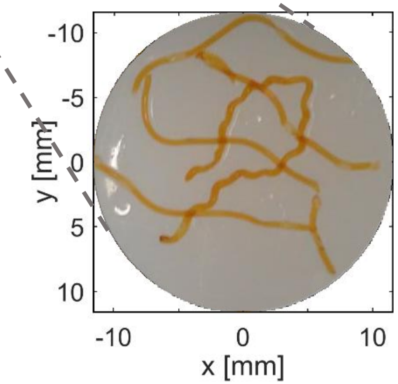
We make use of a projection model with calibration[2]:

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

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

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

PAT-operator

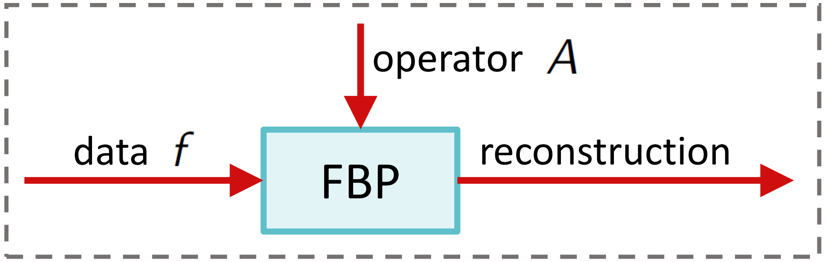


[1] Van Es, Vlieg, Biswas, Hondebrink, Van Hespen, Moens, Steenbergen, Manohar - Coregistered photoacoustic and ultrasound tomography of healthy and inflamed human interphalangeal joints (2015)

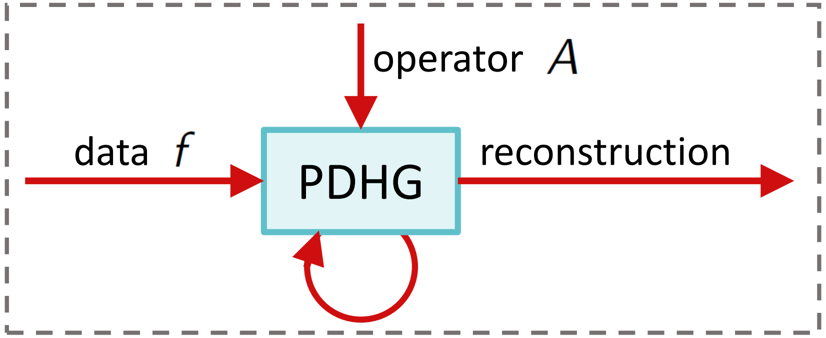
[2] Wang, Xing, Zeng, Chen - Photoacoustic imaging with deconvolution (2004)

INVERSE RECONSTRUCTION METHODS

- **Direct** reconstruction: filtered backprojection (FBP)

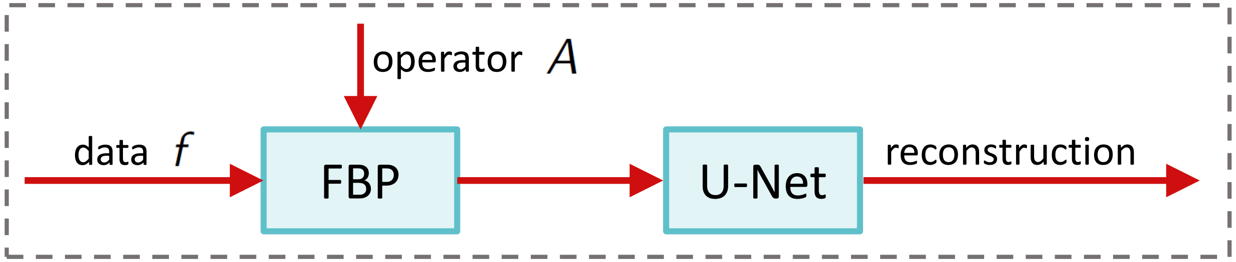


- **Iterative** reconstruction: total variation (TV)
  (solved with PDHG)



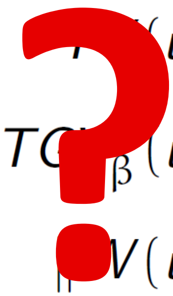
- **Learned** post-processing: U-Net [3]

[3] Jin, McCann, Froustey, Unser - Deep Convolutional Neural Network for Inverse Problems in Imaging (2017)



MODEL-BASED REGULARISED RECONSTRUCTION - CHOOSING FUNCTION SPACES

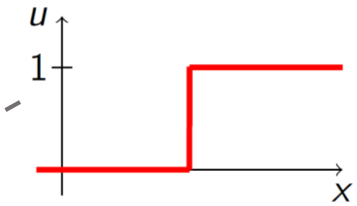
$$\min_{u \in L^2(\Omega)} \left\{ ||Au - f||^2_{L^2(\Sigma)} + \alpha R(u) \right\}$$

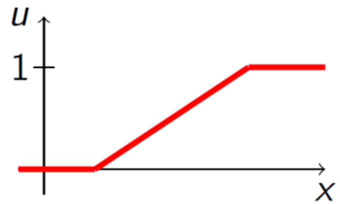


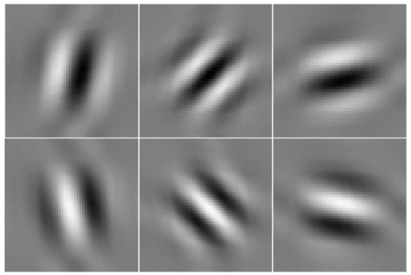
(u)

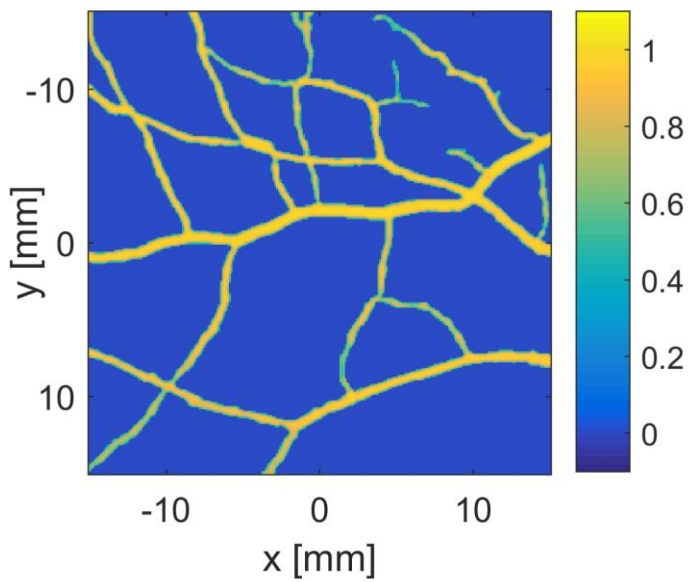
$TG \ldots_\beta(u)$

$\ldots V(u)||_{L_1(\mathcal{W})}$









Rudin, Osher, Fatemi - Nonlinear total variation based noise removal algorithms (1992)

Bredies, Kunisch, Pock - Total Generalised Variation (2010)

Kingsbury - The dual-tree complex wavelet transform a new efficient tool for image restauration and enhancement (1998)

Boink, Lagerwerf, Steenbergen, van Gils, Manohar, Brune - A framework for directional and higher-order reconstruction in photoacoustic tomography (2018)

FROM MODEL-DRIVEN TO DATA-DRIVEN

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

$$q^{n+1} = \mathrm{prox}_{\sigma F_f^*}\left(q^n + \sigma A\left[(1+\theta)u^n - \theta u^{n-1}\right]\right),$$

$$u^{n+1} = \mathrm{prox}_{\tau G}(u^n - \tau A^* q^{n+1}).$$

end for

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

$$q^{n+1} = q^n + \Gamma_{\Theta_n}\left(q^n, A\left[(1+\theta)u^n - \theta u^{n-1}\right], f\right),$$

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

end for

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

$$q^{n+1}_{\{1,\ldots,k\}} = q^n_{\{1,\ldots,k\}} + \Gamma_{\Theta_n}\left(q^n_{\{1,\ldots,k\}}, Au^n_1, f\right),$$

$$u^{n+1}_{\{1,\ldots,k\}} = u^n_{\{1,\ldots,k\}} + \Lambda_{\Theta_n}\left(u^n_{\{1,\ldots,k\}}, A^* q^{n+1}_1\right).$$

end for

- No regularisation parameter;
- Better robustness to noise;
- Faster reconstruction.

However,

- No proven stability or convergence.

Meinhardt, Möller, Hazirbas, Cremers - Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems (2017)

Adler, Öktem – Learned Primal-Dual Reconstruction (2017)

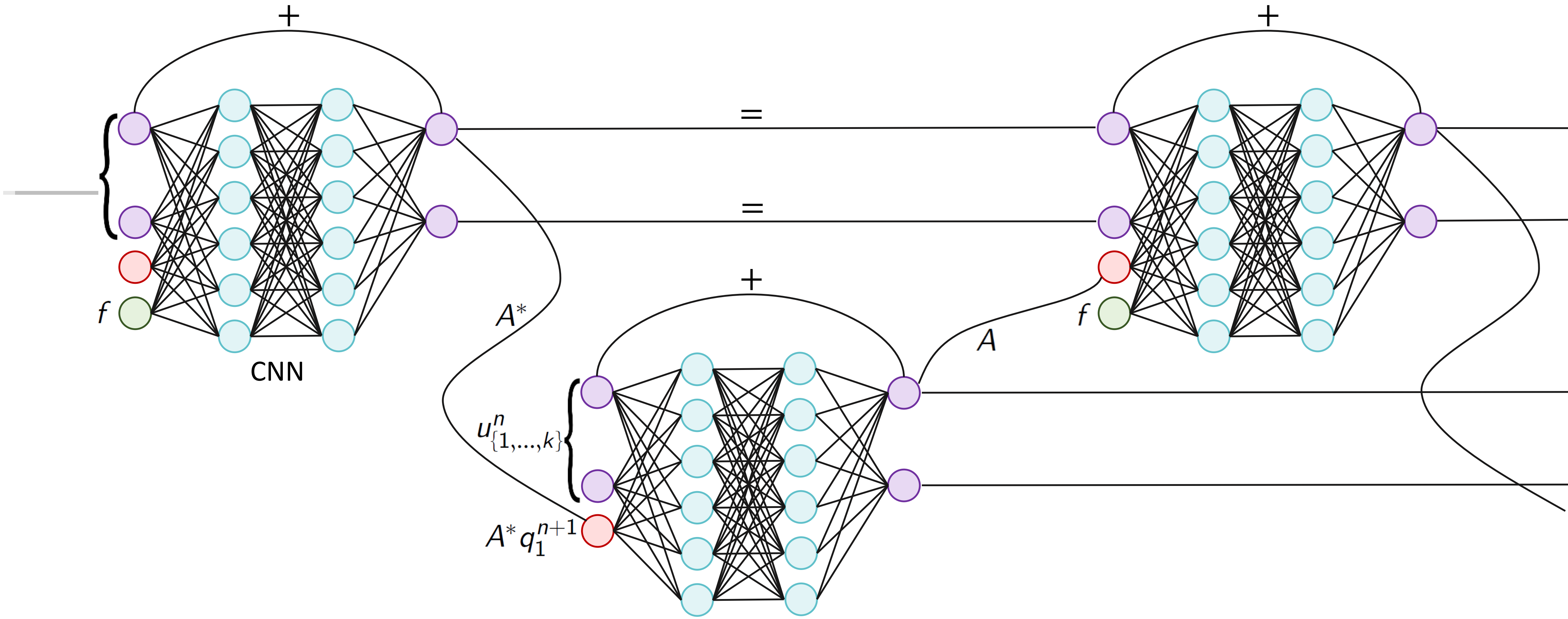
Hauptmann, Lucka, Betcke, Huynh, Cox, Beard, Ourselin, Arridge - Model based learning for accelerated, limited-view 3D photoacoustic tomography (2017)

