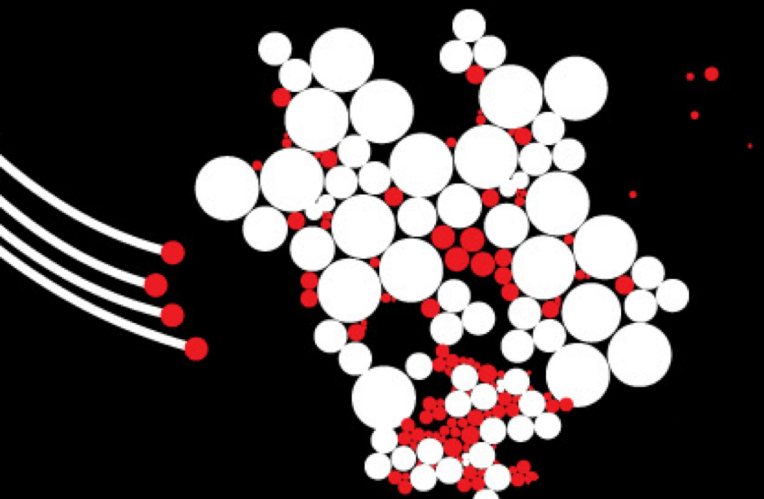
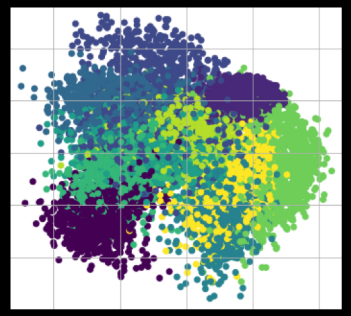
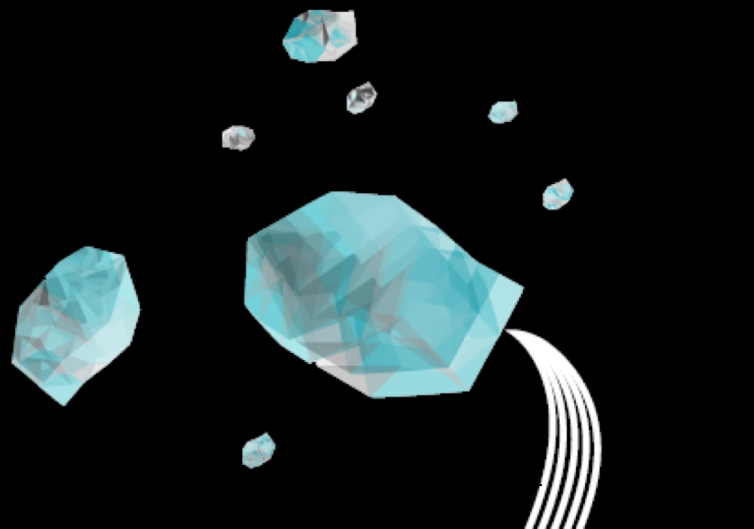
UNIVERSITY OF TWENTE.

Deep Inversion - Latent Space Analysis of Autoencoders for Inverse Problems

Yoeri Boink*†, Srirang Manohar†, Stephan van Gils*, *Christoph Brune**
Applied Mathematics*
Biomedical Photonic Imaging†







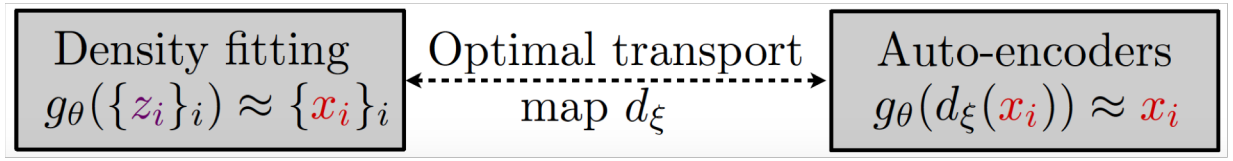
OUTLINE

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

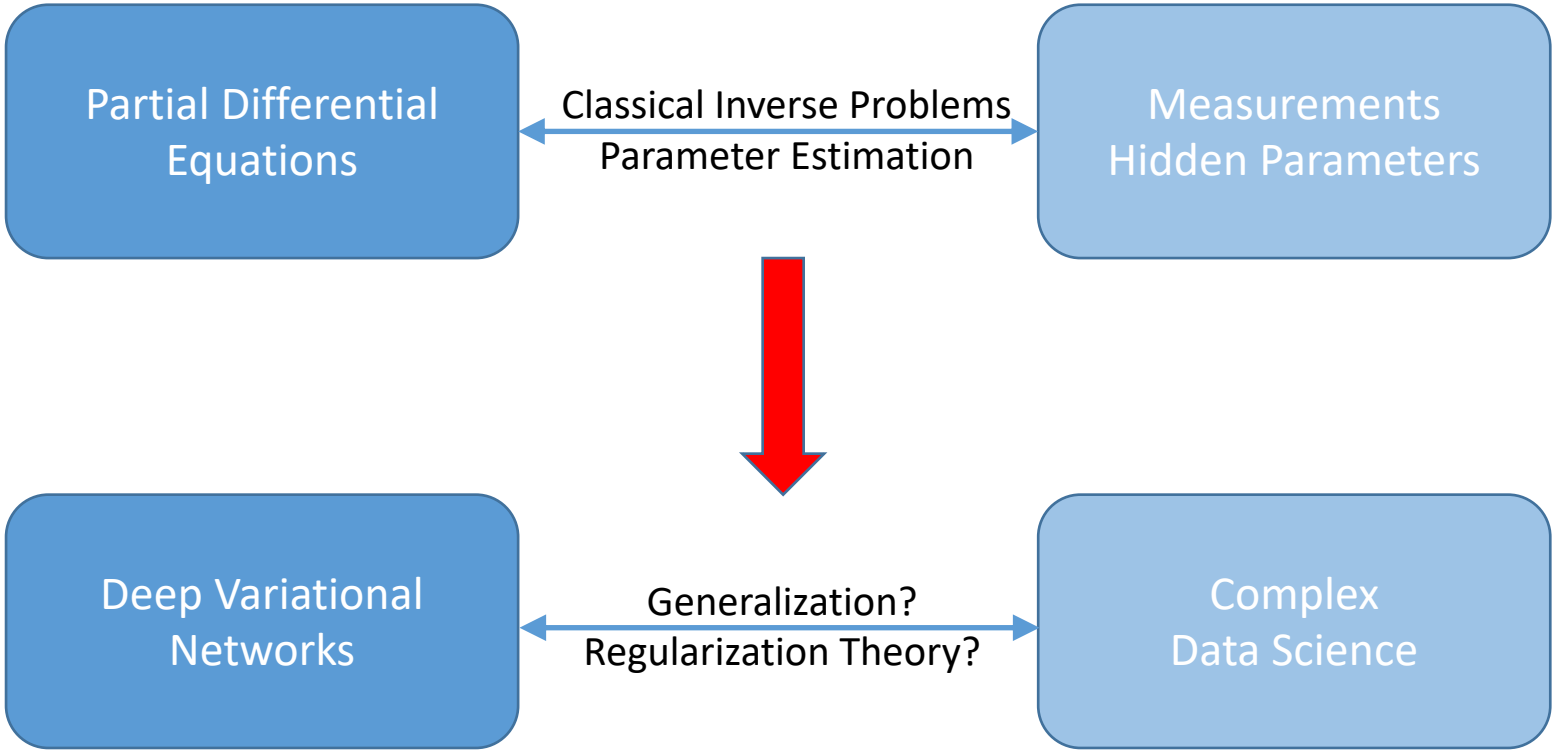
2. **Functional learning** with an unrolled gradient descent (GD) scheme

   → guaranteed convergence and stability!

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

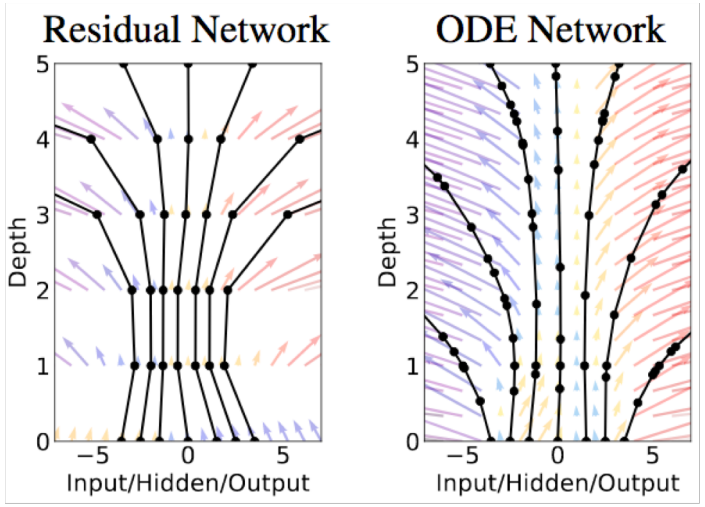


VARIATIONAL METHODS AND DEEP LEARNING



VARIATIONAL METHODS AND DEEP LEARNING

Deep Residual Neural Networks
are connected to
Partial Differential Equations



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

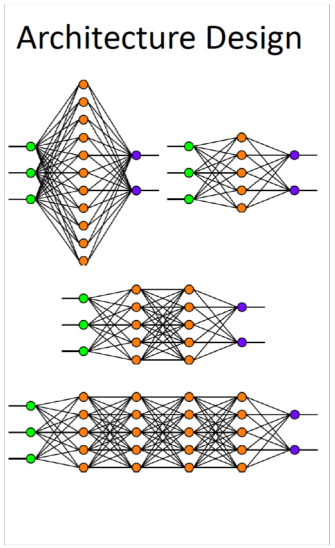
Chen, Pock - Trainable Nonlinear Reaction Diffusion, 2016

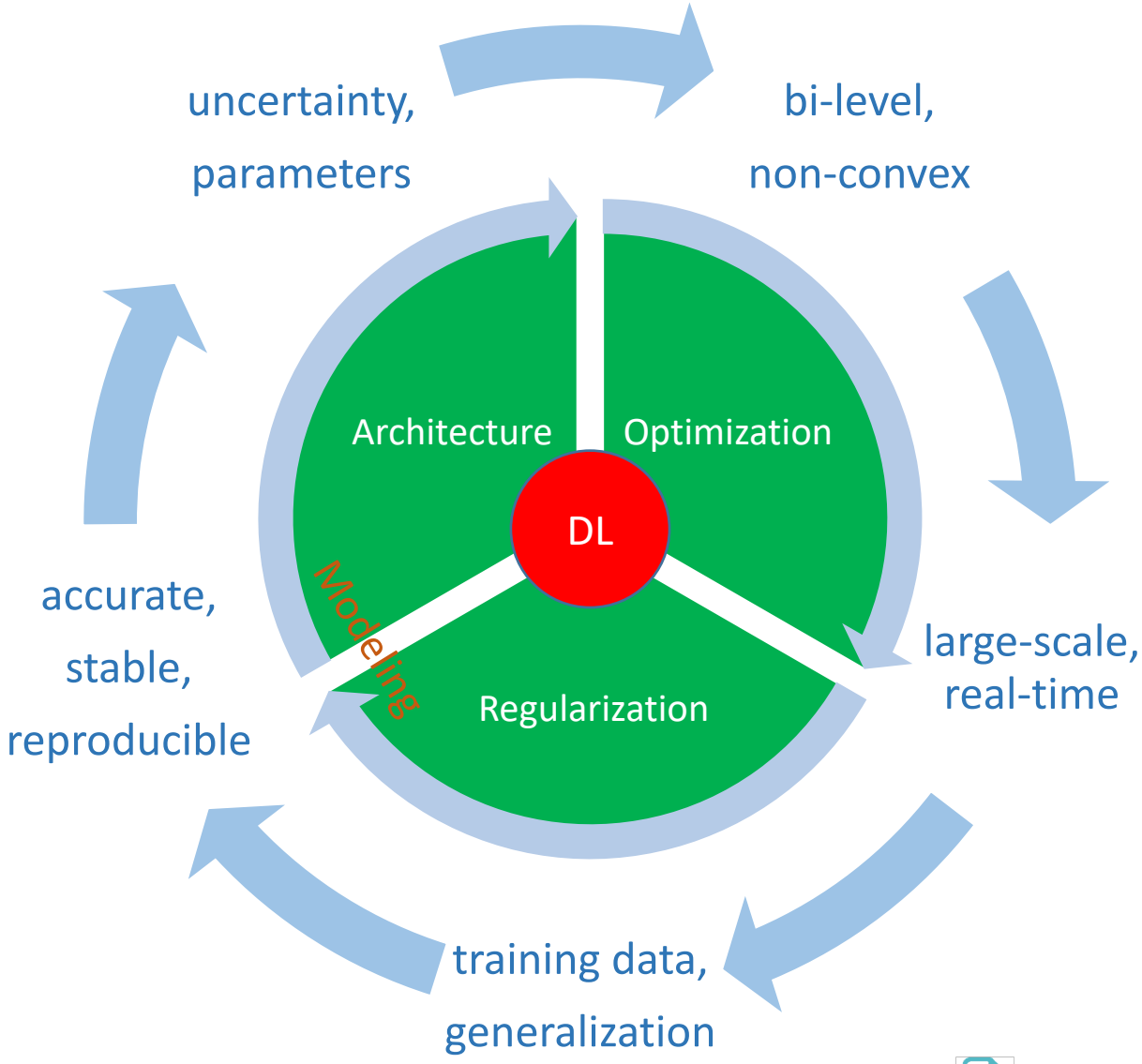
Haber, Ruthotto - Stable architectures for deep neural networks, 2018

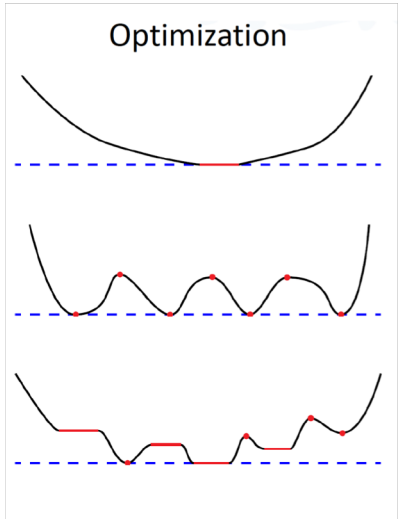
Chen et al - Neural Ordinary Differential Equations, 2018

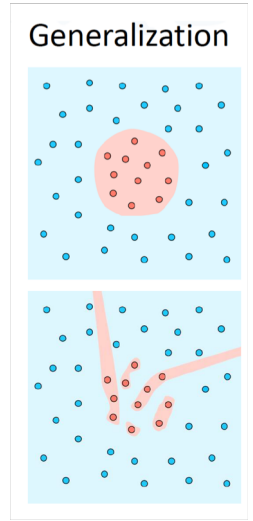
Ciccone et al - Stable Deep Networks from Non-Autonomous DEs, 2018

MATHEMATICS OF DEEP LEARNING





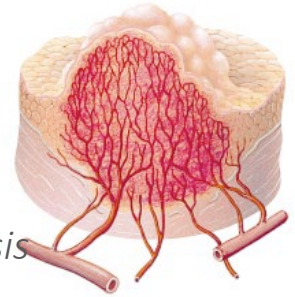




Vidal et al. Mathematics of Deep Learning, 2018

PHOTOACOUSTIC TOMOGRAPHY

Folkman, 1996



angiogenesis

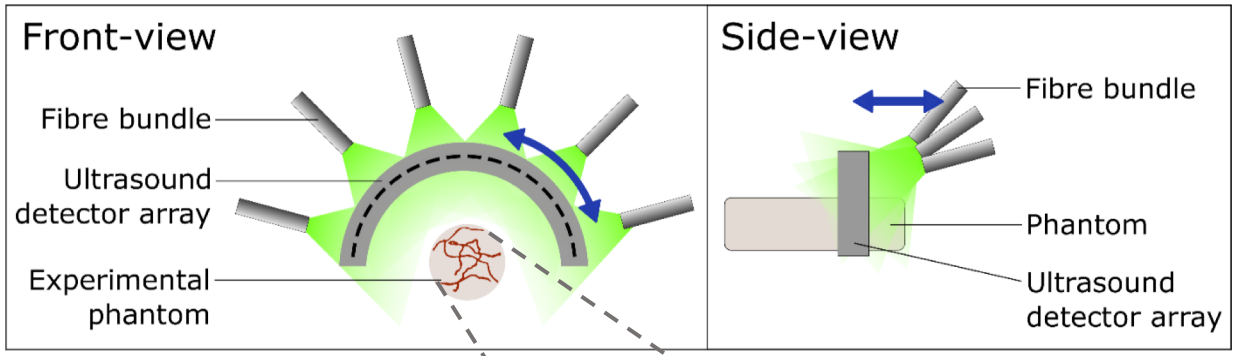


Figure: 2D slice-based imaging with rotating fibres and sensor array.[1]

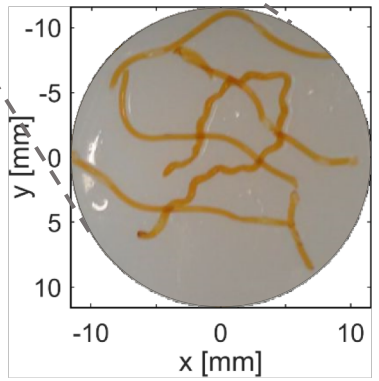
We make use of a projection model with calibration[2]:

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

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

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

PAT-operator

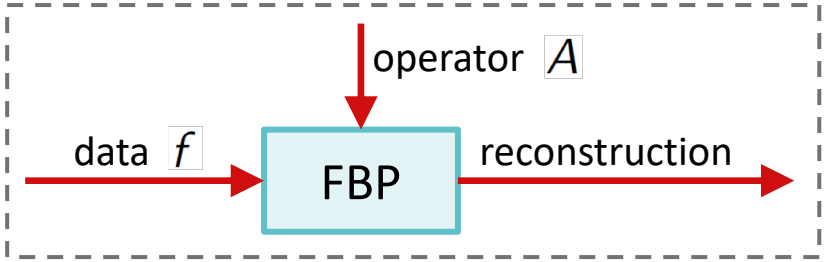


[1] Van Es, Vlieg, Biswas, Hondebrink, Van Hespen, Moens, Steenbergen, Manohar - Coregistered photoacoustic and ultrasound tomography of healthy and inflamed human interphalangeal joints (2015)

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

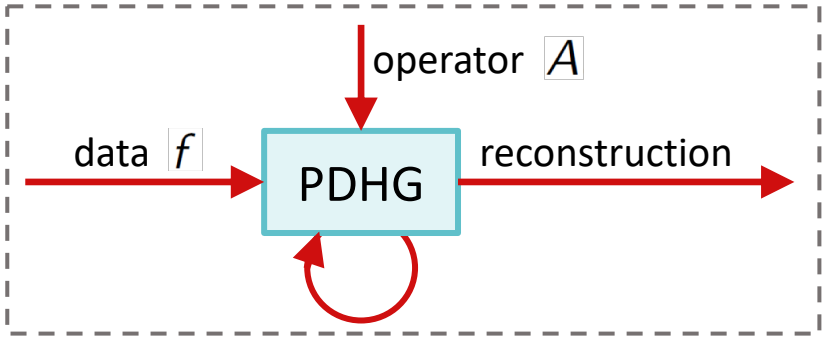
INVERSE RECONSTRUCTION METHODS

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



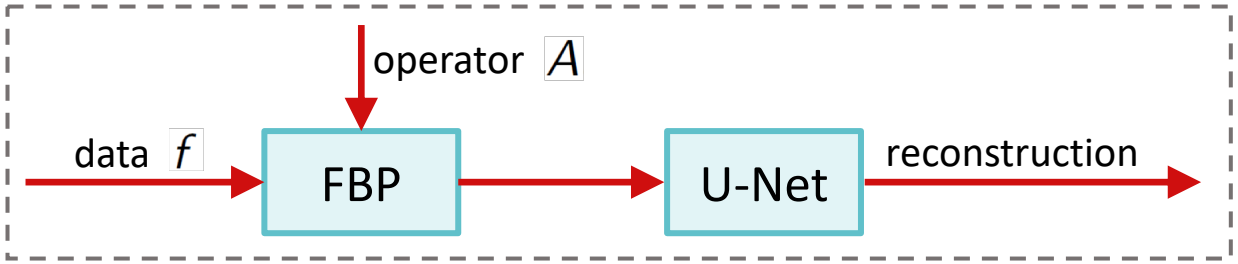
- **Iterative** reconstruction: total variation (TV)

  (solved with PDHG)



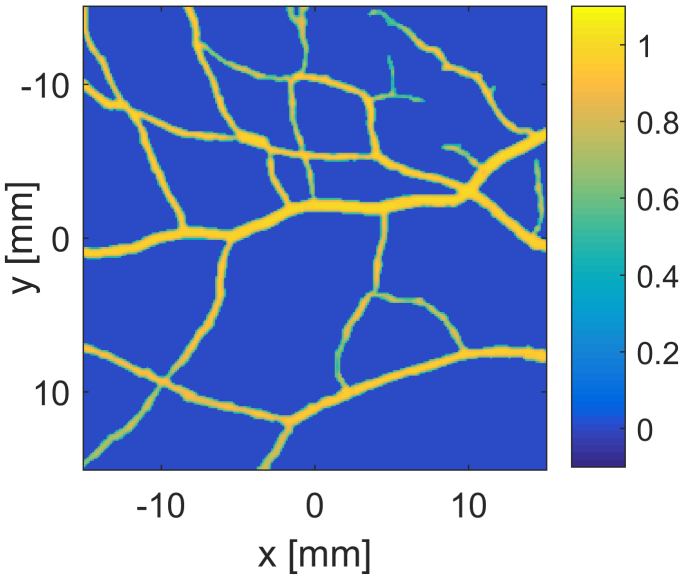
- **Learned** post-processing: U-Net [3]

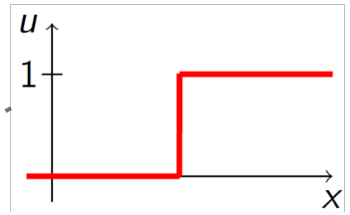
[3] Jin, McCann, Froustey, Unser - Deep Convolutional Neural Network for Inverse Problems in Imaging (2017)

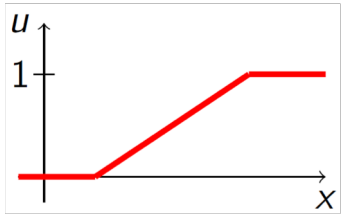


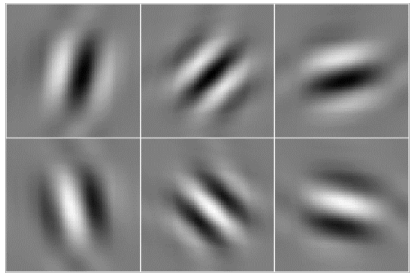
MODEL-BASED REGULARISED RECONSTRUCTION - CHOOSING FUNCTION SPACES

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









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

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)

FROM MODEL-DRIVEN TO DATA-DRIVEN

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

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

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**

$$q^{n+1}_{\{1,\dots,k\}} = q^n_{\{1,\dots,k\}} + \Gamma_{\Theta_n}\left(q^n_{\{1,\dots,k\}}, Au^n_1, f\right),$$
$$u^{n+1}_{\{1,\dots,k\}} = u^n_{\{1,\dots,k\}} + \Lambda_{\Theta_n}\left(u^n_{\{1,\dots,k\}}, A^* q^{n+1}_1\right).$$

end for

- 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)