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

$$q^{n+1}_{\{1,\dots,k\}} = q^n_{\{1,\dots,k\}} + \Gamma_{\Theta_n}\left(\boxed{q^n_{\{1,\dots,k\}}}, \boxed{Au^n_1}, \boxed{f}\right),$$

$$u^{n+1}_{\{1,\dots,k\}} = u^n_{\{1,\dots,k\}} + \Lambda_{\Theta_n}\left(\boxed{u^n_{\{1,\dots,k\}}}, \boxed{A^* q^{n+1}_1}\right).$$

end for



MODEL CHOICES PAT SIMULATION AND RECONSTRUCTION NETWORK

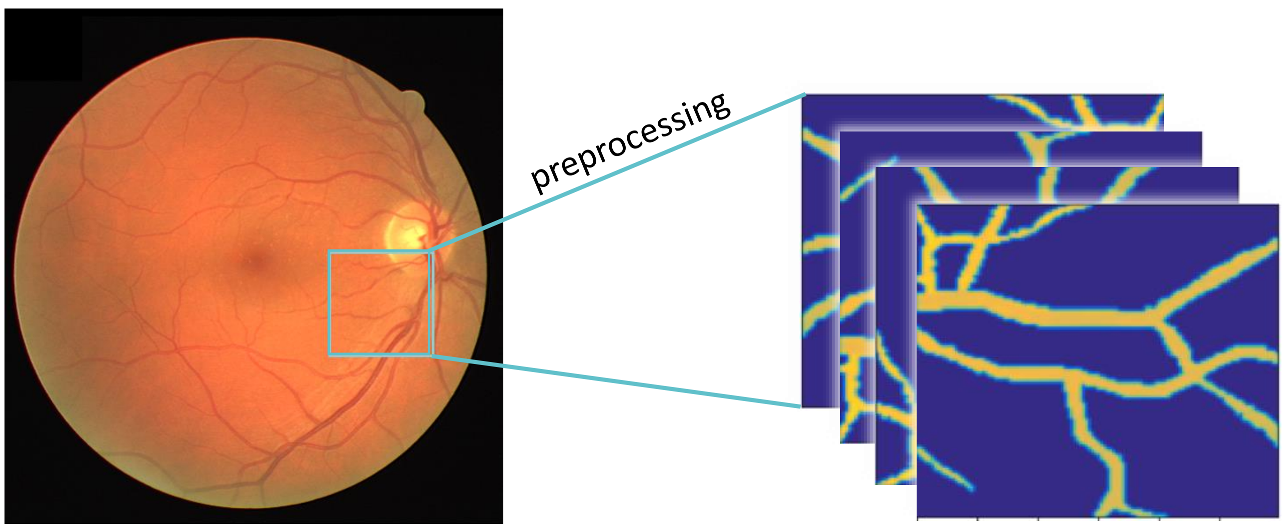
	Training
Resolution	1.5625 mm
Number of pixels	192x192
Number of sensors	32

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

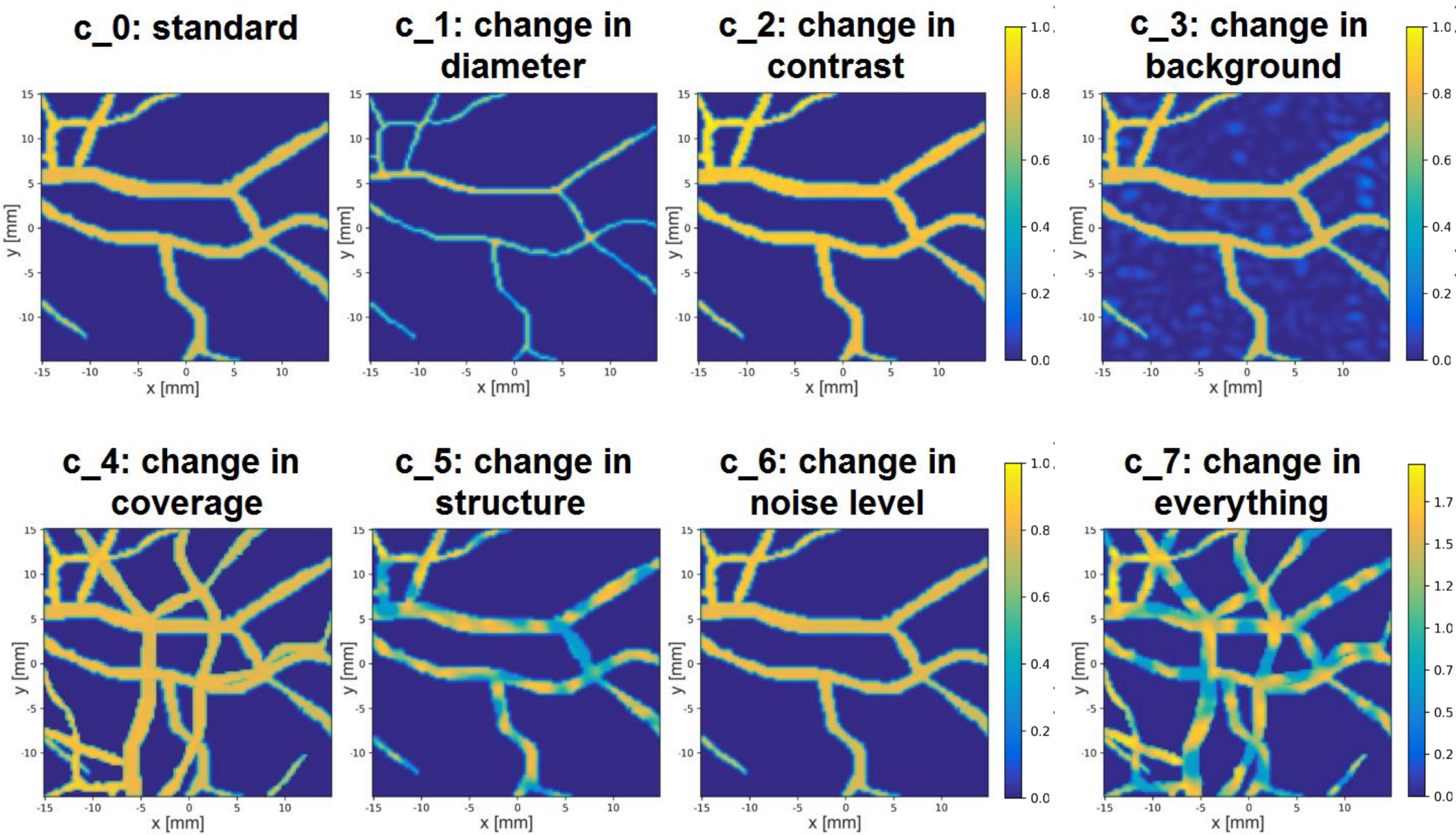


	'large' network	small network
# primal-dual iterations (N)	10	5
# primal/dual channels (k)	5	2
# hidden layers	2	2
# channels in hidden layers	32	32
activation functions	ReLU	ReLU
filter size convolutions	3x3	3x3

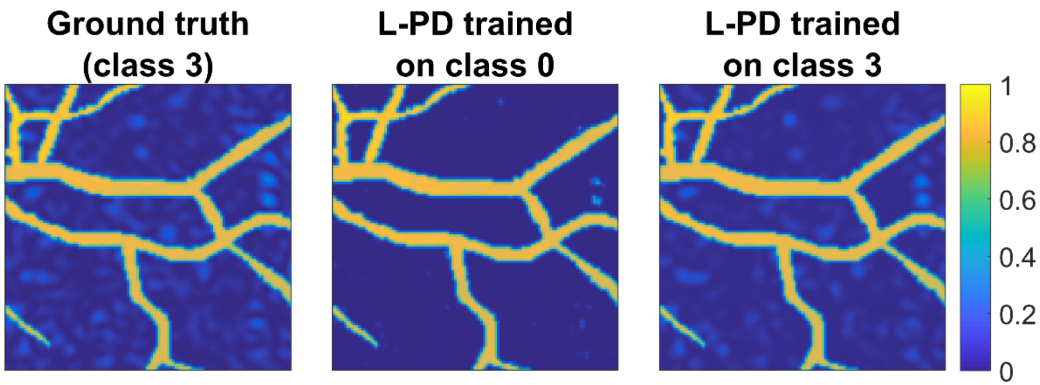
DRIVE dataset: Staal, Abramoff, Niemeijer, Viergever, Ginneken - Ridge based vessel segmentation in color images of the retina (2004)

- 768 training images
- 192 test images
- scaled between 0 and 1

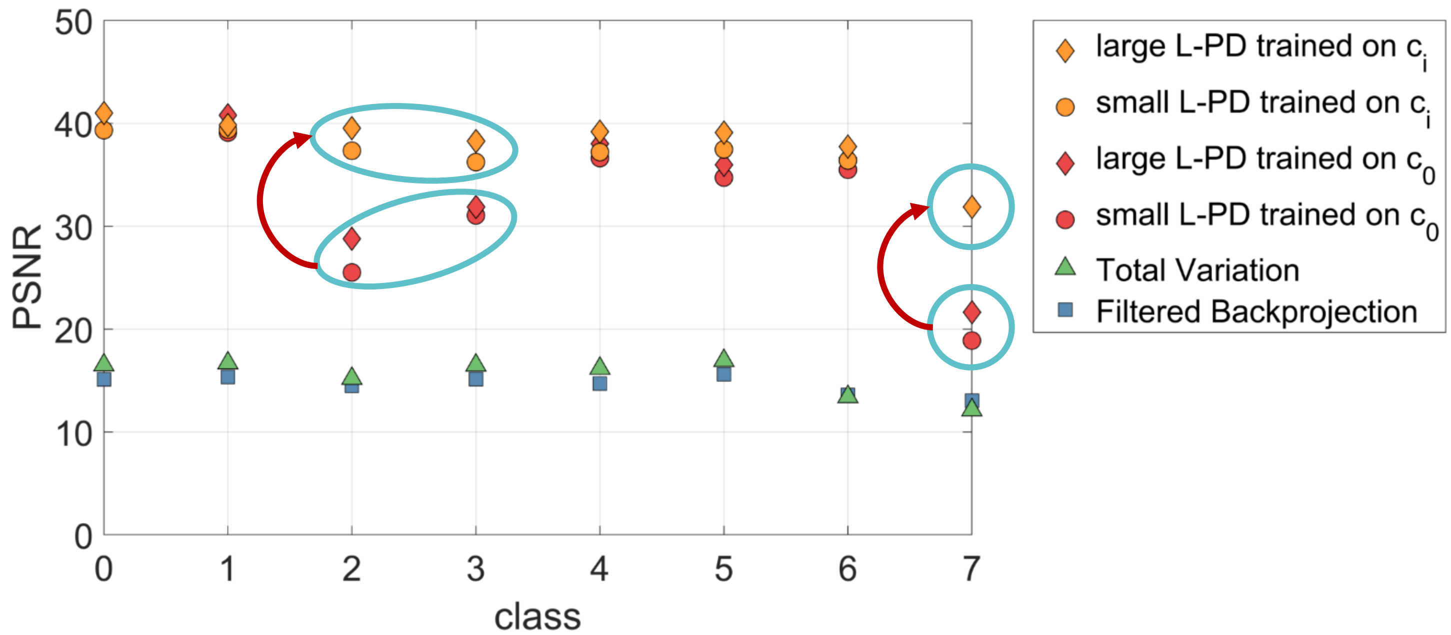
ROBUSTNESS TO IMAGE CHANGES



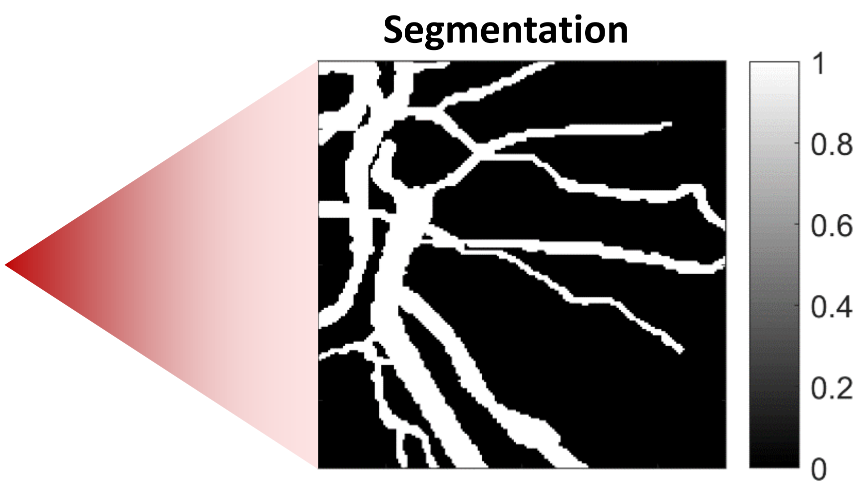
ROBUSTNESS TO IMAGE CHANGES

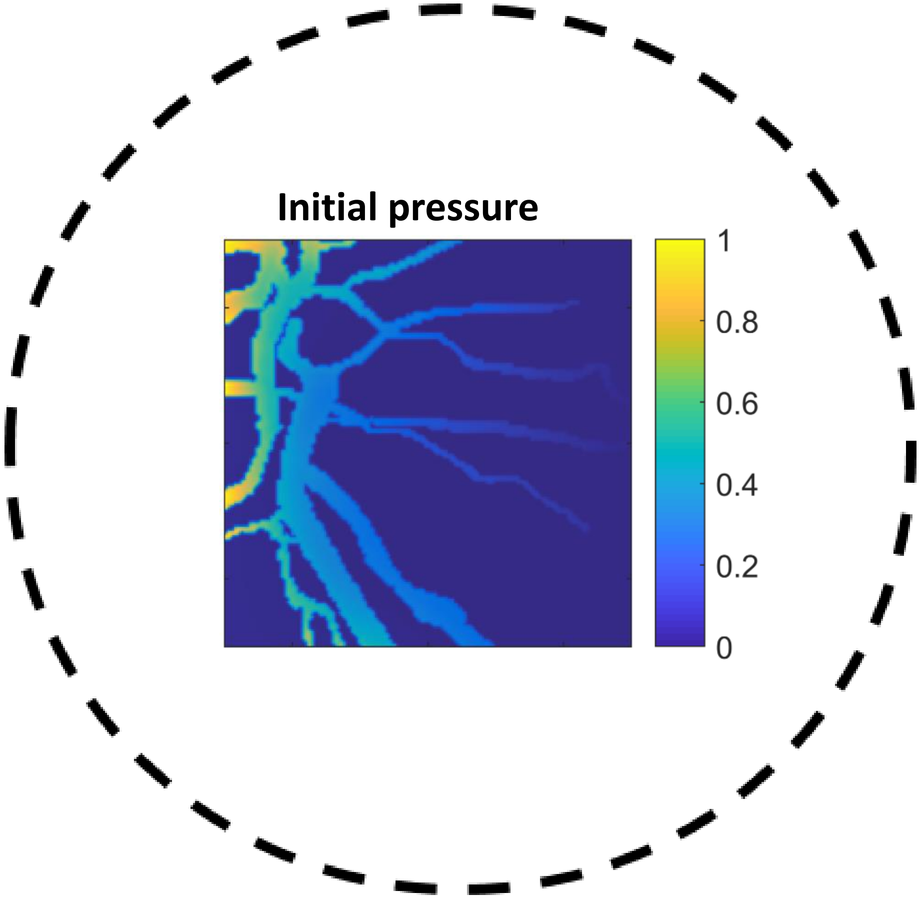


- Strong noise removal and background identification;
- L-PD is robust against many changes in image;
- Training with more variety in data has positive effect on quality.



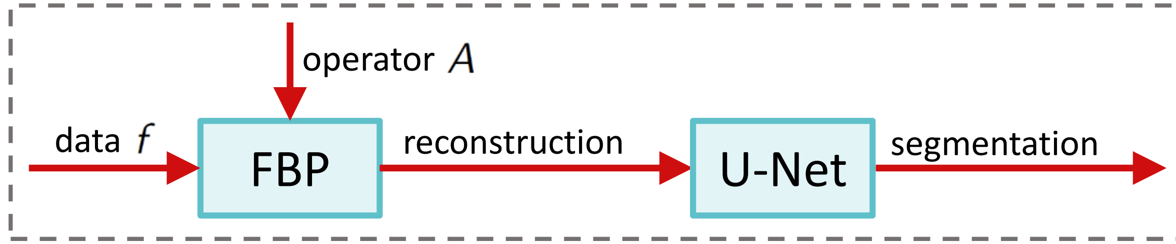
INCLUDING HIGHER-LEVEL TASK: SEGMENTATION



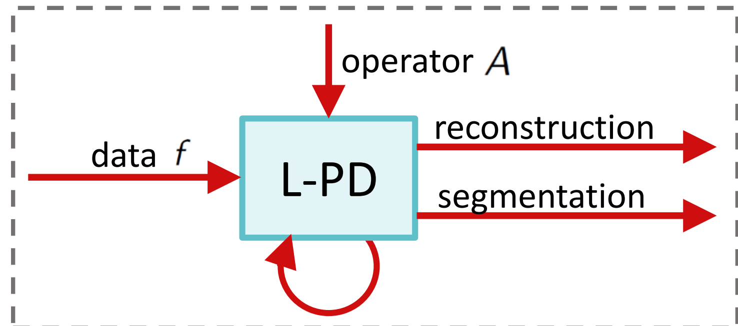


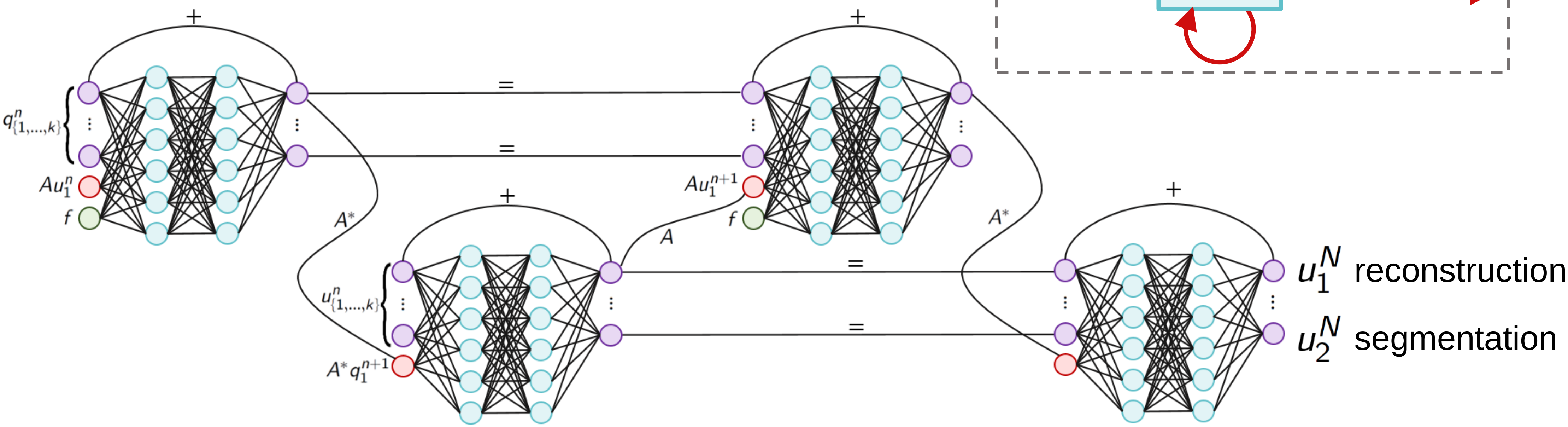
INCLUDING HIGHER-LEVEL TASK: SEGMENTATION