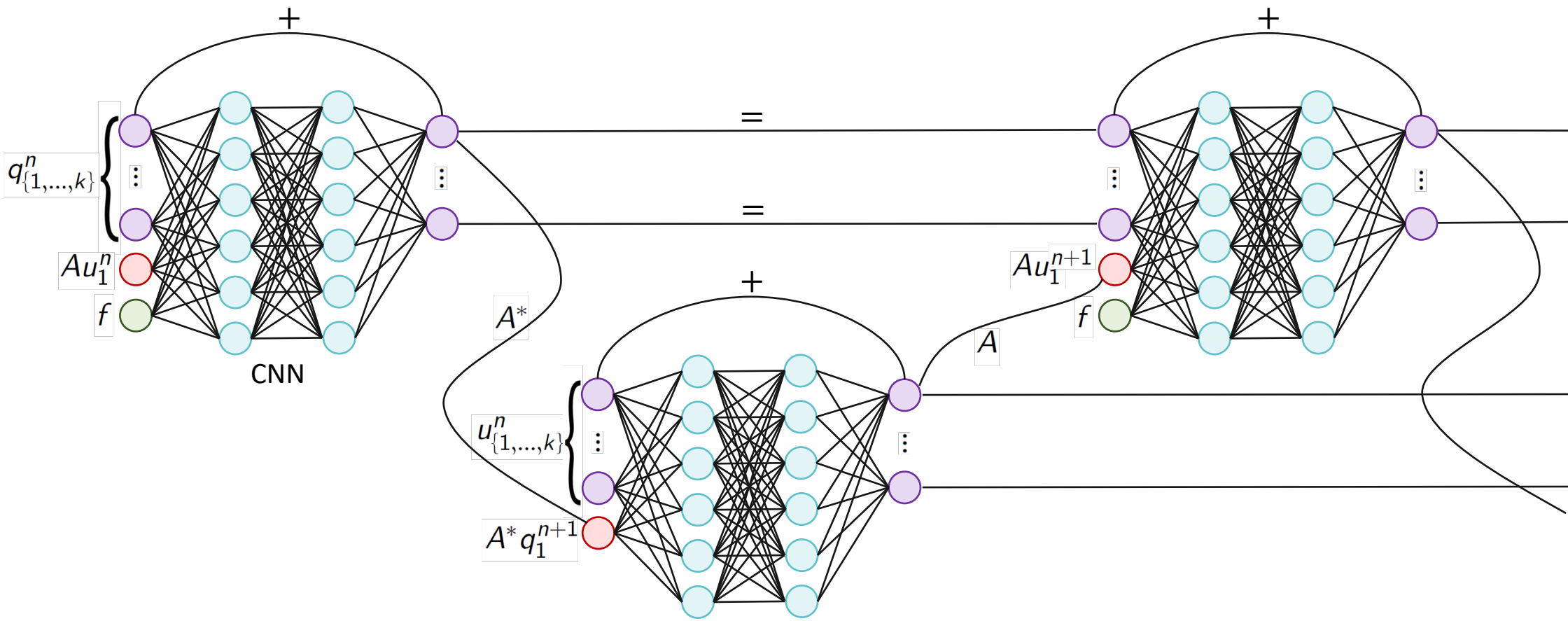
Hauptmann, Lucka, Betcke, Huynh, Cox, Beard, Ourselin, Arridge - Model based learning for accelerated, limited-view 3D photoacoustic tomography (2017)

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

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

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

end for



MODEL CHOICES PAT SIMULATION AND RECONSTRUCTION NETWORK

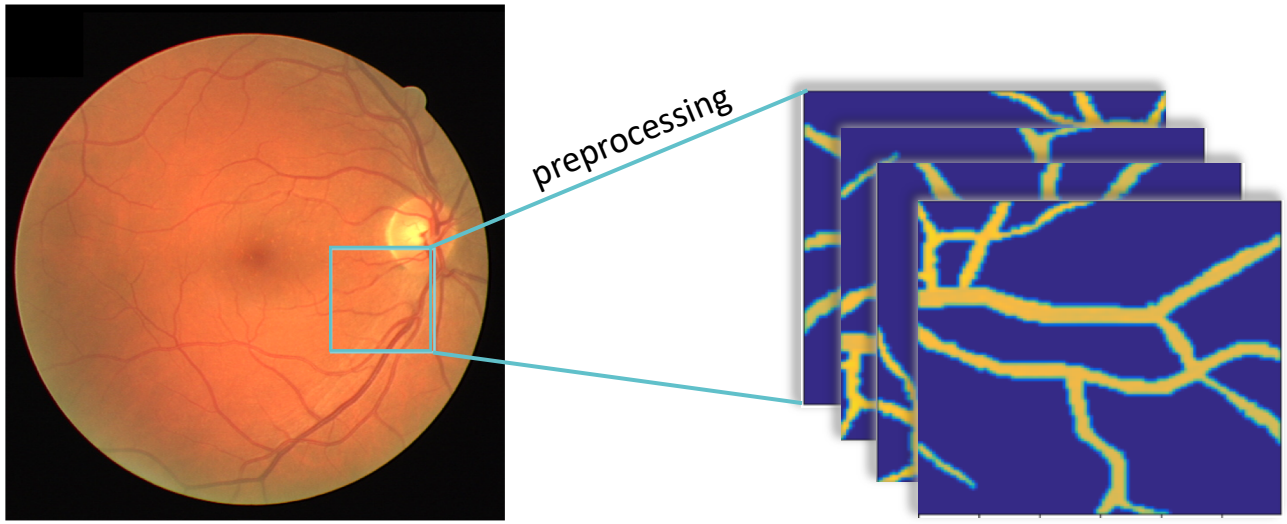
	Training
Resolution	1.5625 mm
Number of pixels	192x192
Number of sensors	32

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

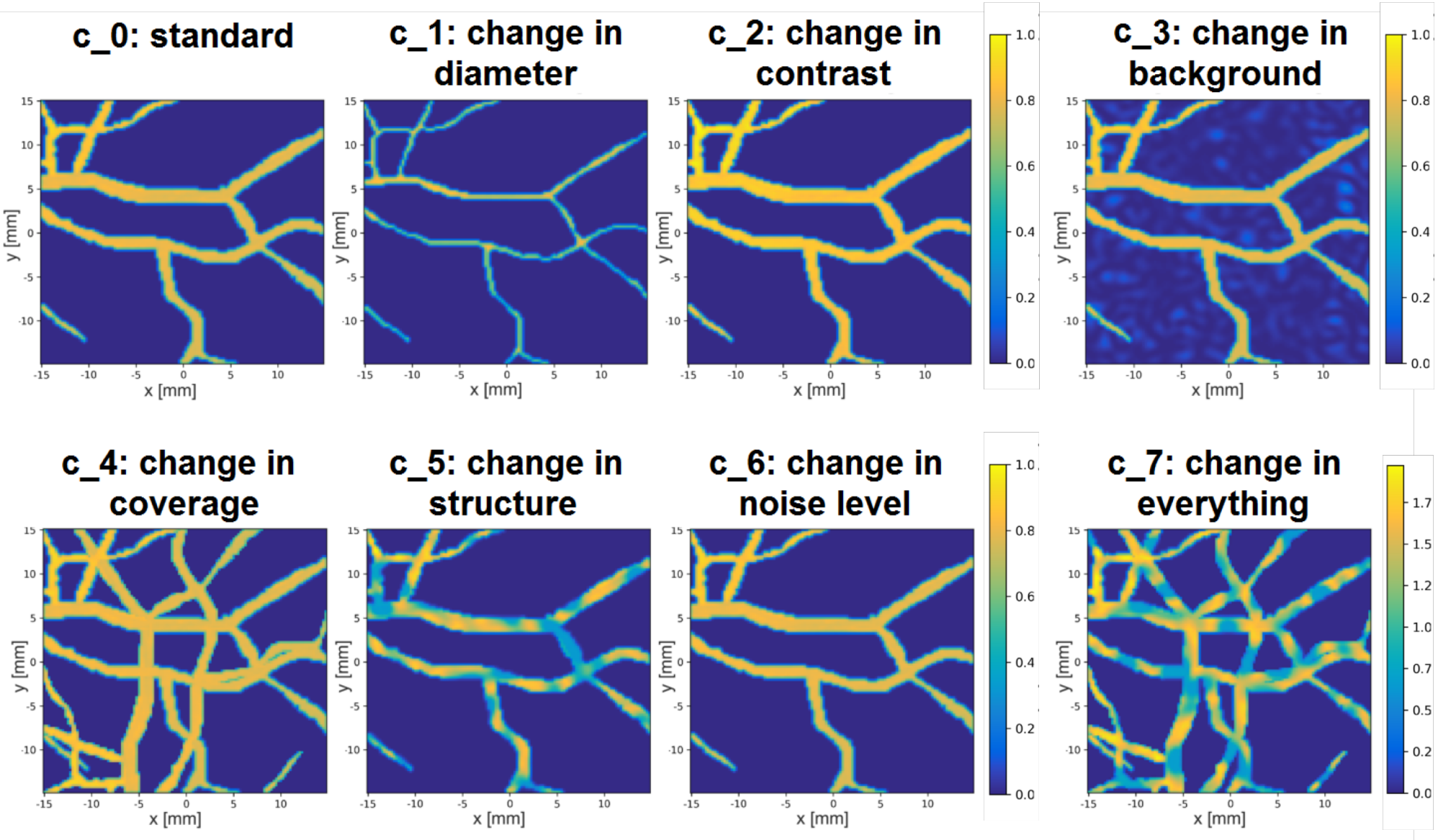


DRIVE dataset: Staal, Abramoff, Niemeijer, Viergever, Ginneken - Ridge based vessel segmentation in color images of the retina (2004)

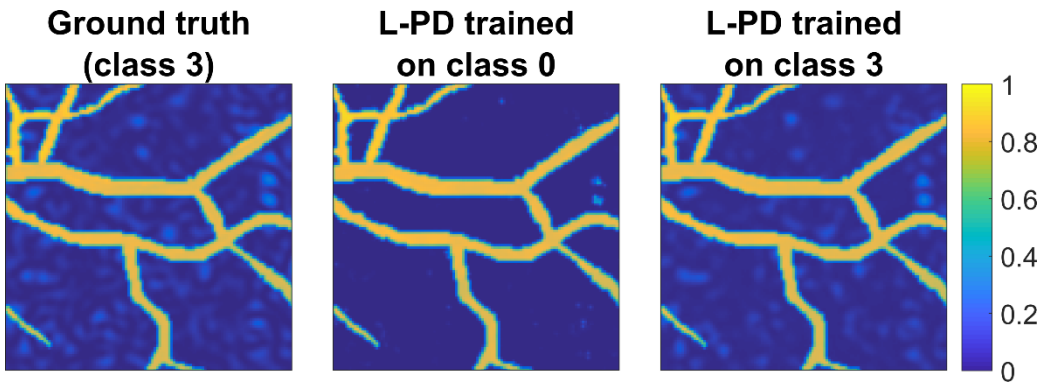
- 768 training images
- 192 test images
- scaled between 0 and 1

	'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

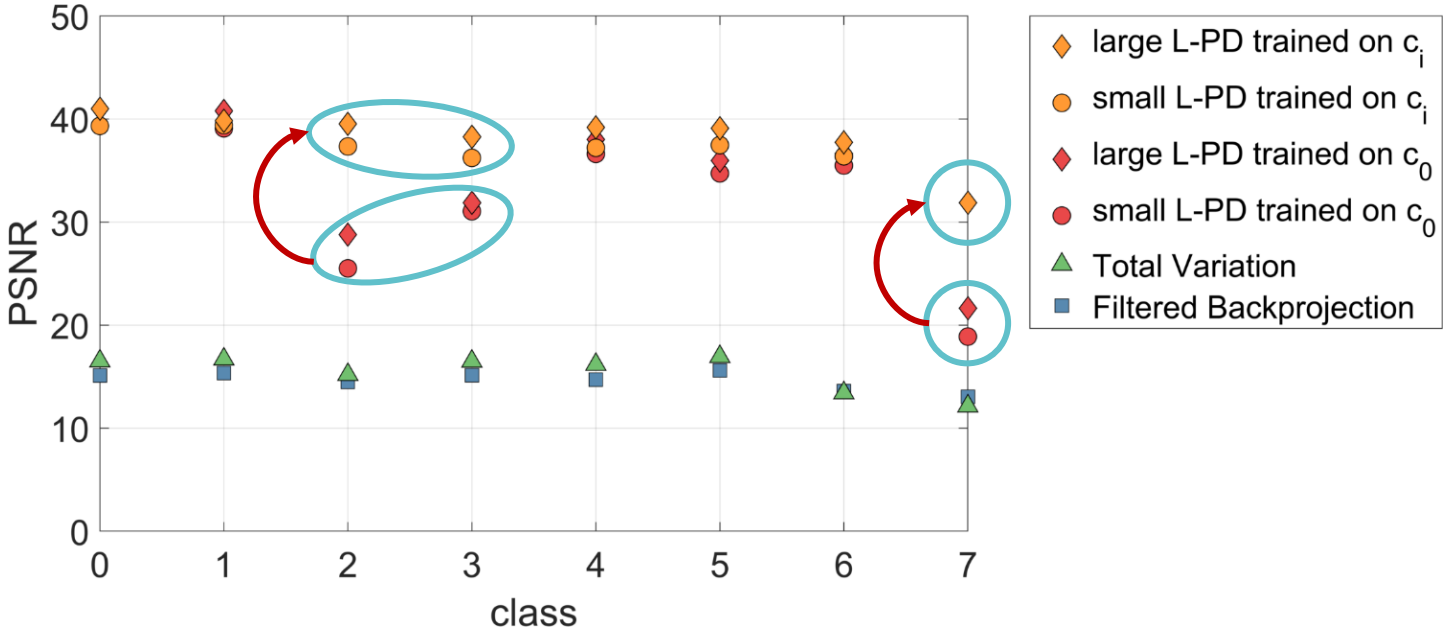
ROBUSTNESS TO IMAGE CHANGES



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.