- Learned **post-processing**: U-Net



- **joint reconstruction and segmentation** with L-PD





INCLUDING HIGHER-LEVEL TASK: SEGMENTATION

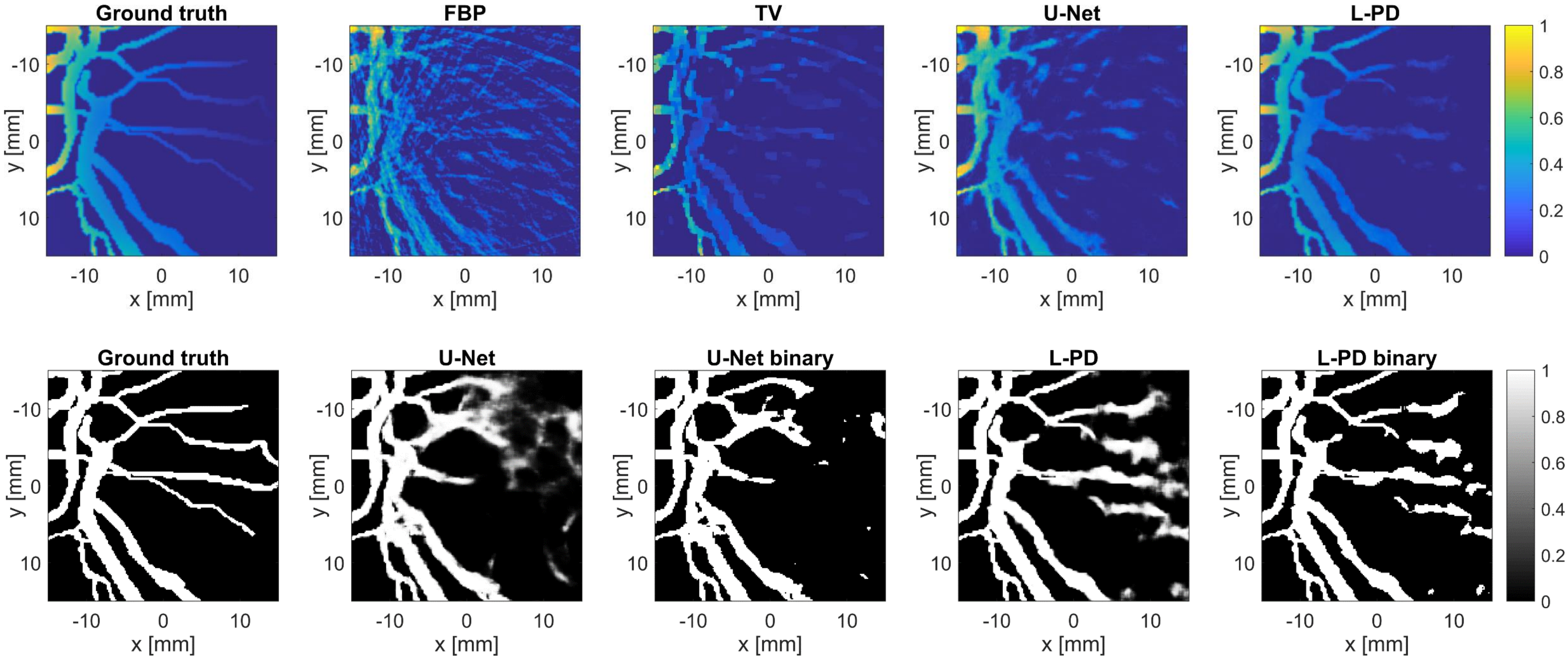
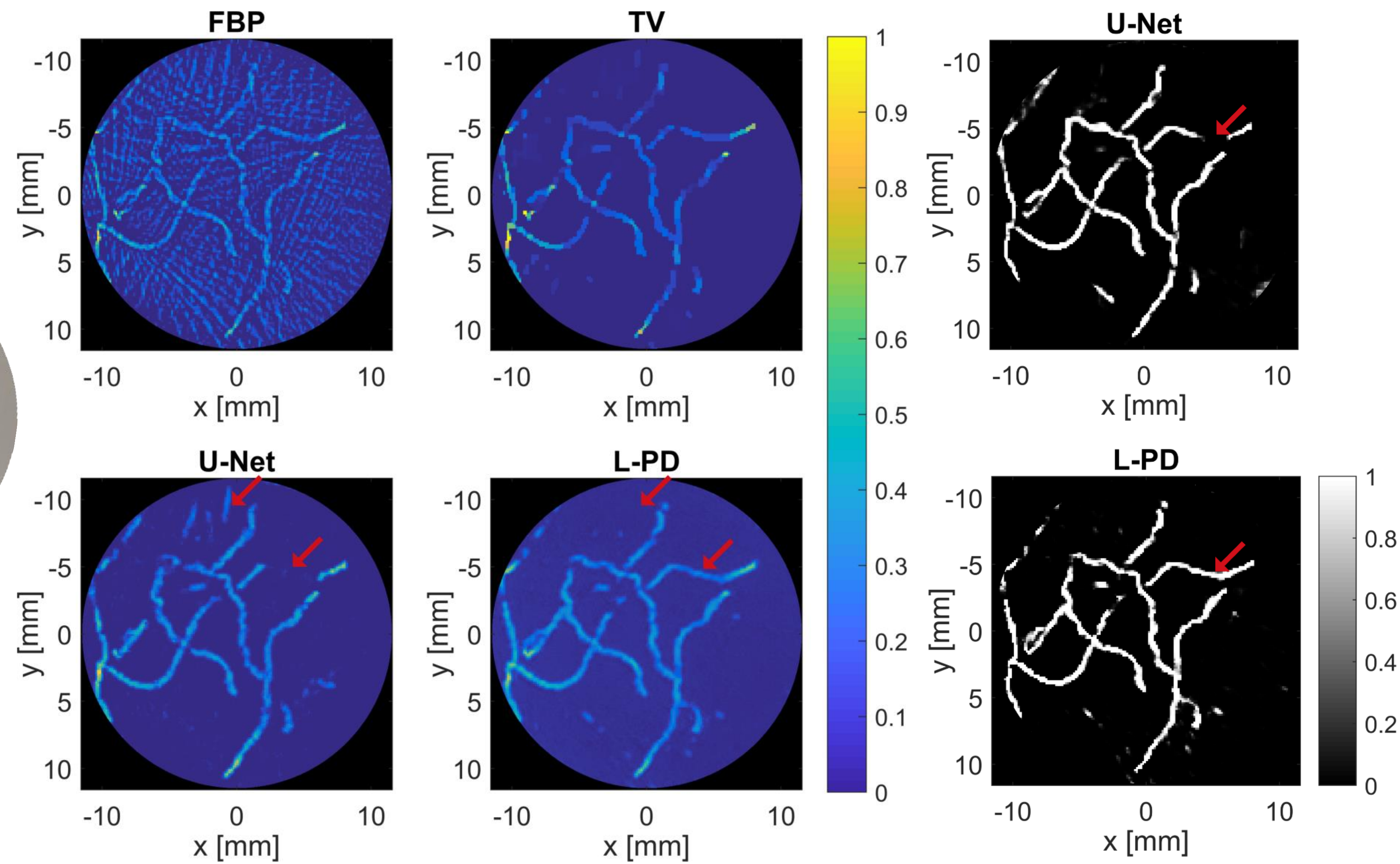
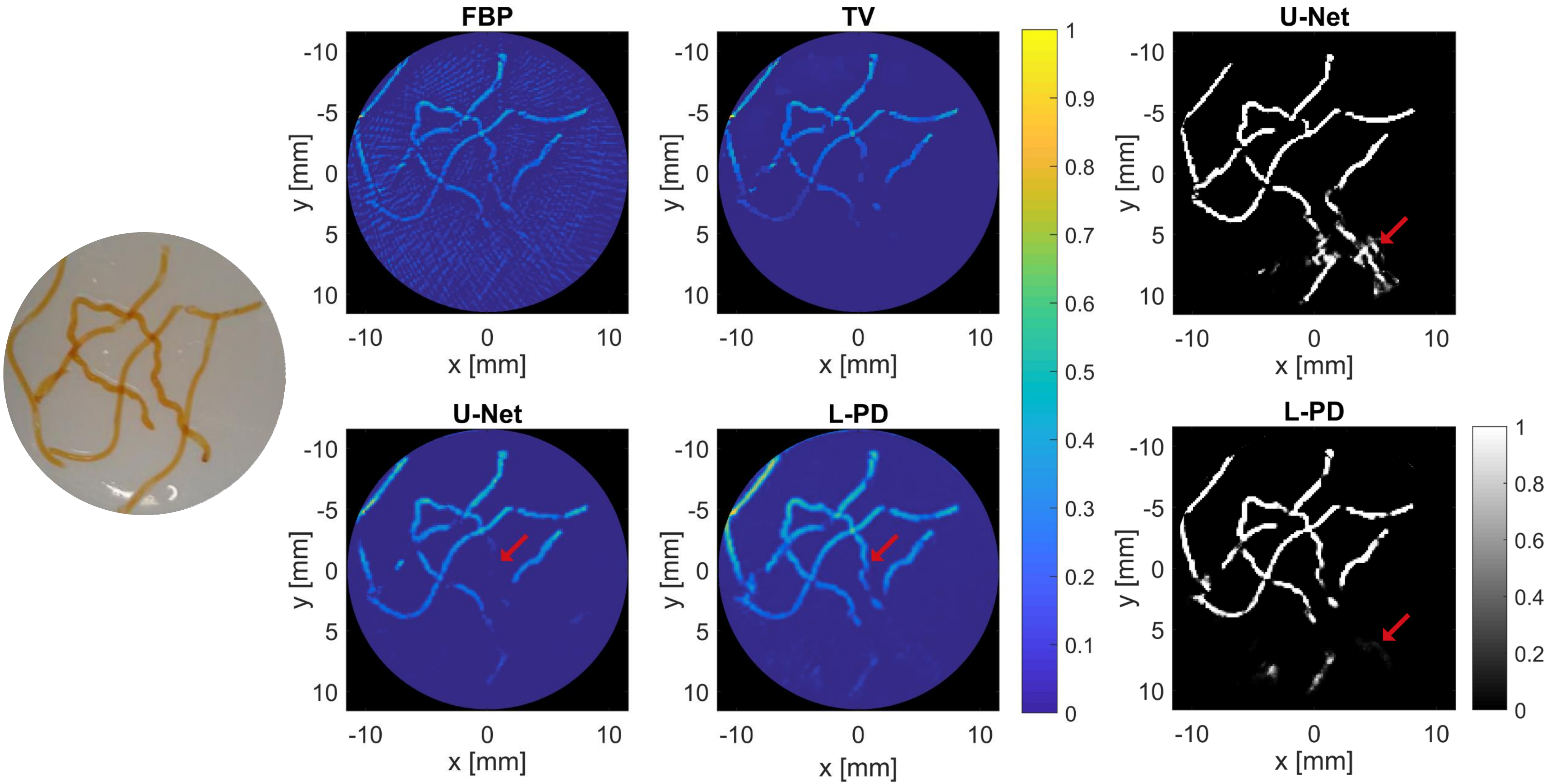


Figure: Reconstructions and segmentations using a 32 detector setting.

EXPERIMENTAL RESULTS: ONE-SIDED SAMPLING



EXPERIMENTAL RESULTS: UNIFORM SAMPLING



ROBUSTNESS TO SYSTEM CHANGES

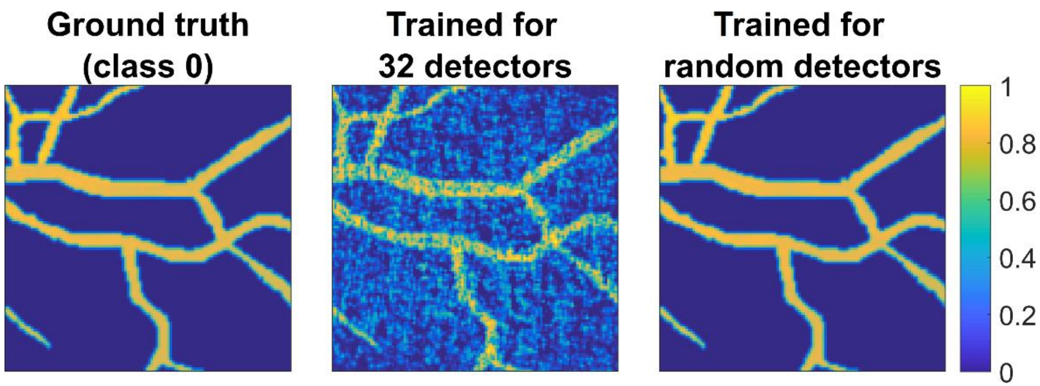


Figure: Reconstruction from data of 64 detectors.

How to achieve generalisability towards imaging operator uncertainty?
Latent structures?!

CONVERGENCE OF PARTIALLY LEARNED METHODS

Several recent papers give convergence proofs for:

- methods that use explicit (Tikhonov-like) regularisation in the form of a neural network.[4]
- methods with a proximal structure.[5]
- methods where the learned part only has influence on the null-space.[6]

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

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

[6] Schwab, Antholzer, Haltmeier – Deep Null Space Learning for Inverse Problems: Convergence Analysis and Rates (2018)

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

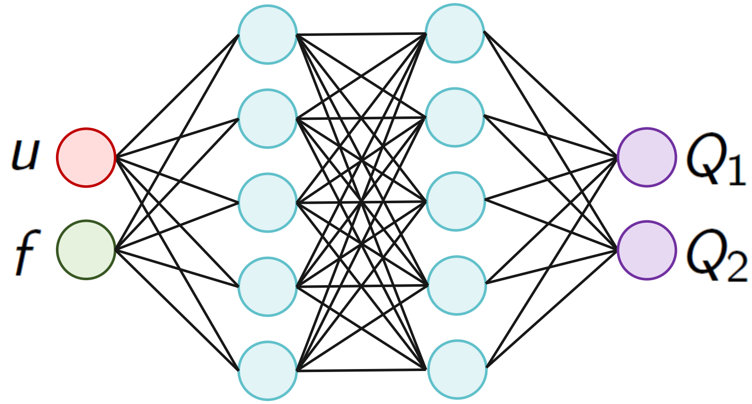
LEARNED (UNROLLED) GRADIENT DESCENT

Goal: learn a nonlinear function (functional) such that its minimiser is our desired reconstruction

$$u^* = \operatorname*{argmin}_u G(u)$$
$$:= \operatorname*{argmin}_u \|Q_1(u,f)\|_1 + \|Q_2(u,f)\|_2^2,$$

where $(Q_1(u), Q_2(u)) = Q(u) := \sigma(H_m\sigma(H_{m-1} \ldots \sigma(H_1(u,f))))$,

and H_i represents a convolutional layer.

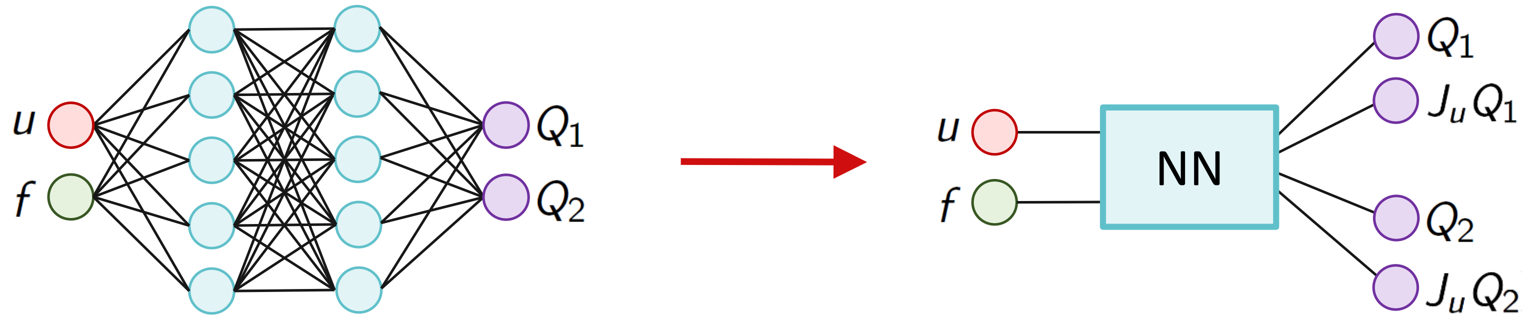


LEARNED (UNROLLED) GRADIENT DESCENT

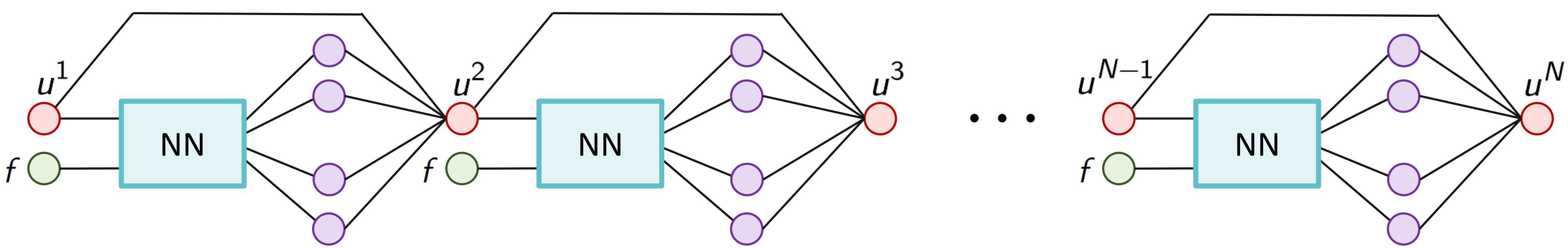
- $G(u, f) = \|Q_1(u, f)\|_1 + \frac{1}{2}\|Q_2(u, f)\|_2^2,$
- $\nabla_u G = \text{sign}(Q_1) J_u Q_1 + Q_2 J_u Q_2$
- $$\begin{aligned} u^{n+1} &= u^n - \eta \nabla_u G \\ &= u^n - \eta\,[\text{sign}(Q_1) J_u Q_1 + Q_2 J_u Q_2] \end{aligned}$$
- Train network with $\eta = 1$ or $n \in \{1, \dots, N = 20\}$

Learning of well-known gradient flows possible?
Diffusion, convection-diffusion, thin-film, etc.

LEARNED (UNROLLED) GRADIENT DESCENT



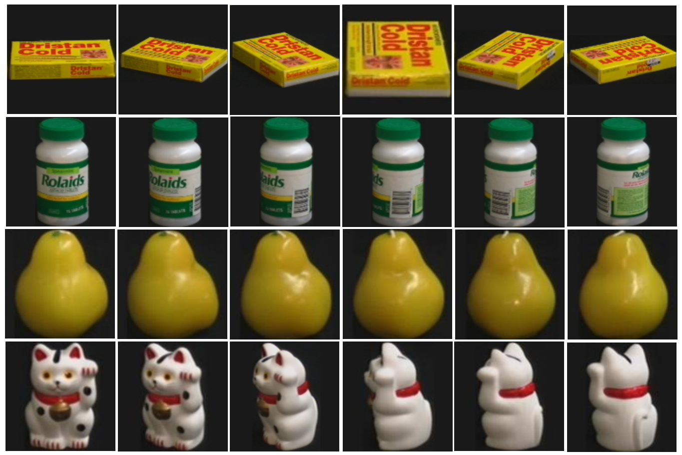
$$
\begin{aligned}
u^{n+1} &= u^n - \eta \nabla_u G \\
&= u^n - \eta \left[\operatorname{sign}(Q_1) J_u Q_1 + Q_2 J_u Q_2\right]
\end{aligned}
$$



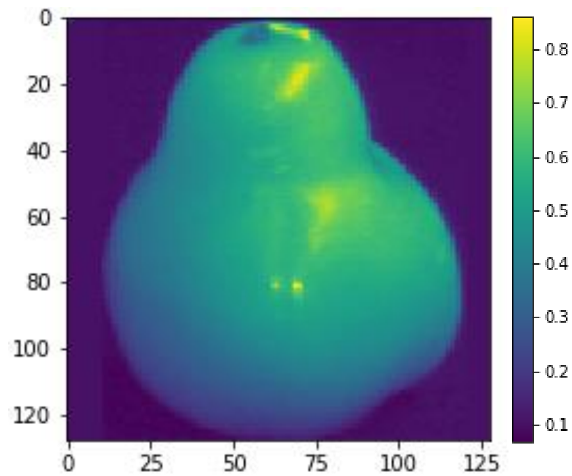
TRAINING DETAILS: COIL DATASET

Columbia Object Image Library

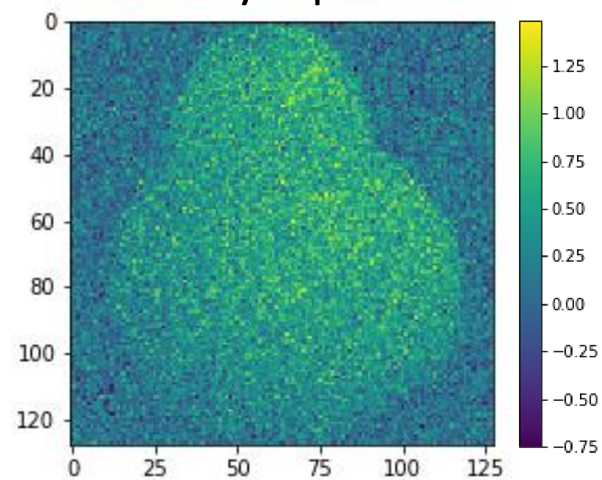
Nene, Nayar, Murase – Columbia Object Image Library (COIL-100) (1996)



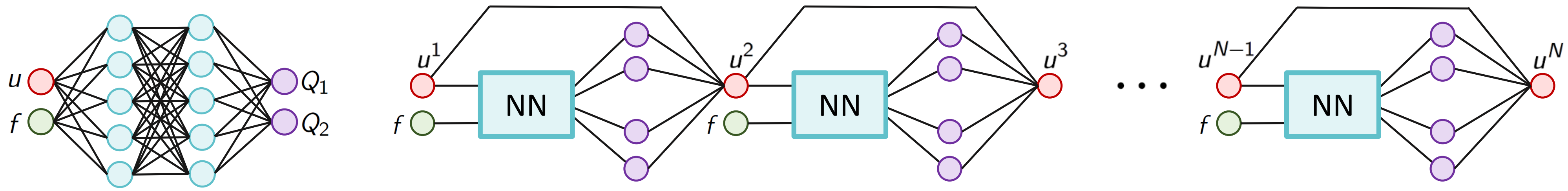
Ground truth



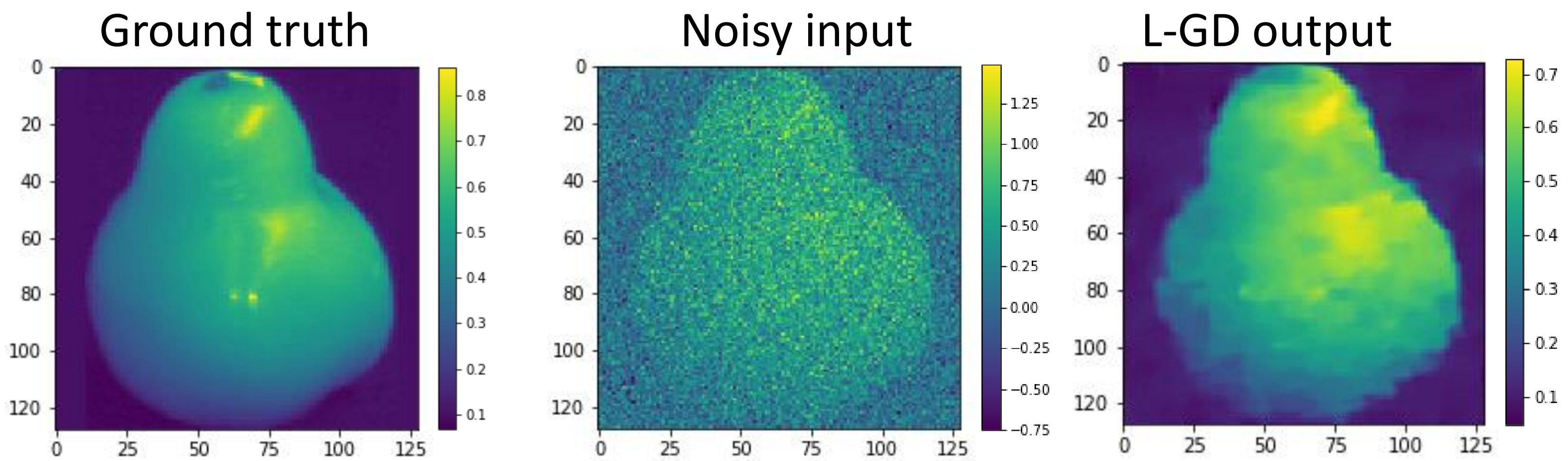
Noisy input



TRAINING PARAMETERS AND RESULT



- #GD-steps = N = 20
- #layers = 3
- #channels = 8
- batch size = 9
- #epochs = 50
- Adam optimiser
  rate $2 \cdot 10^{-2} \rightarrow 10^{-3}$
- kernel size = 3x3