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)

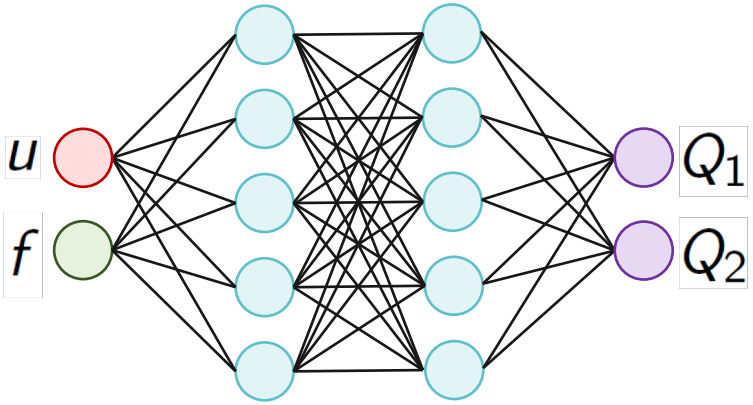
LEARNED (UNROLLED) GRADIENT DESCENT

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

and H_i represents a convolutional layer.

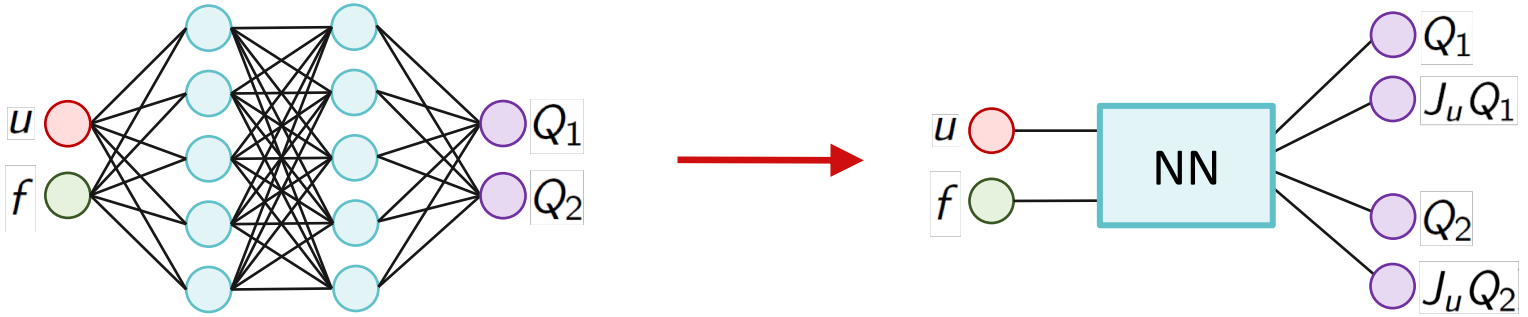


LEARNED (UNROLLED) GRADIENT DESCENT

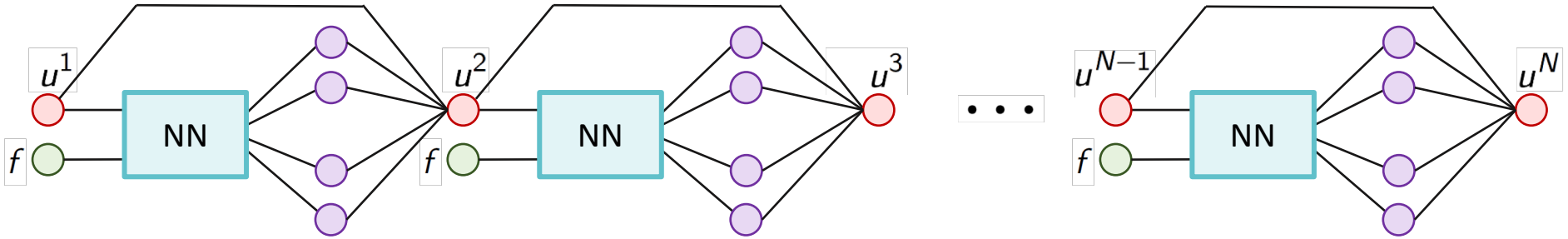
- $$G(u,f) = \|Q_1(u,f)\|_1 + \frac{1}{2}\|Q_2(u,f)\|_2^2,$$
- $$\nabla_u G = \text{sign}(Q_1) J_u Q_1 + Q_2 J_u Q_2$$
- $$\begin{aligned} u^{n+1} &= u^n - \eta \nabla_u G \\ &= u^n - \eta\,[\text{sign}(Q_1) J_u Q_1 + Q_2 J_u Q_2] \end{aligned}$$
- Train network with $\eta = 1$ or $n \in \{1, \ldots,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

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

LEARNED (UNROLLED) GRADIENT DESCENT



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

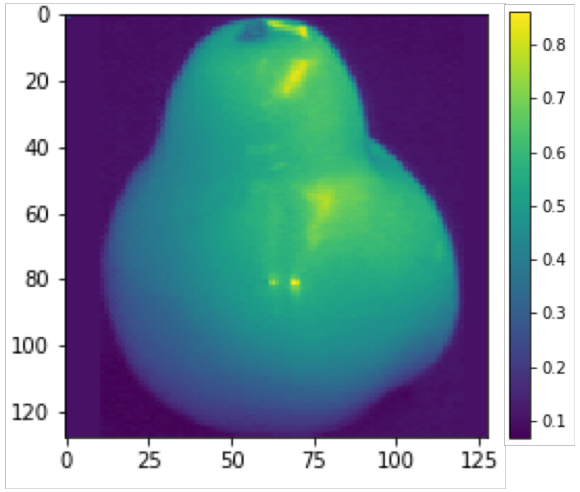


TRAINING DETAILS: COIL DATASET

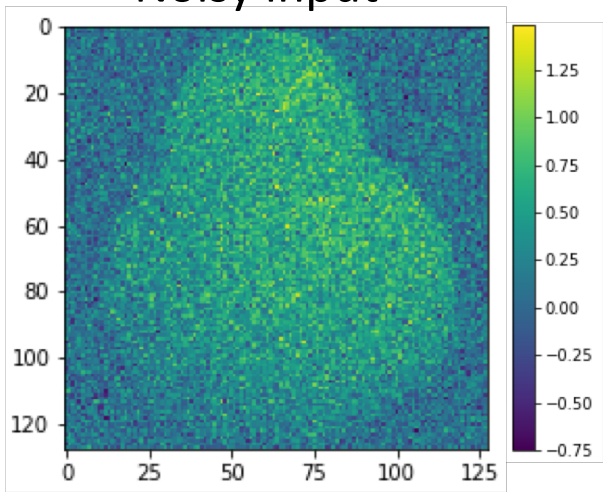
Columbia Object Image Library

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

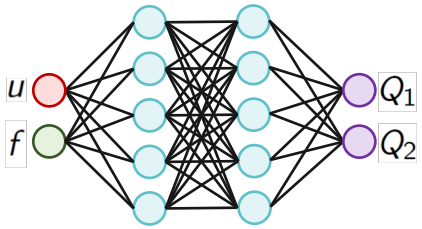
Ground truth

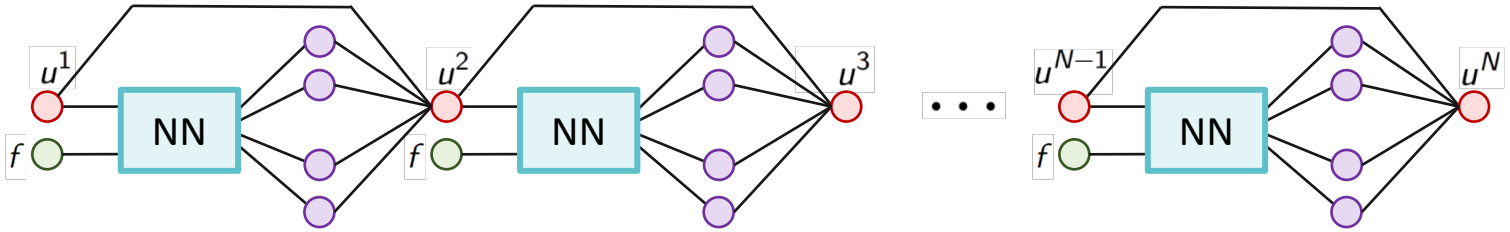


Noisy input



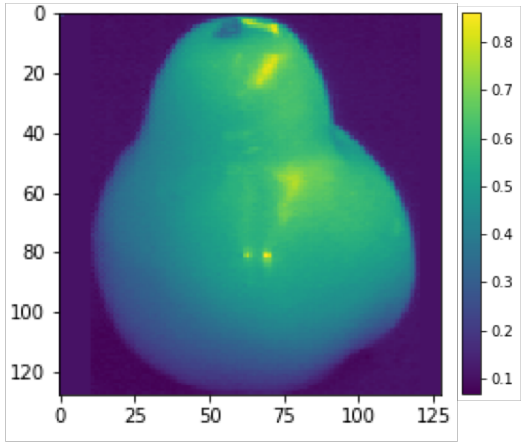
TRAINING PARAMETERS AND RESULT



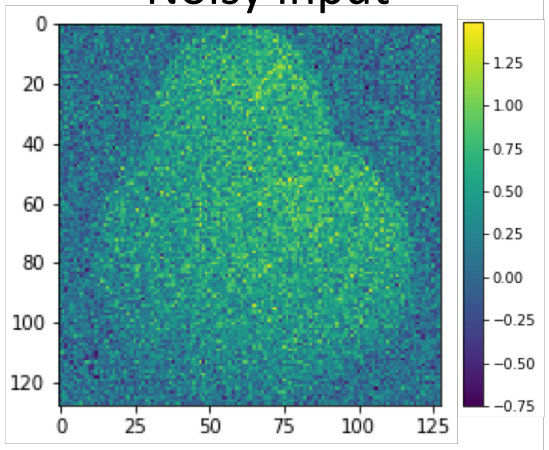


- #GD-steps = N = 20
- #layers = 3
- #channels = 8
- batch size = 9
- #epochs = 50
- Adam optimiser
  rate $2 \cdot 10^{-2} \rightarrow 10^{-3}$
- kernel size = 3x3

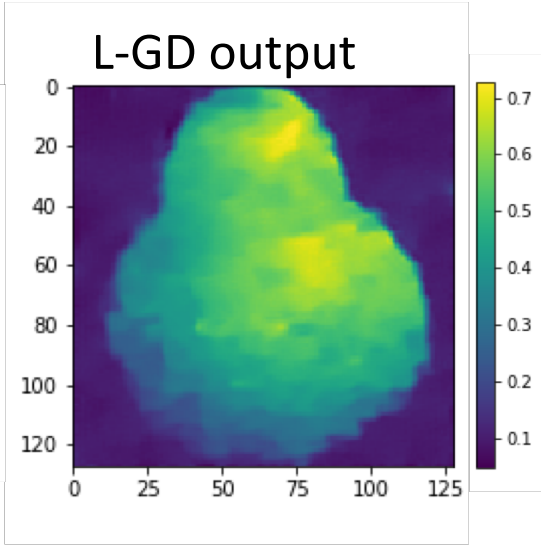
Ground truth



Noisy input

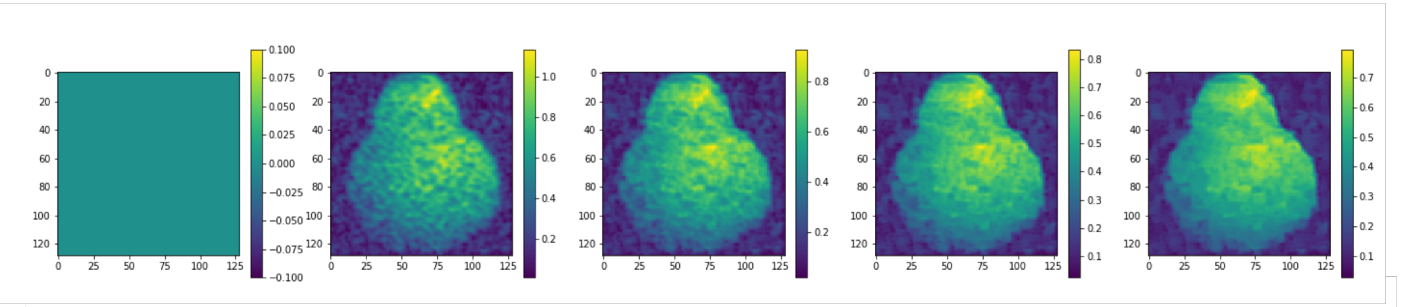


L-GD output

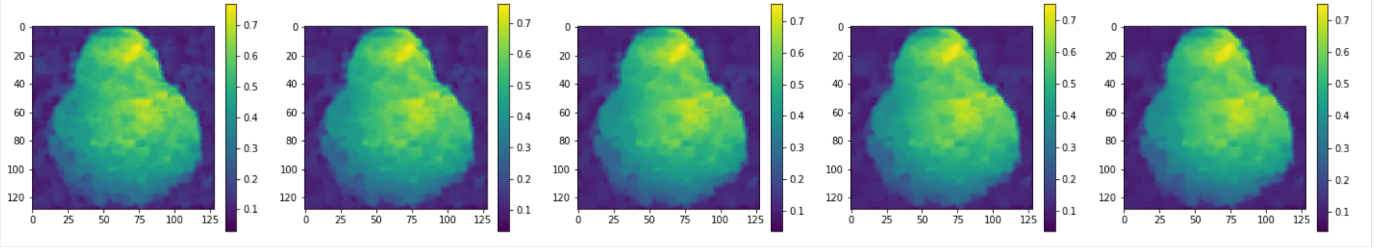


LEARNED (UNROLLED) GRADIENT DESCENT

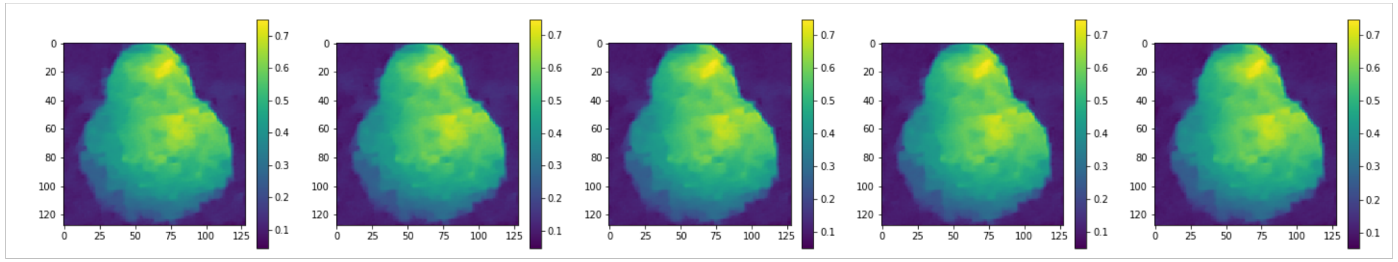
iterations 1-5



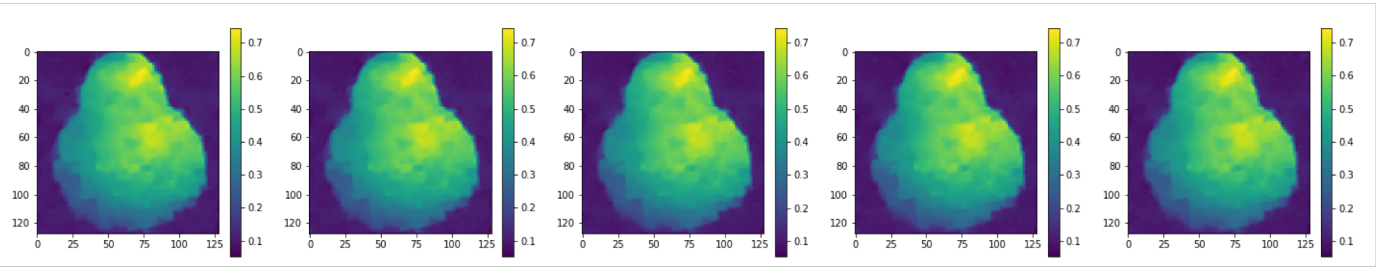
iterations 6-10



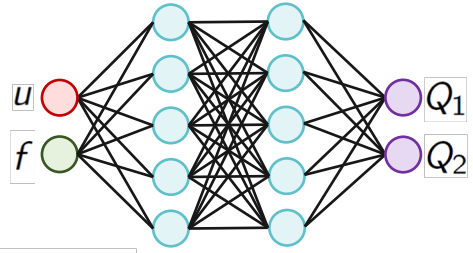
iterations 11-15



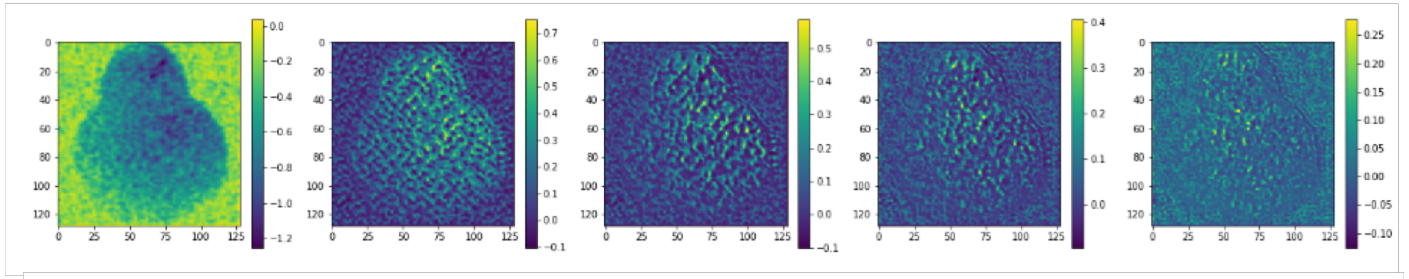
iterations 16-20



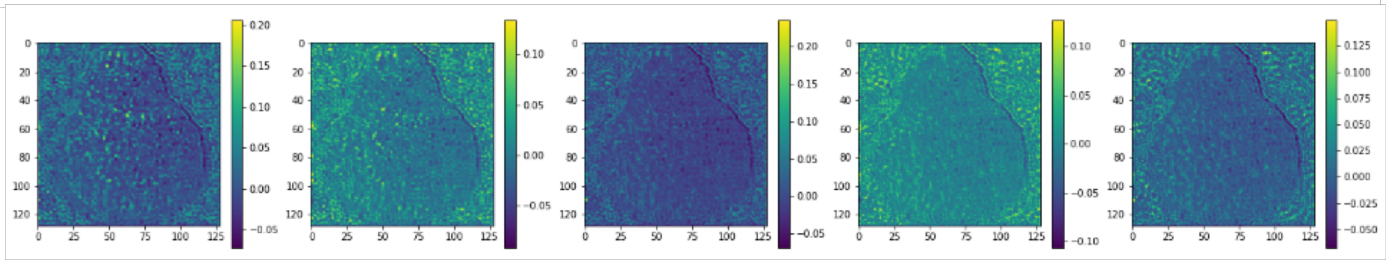
LEARNED (UNROLLED) GRADIENT DESCENT: Q_2