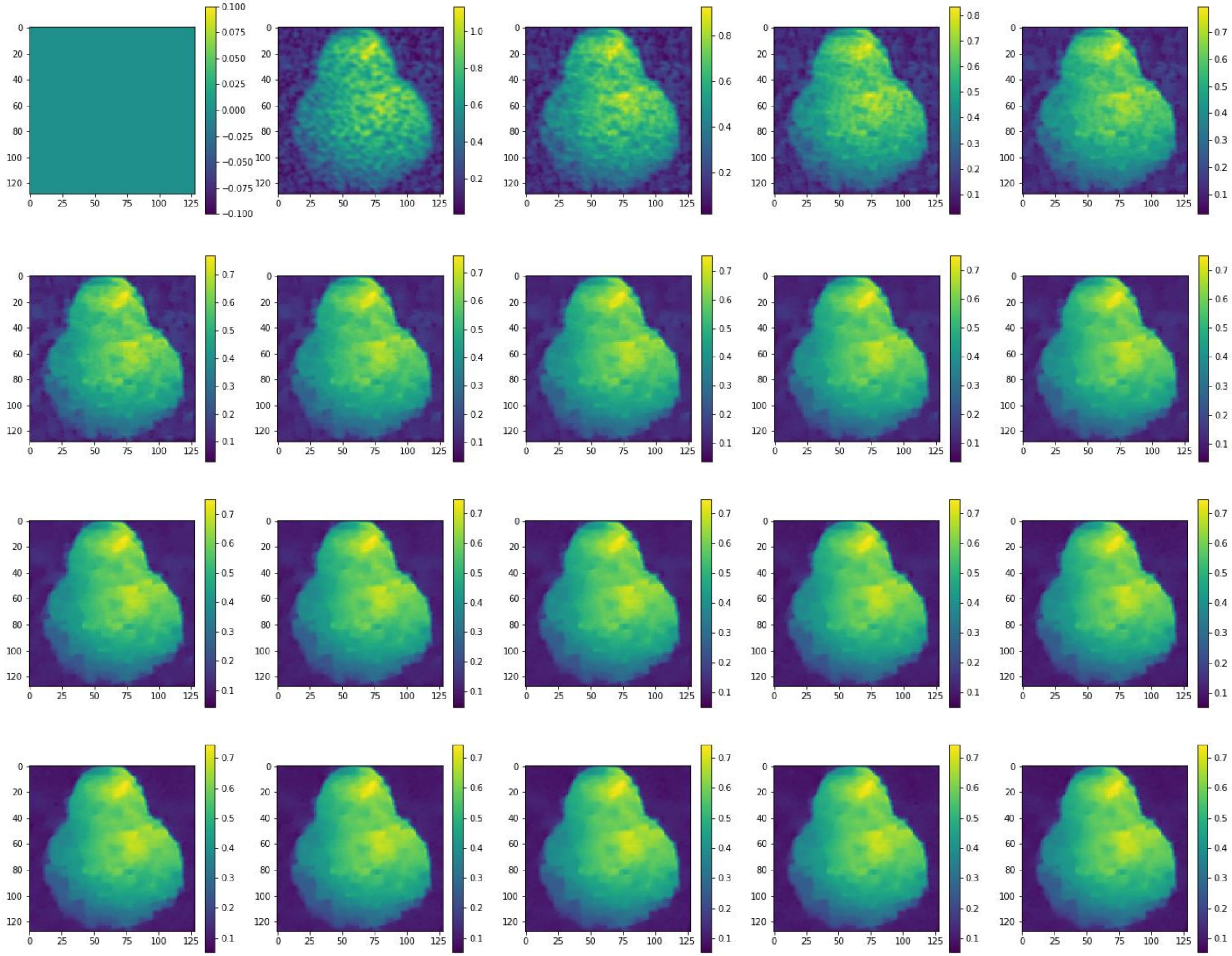
LEARNED (UNROLLED) GRADIENT DESCENT

iterations 1-5

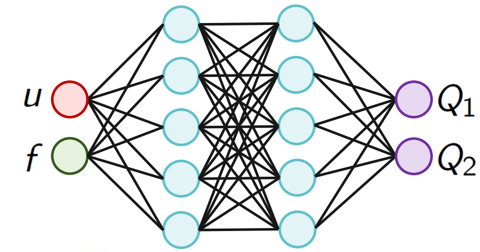
iterations 6-10

iterations 11-15

iterations 16-20



LEARNED (UNROLLED) GRADIENT DESCENT: Q_2

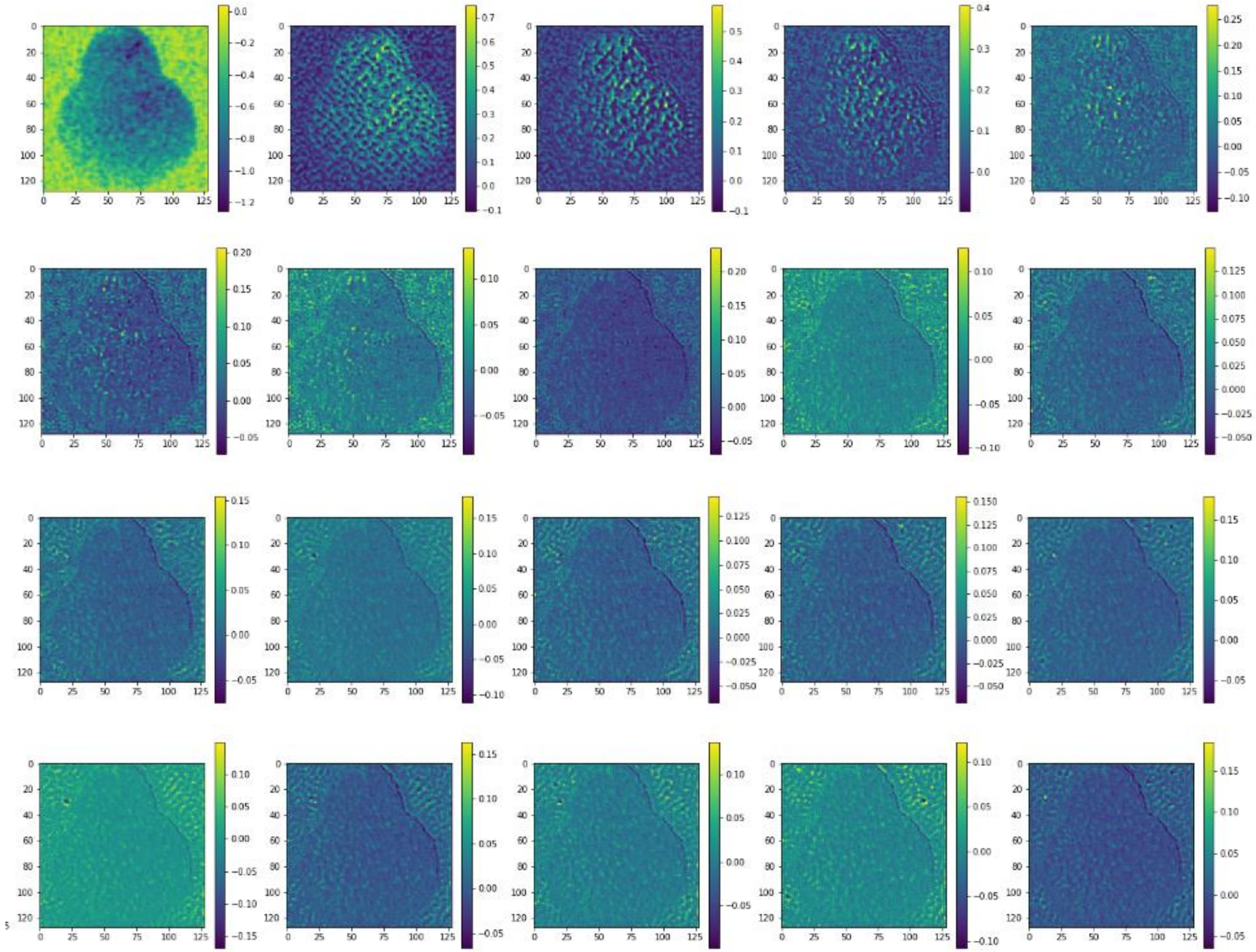


iterations 1-5

iterations 6-10

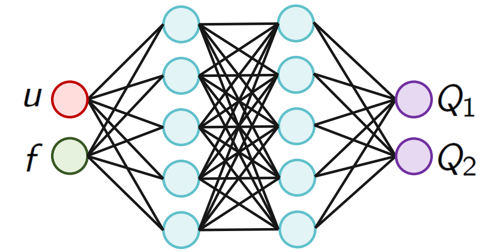
iterations 11-15

iterations 16-20



LEARNED
“DATA FIDELITY”
OPERATOR

LEARNED (UNROLLED) GRADIENT DESCENT: Q_1

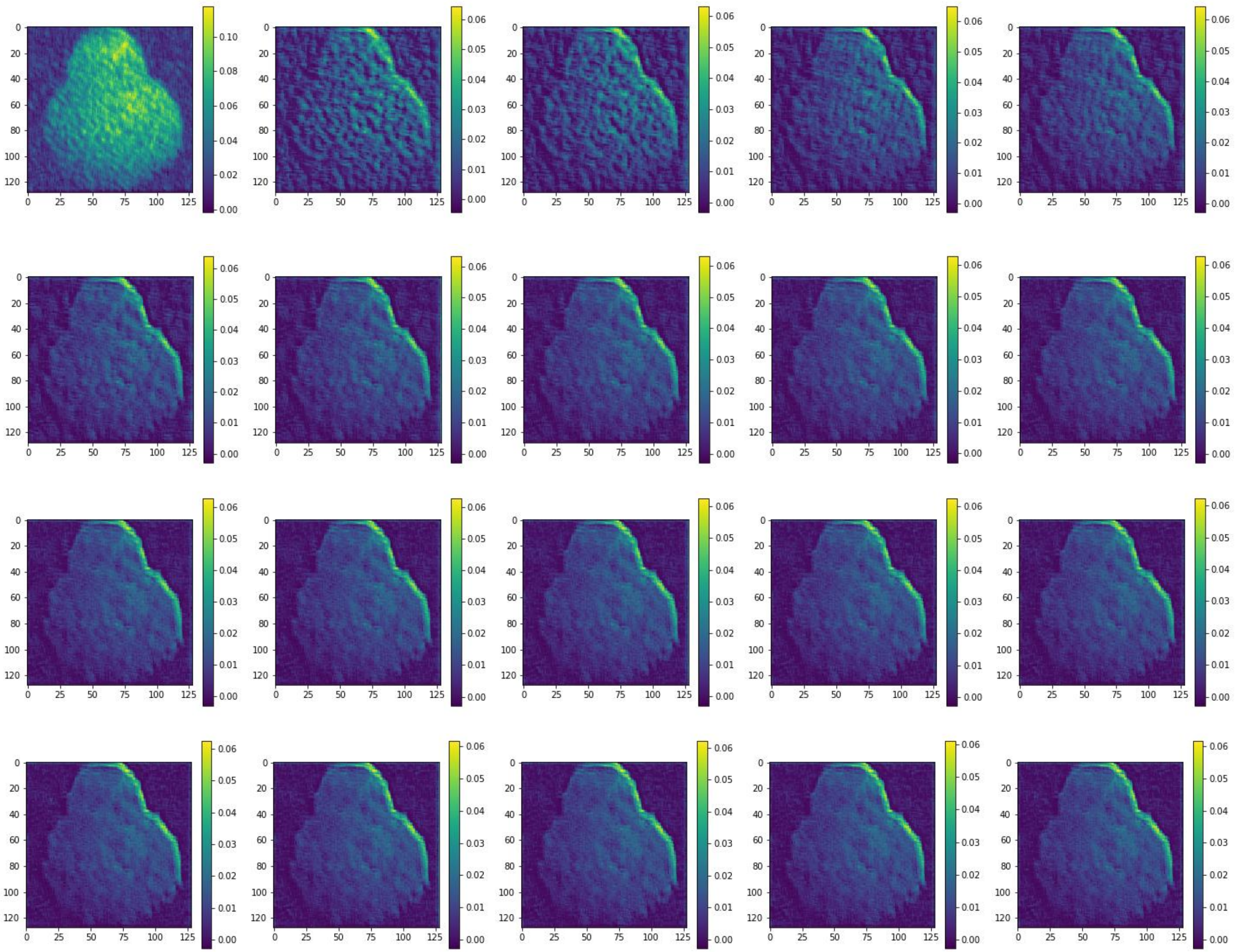


iterations 1-5

iterations 6-10

iterations 11-15

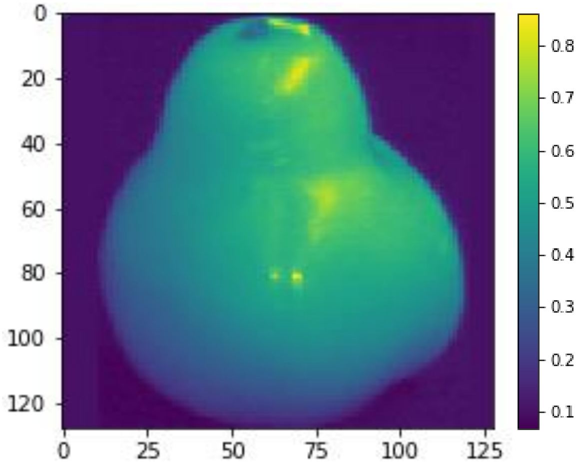
iterations 16-20



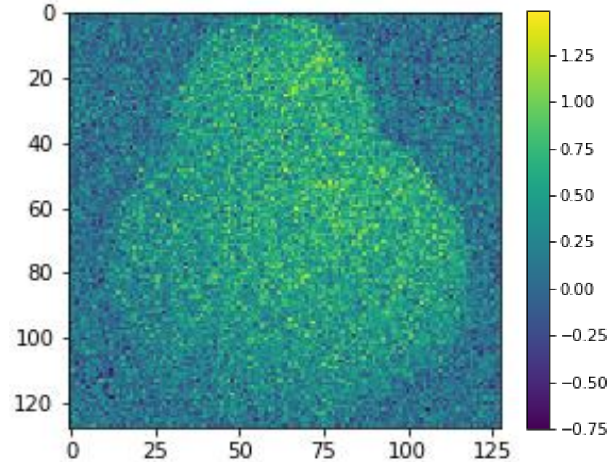
LEARNED "REGULARIZATION" OPERATOR

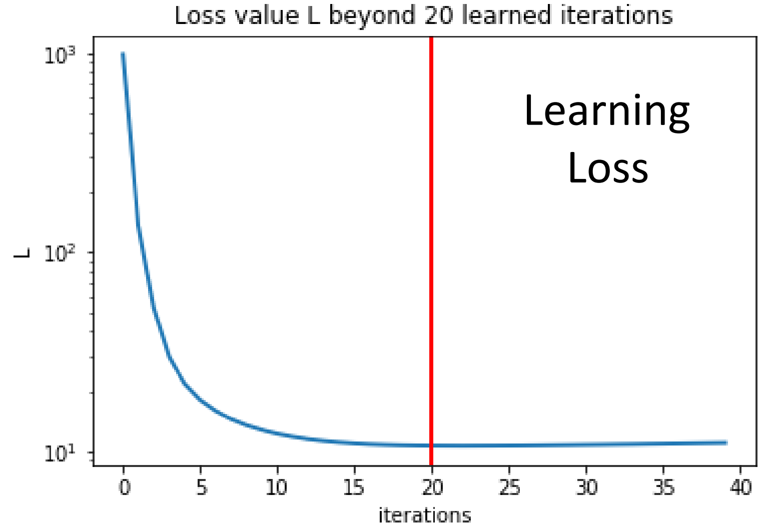
LEARNED (UNROLLED) GRADIENT FLOW

Ground truth

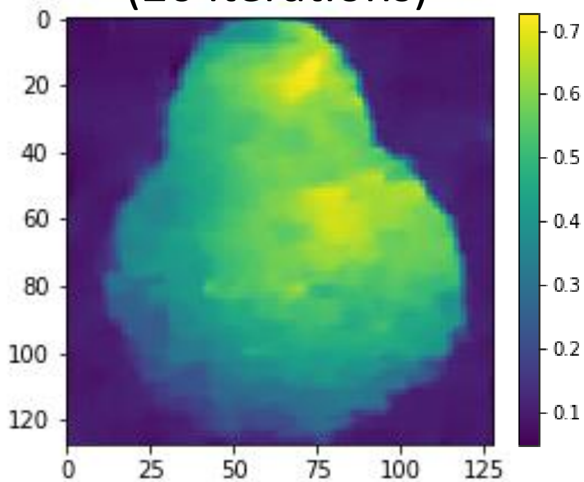


Noisy input

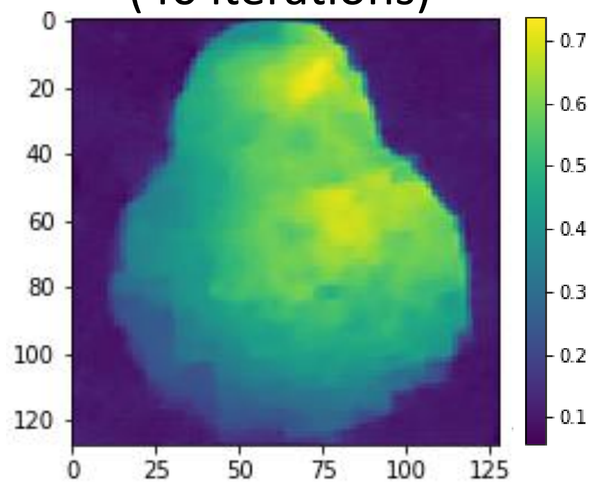


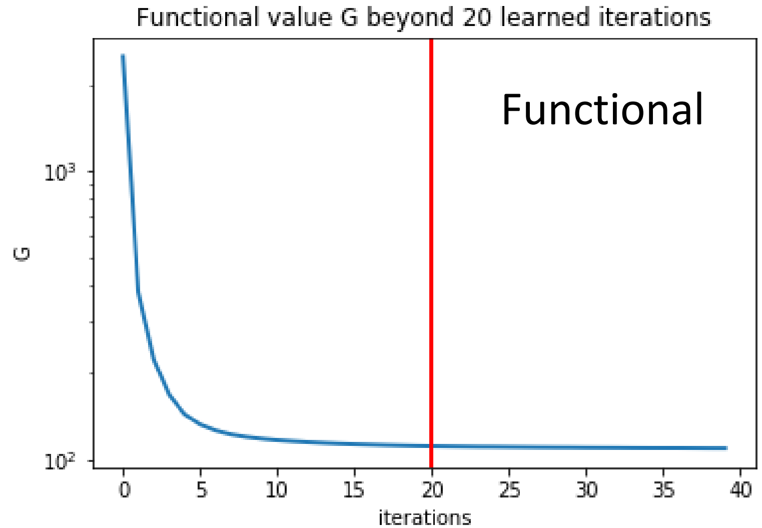


L-GD output (20 iterations)



L-GD output (40 iterations)





LEARNING GRADIENT FLOWS, VARIATIONAL NETWORKS (PHYSICS-INFORMED)

$\partial_t \rho = \Delta \rho$	heat equation
$\partial_t \rho = \Delta \rho + \operatorname{div} \rho \nabla V + \operatorname{div} \rho \nabla W * \rho$	convection-diffusion
$\partial_t \rho = \Delta \rho^m$	porous medium equation
$\partial_t \rho = \operatorname{div} \rho \nabla \Delta \rho$	thin-film equation
$\partial_t \rho = -\operatorname{div} \rho \nabla \left[\rho^{\alpha-1} \Delta \rho^\alpha\right]$	DLSS equation
$\partial_t \rho = -\operatorname{div} D(\rho) \nabla \left[\Delta \rho + f(\rho)\right]$	Cahn-Hilliard equation