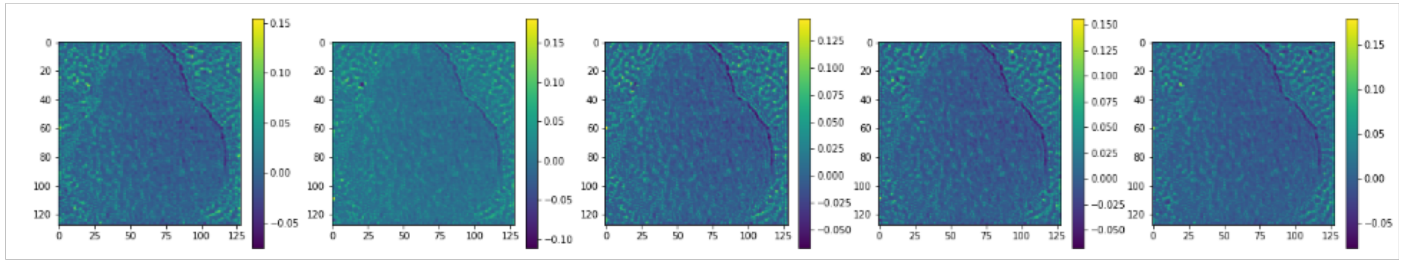
iterations 1-5



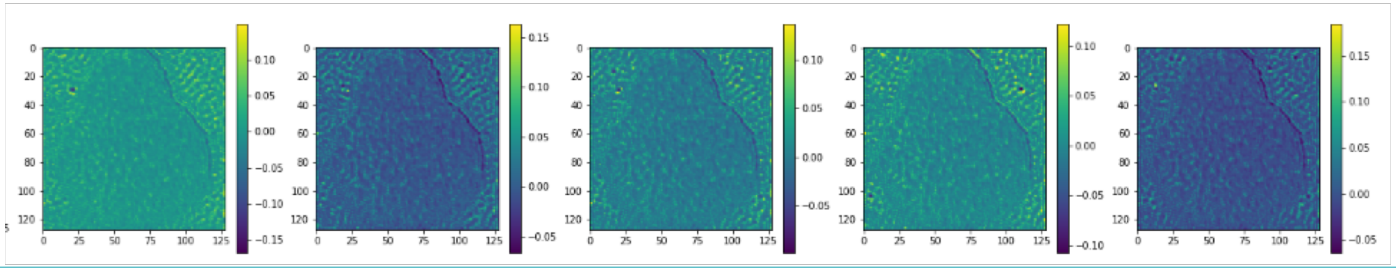
iterations 6-10



iterations 11-15

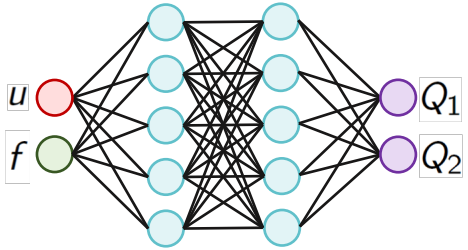


iterations 16-20

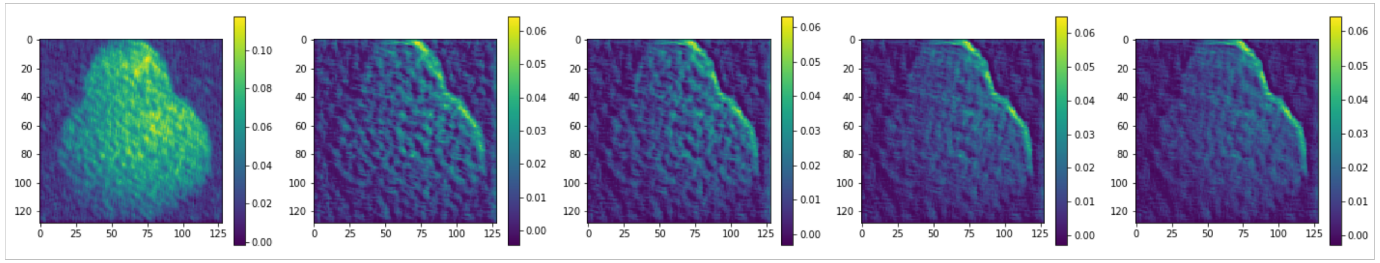


LEARNED "DATA FIDELITY" OPERATOR

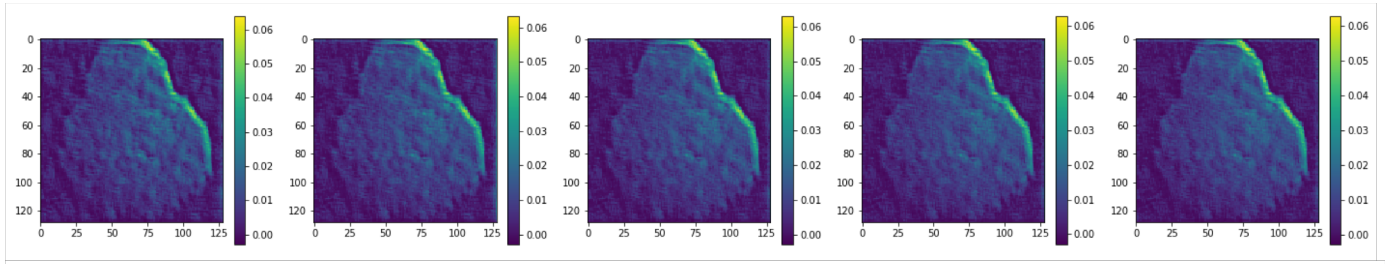
LEARNED (UNROLLED) GRADIENT DESCENT: Q_1



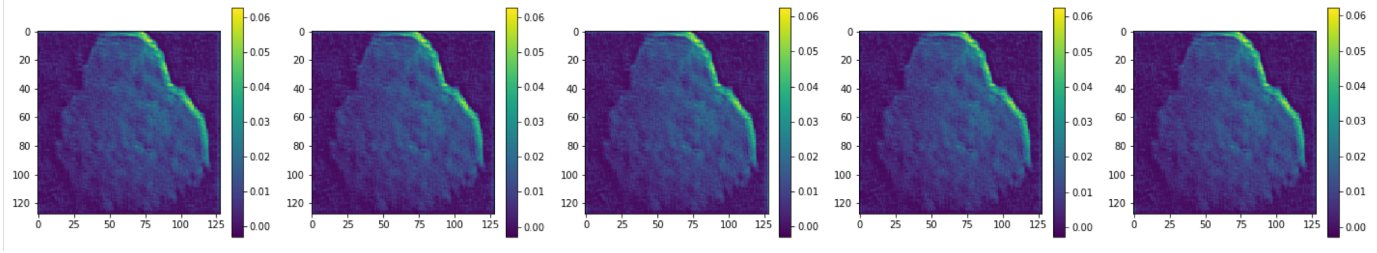
iterations 1-5



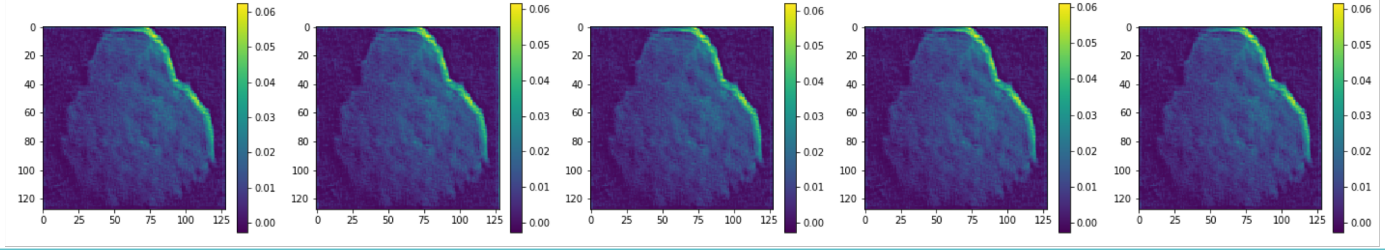
iterations 6-10



iterations 11-15



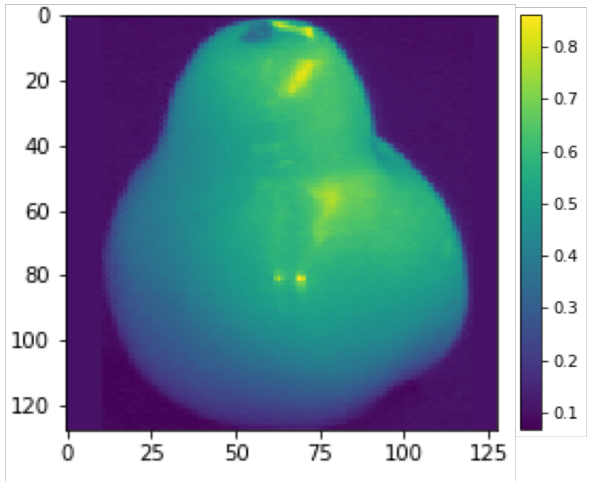
iterations 16-20



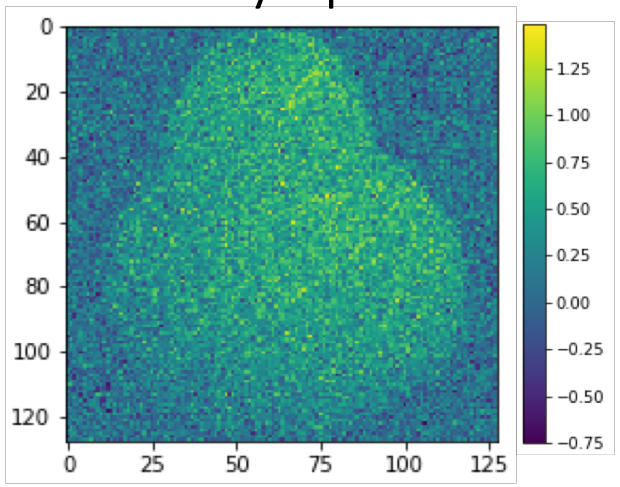
LEARNED "REGULARIZATION" OPERATOR

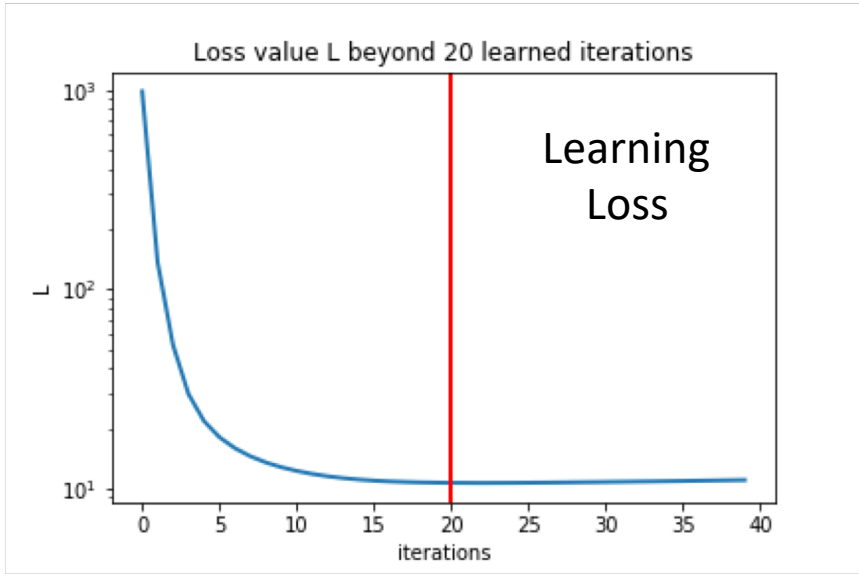
LEARNED (UNROLLED) GRADIENT FLOW

Ground truth

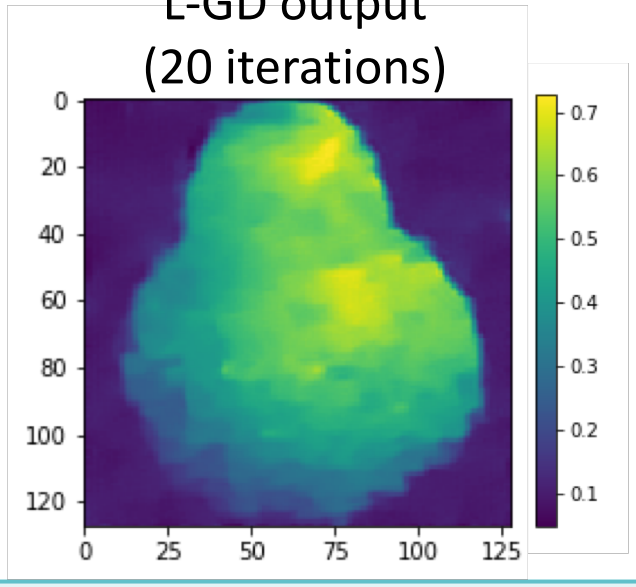


Noisy input

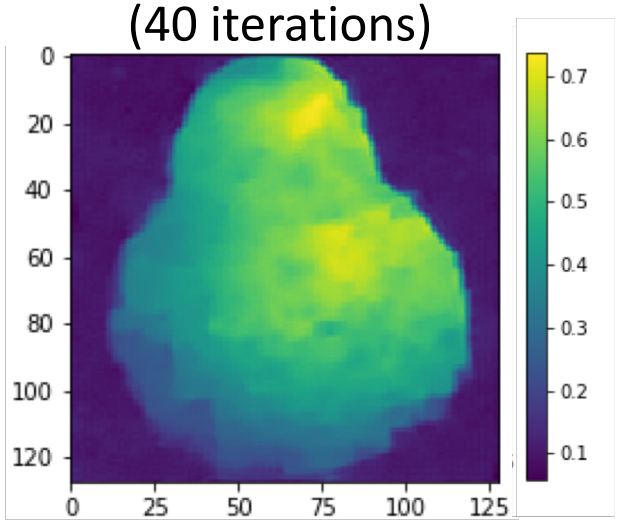


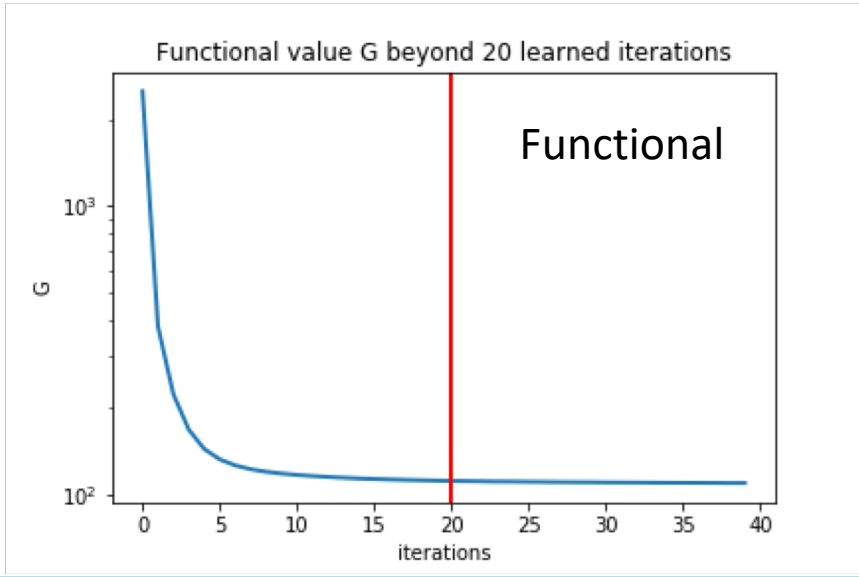


L-GD output (20 iterations)