$\partial_t \rho = -\operatorname{div} \rho u, \ \operatorname{div} u = 0, \ u = f(\rho)\left[\nabla p + \rho e_z\right]$	two-phase porous-media flow

Chen, Pock – Trainable Nonlinear Reaction Diffusion (2015)

Lusch et al - Deep learning for universal linear embeddings of nonlinear dynamics (2017)

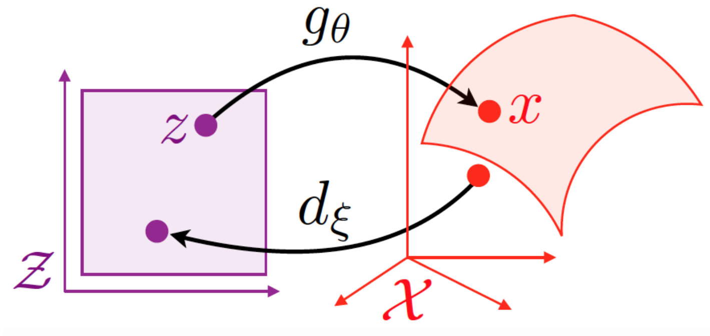
Kobler et al – Variational Networks: Connecting Variational Methods and Deep Learning (2017)

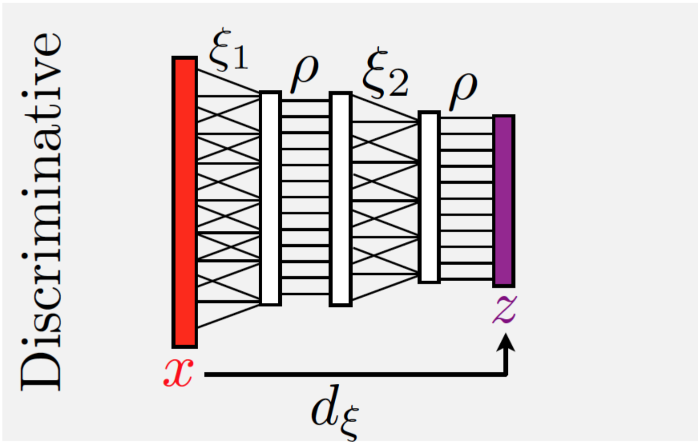
Yang et al - Physics-informed deep generative models (2018)

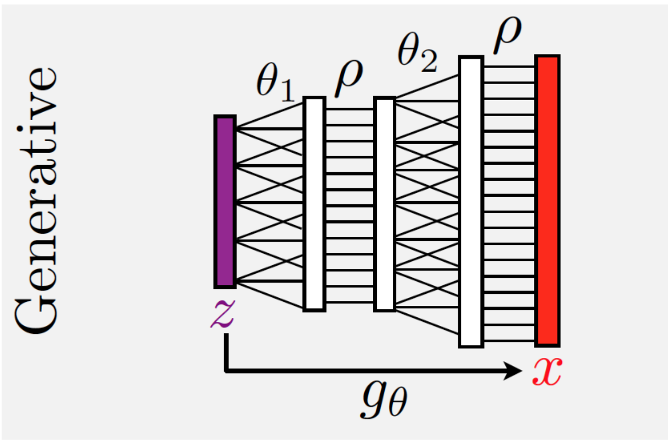
DEEP DISCRIMINATIVE VS GENERATIVE MODELS

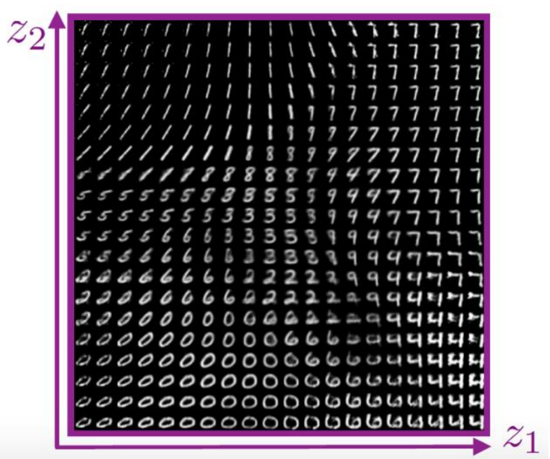
Deep networks:

$$d_\xi(x) = \rho(\xi_K(\dots\rho(\xi_2(\rho(\xi_1(x)\dots)$$
$$g_\theta(z) = \rho(\theta_K(\dots\rho(\theta_2(\rho(\theta_1(z)\dots)$$









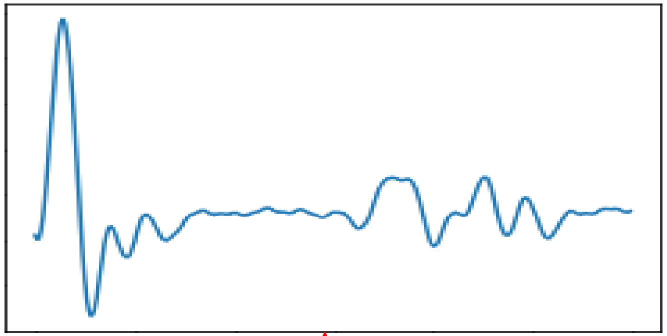
Density fitting	Optimal transport	Auto-encoders
$g_\theta(\{z_i\}_i) \approx \{x_i\}_i$	map d_ξ	$g_\theta(d_\xi(x_i)) \approx x_i$

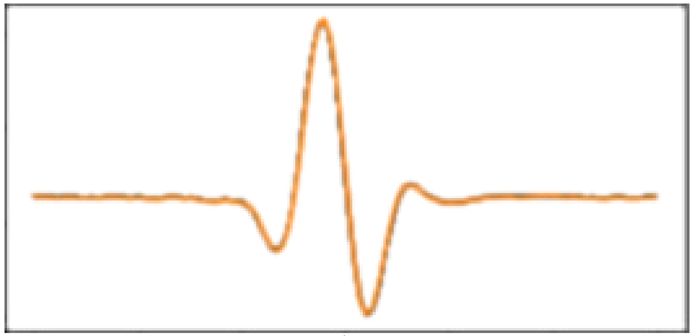
Genevay, Peyre, Cuturi – GAN and VAE from an Optimal Transport Point of View (2017)

Mescheder, Nowozin, Geiger – Adversarial Variational Bayes: Unifying VAEs and GANs (2017)

(BLIND) DECONVOLUTION PROBLEM IN PHOTOACOUSTICS

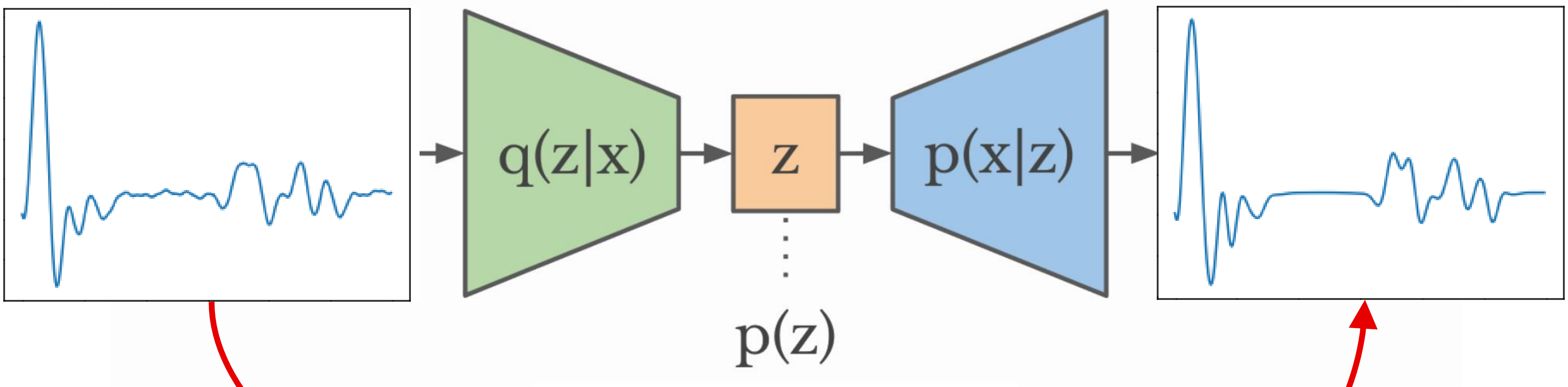
LEARNING UNCERTAINTY - CALIBRATION IN FORWARD MODEL

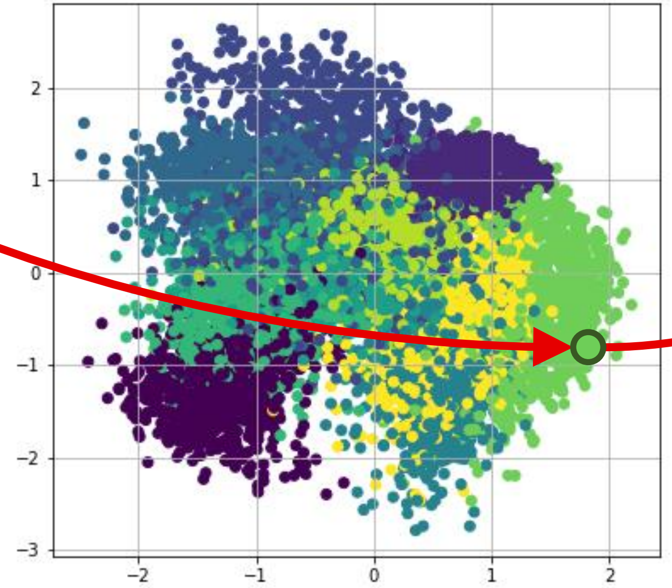




$$\tilde{p}(x,t) = \left(\frac{1}{t} \iint_{|x-\tilde{x}|=ct} u(\tilde{x})\, \mathrm{d}\tilde{x} \right) *_t p_{\mathrm{cal}}(t),$$