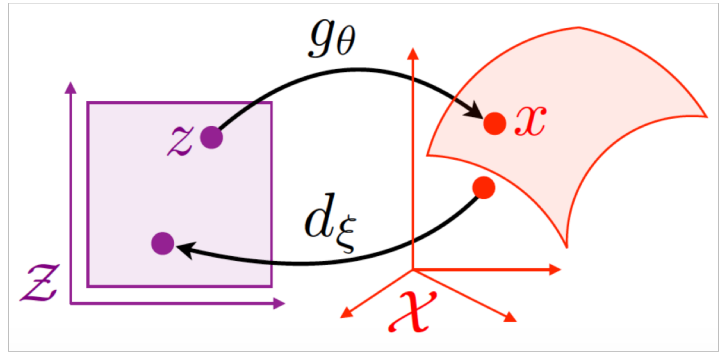


L-GD output (40 iterations)

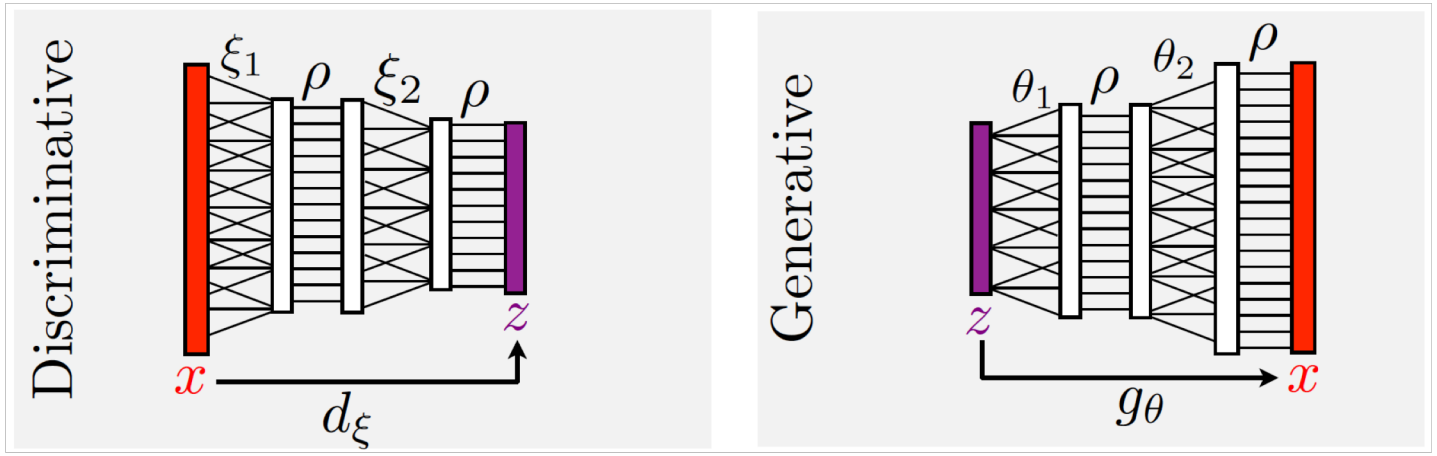


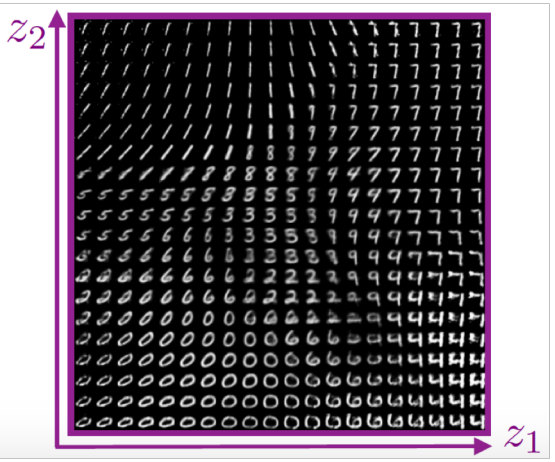


DEEP DISCRIMINATIVE VS GENERATIVE MODELS



$$\text{Deep networks:} \begin{cases} d_\xi(x) = \rho(\xi_K(\dots\rho(\xi_2(\rho(\xi_1(x)\dots) \\ g_\theta(z) = \rho(\theta_K(\dots\rho(\theta_2(\rho(\theta_1(z)\dots) \end{cases}$$





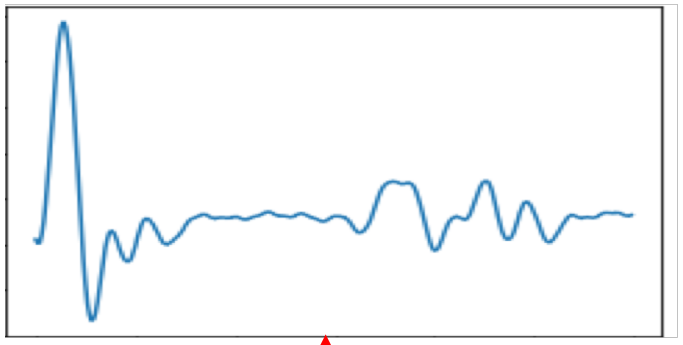
Density fitting	Optimal transport	Auto-encoders
$g_\theta(\{z_i\}_i) \approx \{x_i\}_i$	map d_ξ	$g_\theta(d_\xi(x_i)) \approx x_i$

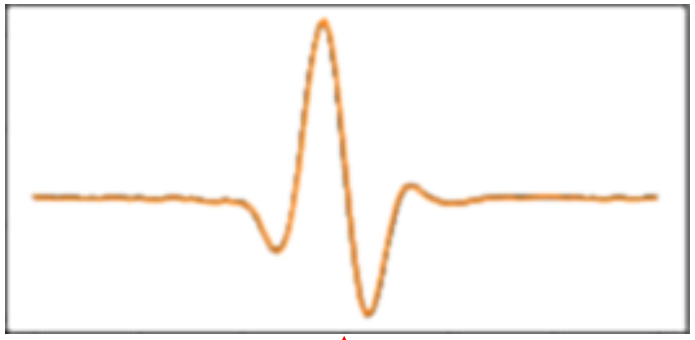
Genevay, Peyre, Cuturi – GAN and VAE from an Optimal Transport Point of View (2017)

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

(BLIND) DECONVOLUTION PROBLEM IN PHOTOACOUSTICS

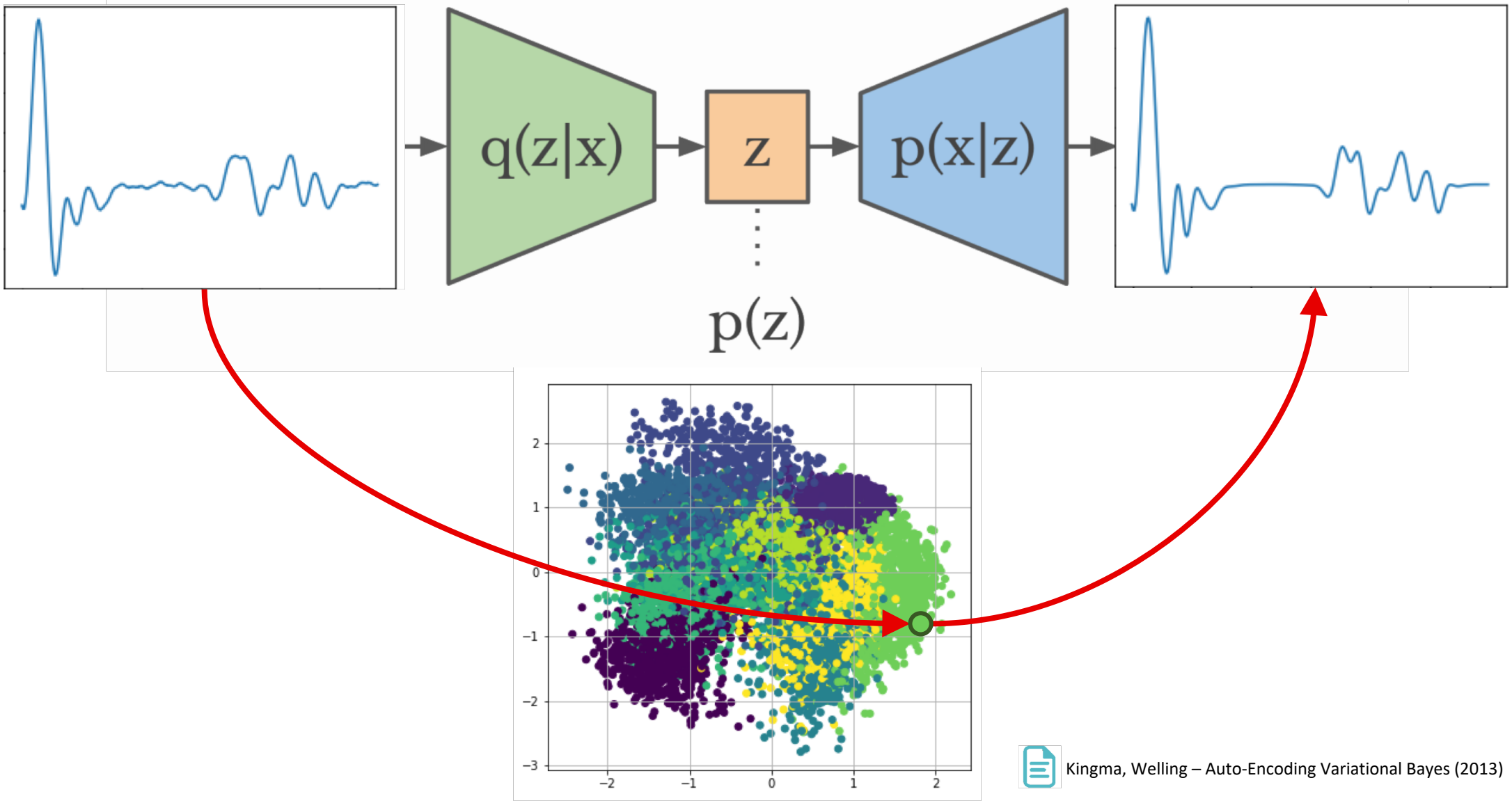
LEARNING UNCERTAINTY - CALIBRATION IN FORWARD MODEL





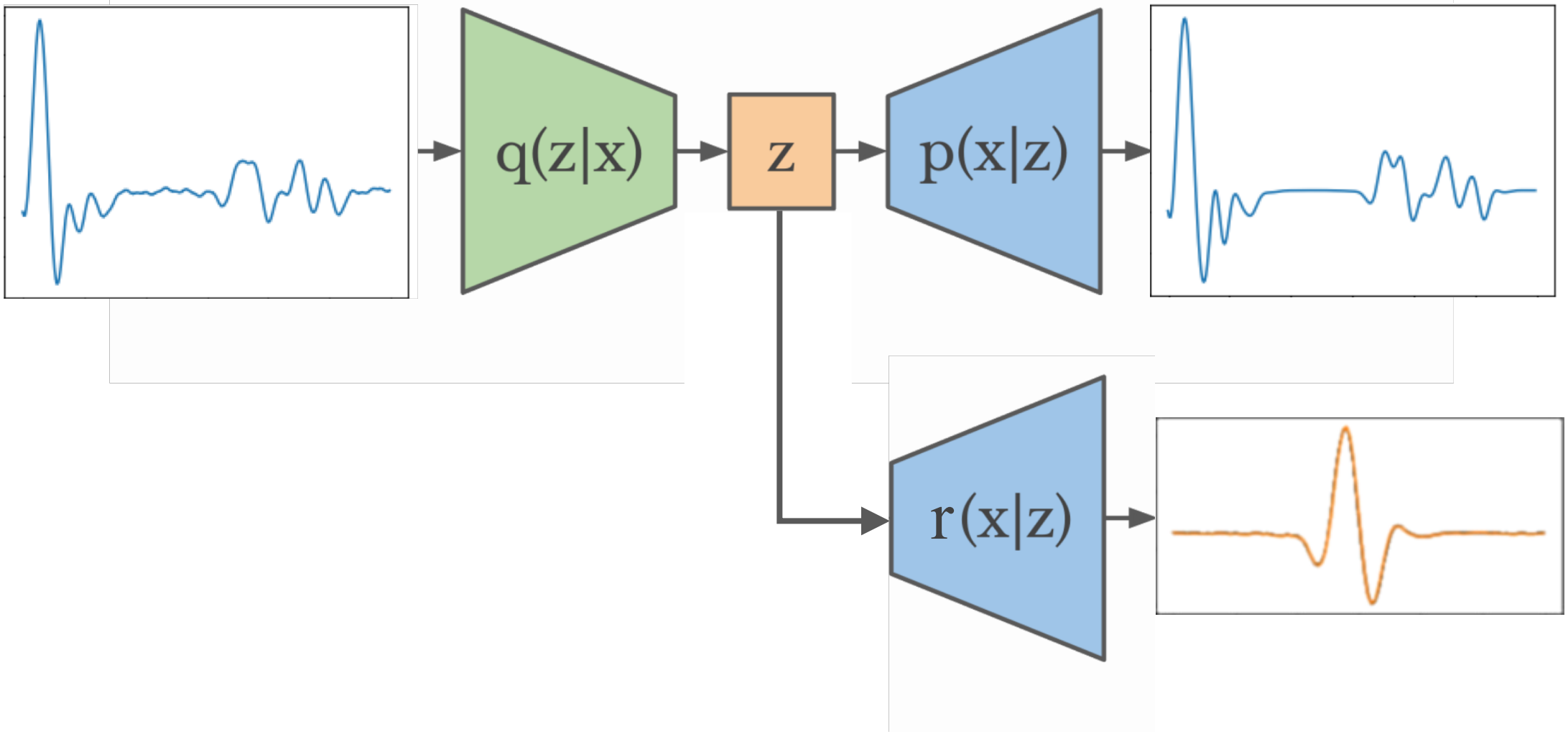
$$\tilde{p}(x,t) = \left(\frac{1}{t} \iint_{|x-\tilde{x}|=ct} u(\tilde{x})\, \mathrm{d}\tilde{x} \right) *_t p_{\mathrm{cal}}(t),$$

LEARN DATA-SPACE WITH VARIATIONAL AUTOENCODER

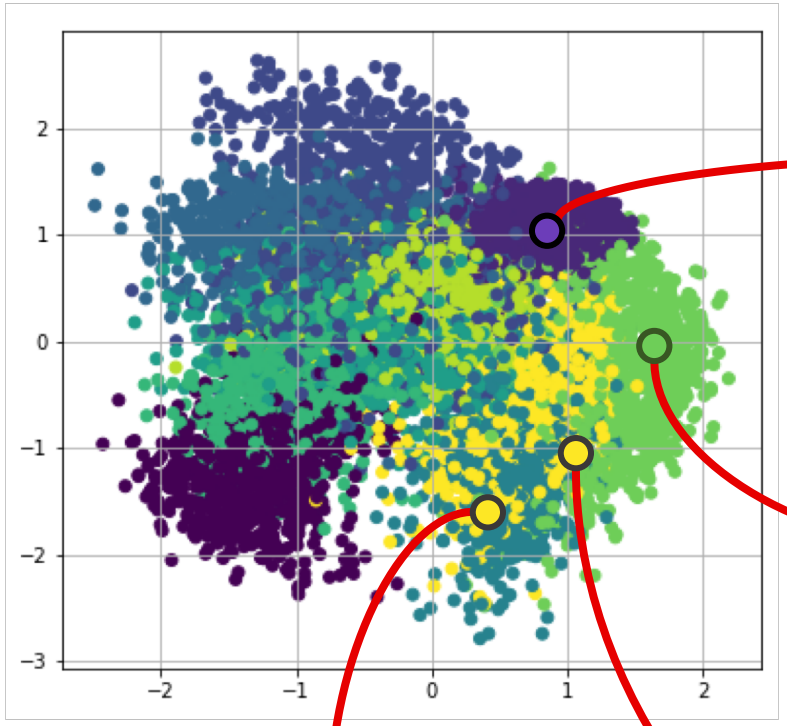


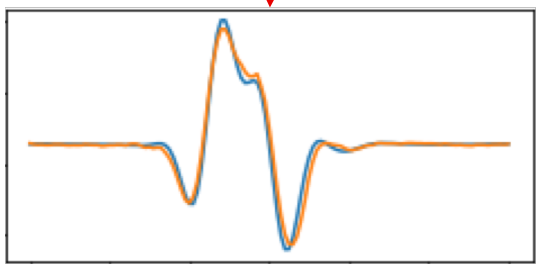
Kingma, Welling – Auto-Encoding Variational Bayes (2013)

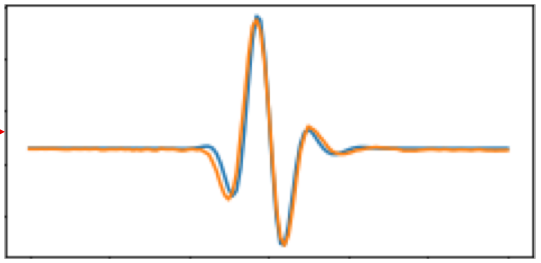
LEARN DATA-SPACE WITH VARIATIONAL AUTOENCODER

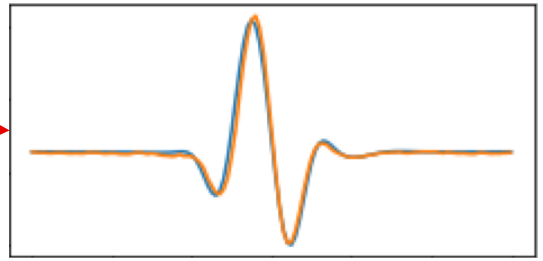


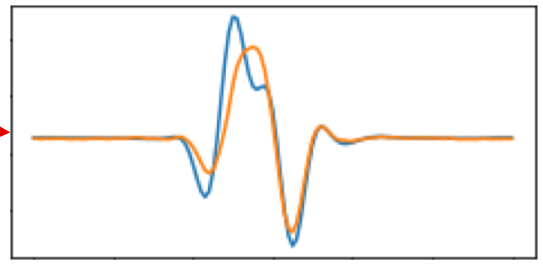
LEARNED DATA-SPACE WITH VARIATIONAL AUTOENCODER









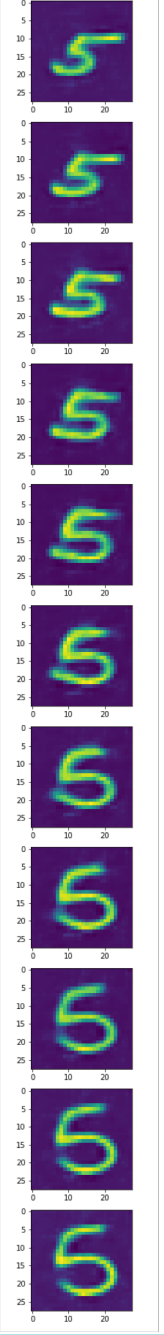


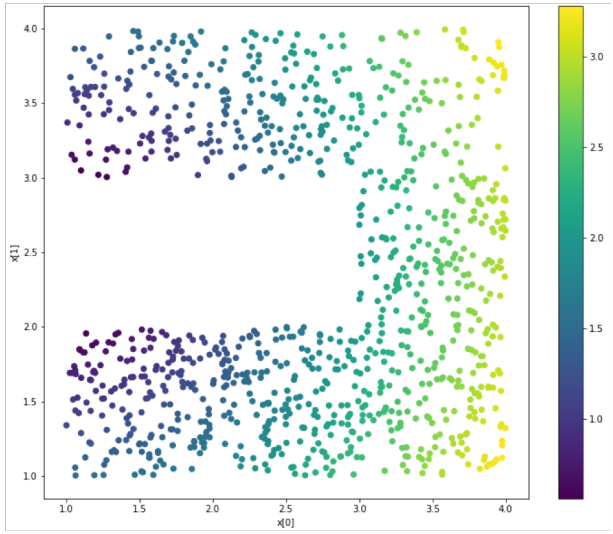
$p_{\text{cal}}(t)$

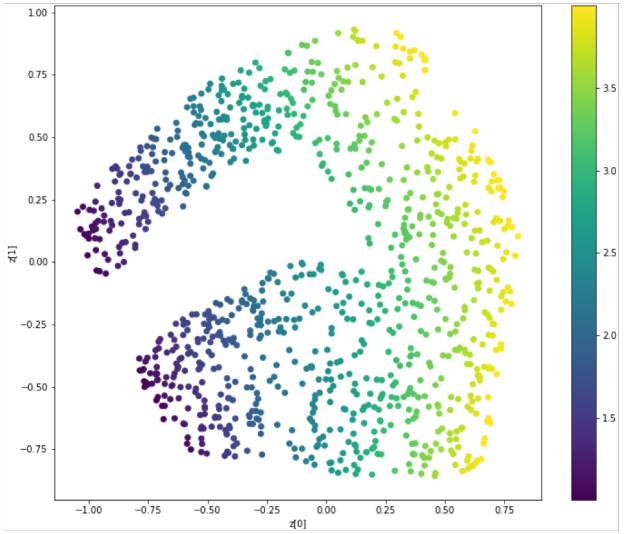
ground truth

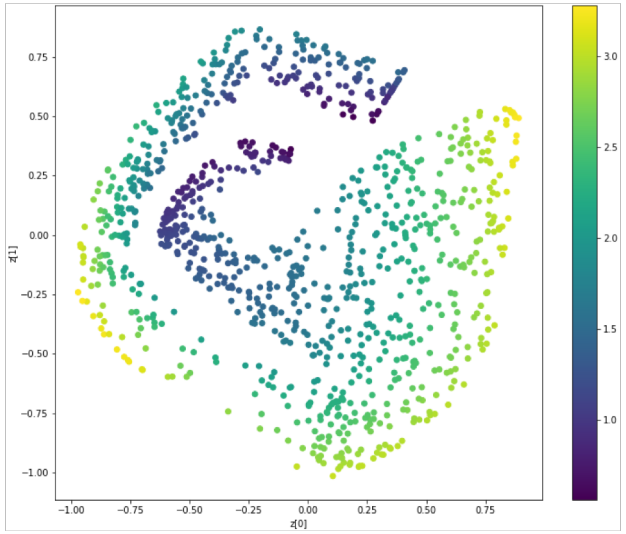
reconstruction

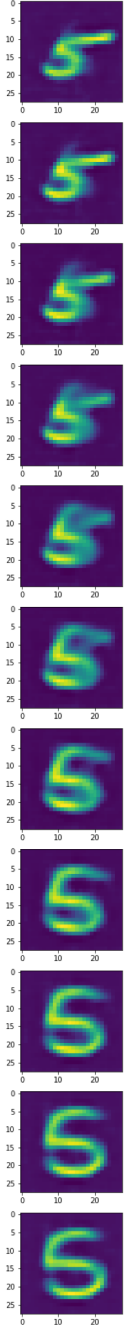
LIMITATION: GEOMETRIC VAE LATENT SPACE REGULARIZATION











Regularization of latent variables z

CONCLUSIONS

1. **Robustness**. Deep learning improves PAT reconstruction quality particularly in uncertain cases.

2. **Functional learning**. Mathematical theory on stability and convergence not available. Construct learned methods that learn the functional to be minimised via a gradient flow.

3. **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

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)

GRADIENT FLOWS

$\partial_t \rho = \Delta \rho$	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