LEARNED DATA-SPACE WITH VARIATIONAL AUTOENCODER

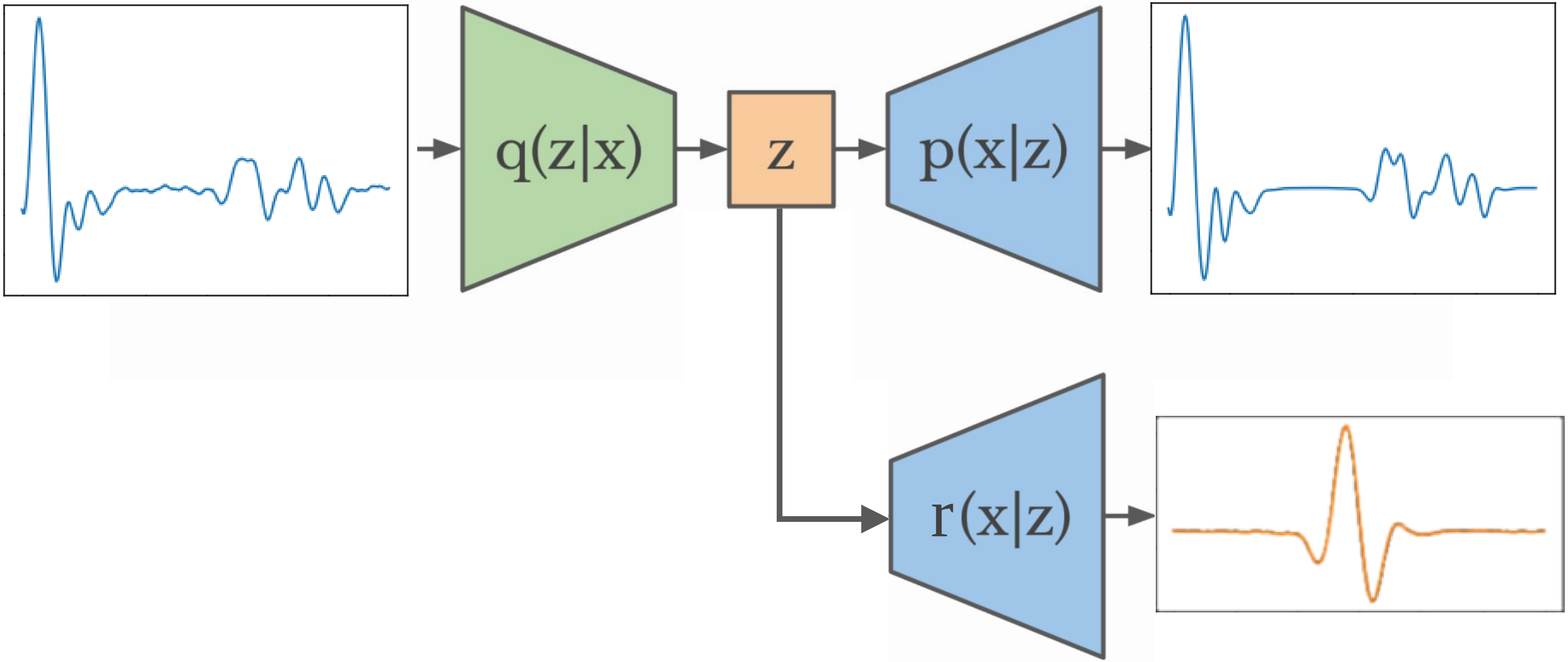




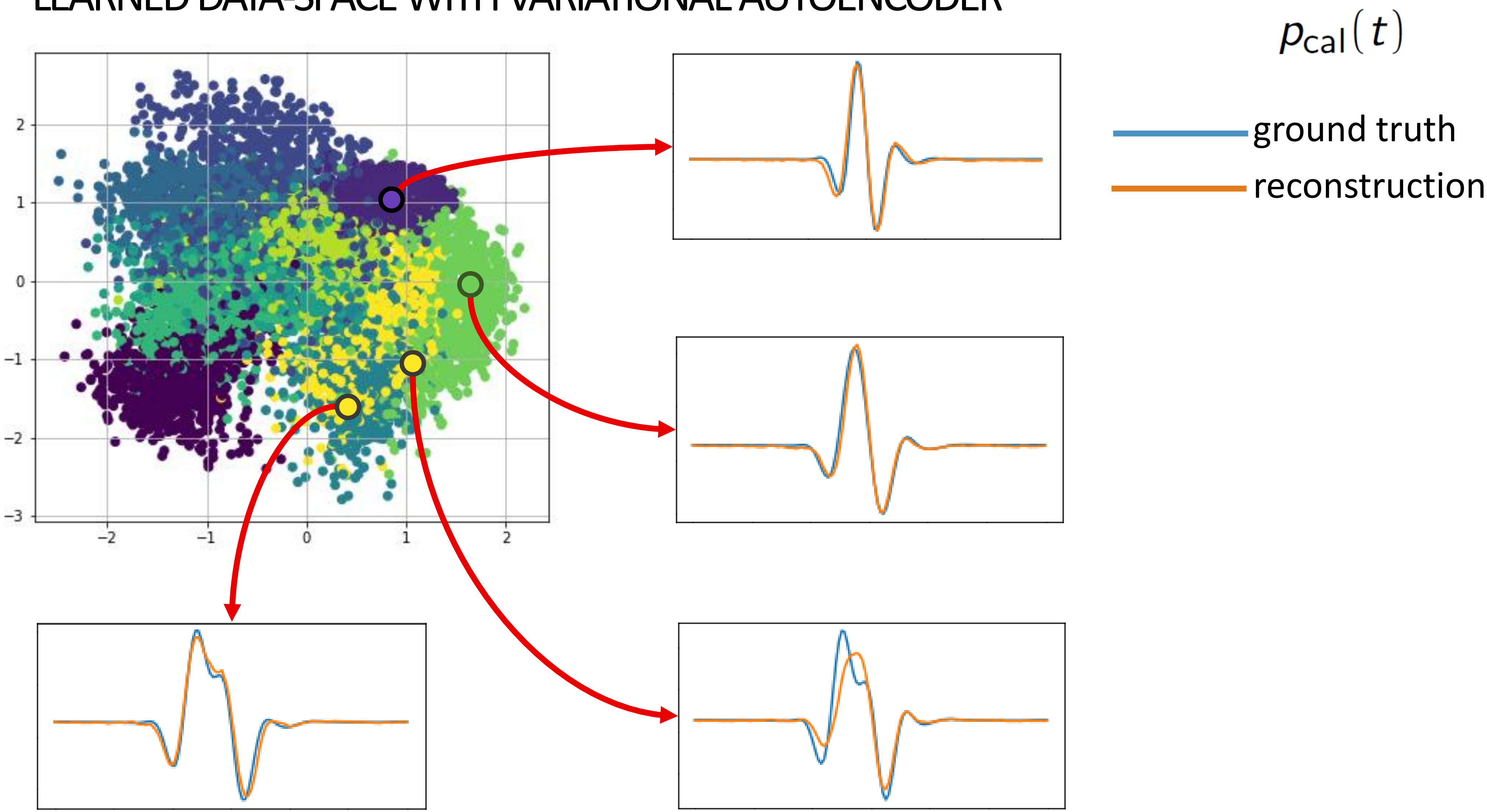
Kingma, Welling – Auto-Encoding Variational Bayes (2013)

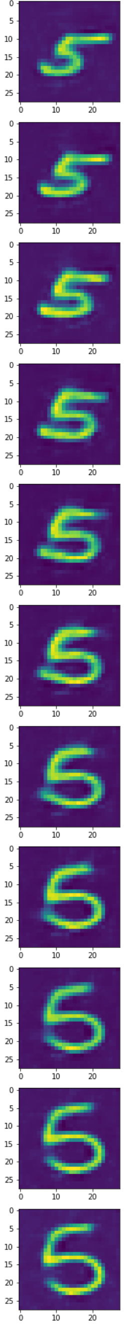
Dai, Wipf - Diagnosing and Enhancing VAE Models (2019)

LEARNED DATA-SPACE WITH VARIATIONAL AUTOENCODER

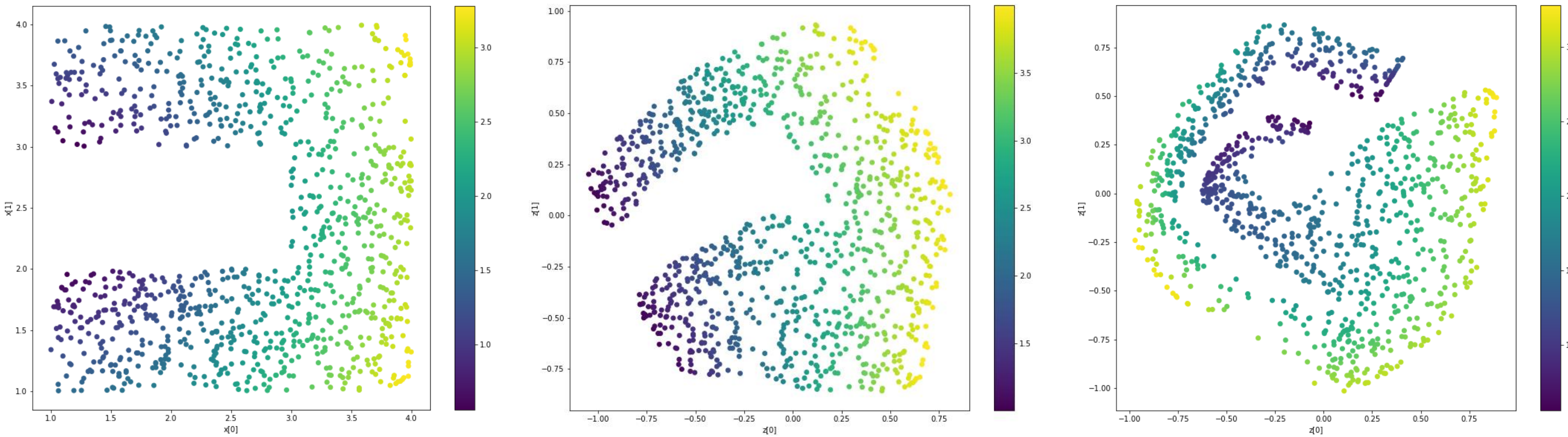


LEARNED DATA-SPACE WITH VARIATIONAL AUTOENCODER

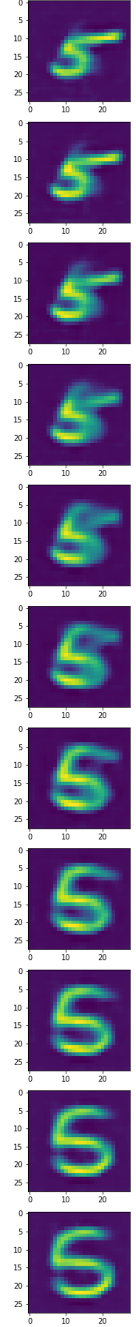




LIMITATION: GEOMETRIC VAE LATENT SPACE REGULARIZATION



Regularization of latent variables z



CONCLUSIONS

- **Nonlinear eigenvalue problems** offer spectral decompositions similar to Autoencoders

- **Robustness**. Deep learning improves PAT reconstruction quality particularly in uncertain cases.
- **Functional learning**. Mathematical theory on stability and convergence not available. Construct learned methods that learn the functional to be minimised via a gradient flow.

- **Generative networks**. Generative networks can be successfully used for learning latent variables in inverse problems. VAEs show geometric limitations for latent space interpolation.

Thanks for your attention

Zeune et al - Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis (2017)

Zeune et al - Deep learning for tumor cell classification (2019)

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

Boink, Manohar, Brune - A partially learned algorithm for joint photoacoustic reconstruction and segmentation (2019)