Deep learning for medical image reconstruction

Models, priors and data

Ruud van Sloun
Image reconstruction

Magnetic Resonance Imaging (faster, low-field etc.)

Computed Tomography (low-dose, sparse-view)

Ultrasound (high-quality, reduced data-rates)
Ultrasound imaging basics

- Echoes result from scattering in the tissue (Coherent & incoherent - speckle)

- The image is formed by estimating (reflected) signal amplitude from a set of spatial locations (i.e. pixels) using an array of sensors = Beamforming
**Ultrasound imaging basics**

**Transmission**

**Channel Data**

- High rate ADC (~20-50MHz / element)
- Hundreds to thousands of channels/firing
- High-data rates

**Reconstructed RF scanline**

**Spatial filtering**

- Sidelobes & grating lobes
- Mainlobe width (resolution)

**Spatial filtering = beamforming**

- The better the spatial selectivity,
  - The better the tissue contrast and resolution

How much signal is reflected from this particular pixel?

*Image from “Speckle reduction imaging” by Milkowski et al.*
Ultrasound imaging basics

Image quality is a function of:

1. Physics: array geometry and probe bandwidth
   *Increased array aperture – increased angular resolution*

   - Better spatial selectivity,
   - Better tissue contrast and resolution

2. Algorithms: powerful digital signal processing and beamforming on RF channel data

   ![Graphs showing beampatterns with different aperture sizes](image)

   - Adaptive apodization calculator

   ![Diagram of beamformed value for point scatterer](image)

   → AI opportunity
Ultrasound imaging AI opportunities

**Parallel ultrafast acquisition**
- High time-resolution
- Compromising spatial resolution and contrast
- Relies more heavily on receive spatial filtering/beamforming

**High image quality under minimal data rates**
- Improve tissue contrast (*accurate* contrast)
- Resolution depends on array aperture -> high-res with small/sparse aperture?
- Compressed sensing to reduce data rates at the probe

---

Focused transmit

Parallel ultrafast transmit

- Full aperture
- Small aperture (compact)
- Sparse aperture (less data)
Image basics: MRI

Magnetic Resonance Imaging (faster, low-field etc.)

Inverse FT

K-space

Pulse sequence

Total time $\sim TR \times \text{amount of repetitions/lines}$

Can we go faster?
Image basics: MRI

Magnetic Resonance Imaging (faster, low-field etc.)

- Reconstruction algorithm
- Sampling/acquisition
- Priors
- How?

Undersampled K-space

- \( p(x|y) \)?
- \( p(x) \)?

Full K-space

\( p(y|x) \)?

= less acquisitions = faster
Image reconstruction

Fast MRI
- Compressed sensing/fewer acquisitions
- Acceleration
- Aliasing artefacts

Low-field MRI
- Lower field strengths: compact MRI machines
- Low SNR k-space -> less clear high-frequencies -> low resolution
Image reconstruction

Low-dose CT
- Safer – less ionizing radiation
- Low SNR -> limited fidelity

Sparse CT
- Reduced scan time and improved time resolution
- Lower dose
- Undersampling artefacts
Image reconstruction

Magnetic Resonance Imaging
(faster, low-field etc.)

Computed Tomography
(low-dose, sparse-view)

Ultrasound
(high-quality, reduced data-rates)
What gaps can we fill by learning?

Challenges in classical image recon based on models

- Acquisition model assumptions incorrect/imprecise
- Statistical image priors not sufficiently expressive/accurate
- Slow image reconstruction with high complexity (e.g. iterative, matrix inversions etc)
Models and Priors: ultrasound

\[ p(y|x) : \text{Likelihood of measurements } y \text{ given object } x \]

Across our imaging pipeline and set of applications, \( y \) can be:
(1) RF channel data (array response)
(2) Beamformed RF data
(3) Image data (beamformed + envelope detected)

Acquisition model assumptions incorrect/imprecise
(multiple scattering, aberration etc.)
Models and Priors: ultrasound

Array response (narrowband) single target:
\[ y(t) = a(\theta)x(t) + n(t) \]

- **Noise vector:**
  - Sensor noise
  - Off-axis scattering/interference
  - Reverberation

- **Array response vector:** (assumed constant SoS, no aberration)

**Typical assumption:** \( n(t) \sim \mathcal{N}(0, \sigma_n^2 I) \)

**MSE-optimal matched filter:**
\[ \hat{x}(t) = w^H y(t) \]

\((w = a)\)

\(= \) delay-and-sum (DAS) for wideband

Acquisition model assumptions incorrect/imprecise (multiple scattering, aberration etc.)
Models and Priors: ultrasound

Array response (narrowband) single target:
\[ y(t) = a(\theta)x(t) + n(t) \]

- **Array response vector:** (assumed constant SoS, no aberration)
- **Noise vector:**
  - Sensor noise
  - Off-axis scattering/interference
  - Reverberation

**Model-based alternatives to DAS:**
- **MVDR** (minimize total (noise) power but retain unity gain)
- **iMAP** (assume \( x(t) \sim \mathcal{N}(0, \sigma_x^2) \); iteratively estimate \( \sigma_x, \sigma_n \) per pixel)
- **ADMIRE** (aperture domain model of signal and noise/interference/clutter; separate using optimization methods)

Jensen et al.
Eldar et al.
Byram et al.
Models and Priors: ultrasound

\[ y(t) = a(\theta)x(t) + n(t) \]

- **DAS** (matched filter)
- **MVDR** (minimize total (noise) power but retain unity gain)
- **iMAP** (assume \( x(t) \sim \mathcal{N}(0, \sigma_x^2) \); iteratively estimate \( \sigma_x, \sigma_n \) per pixel)
- **ADMIRE** (aperture domain model of signal and noise/interference/clutter; separate using optimization methods)

**General problem:** model \( p(y|x) \) too simple / incomplete / or hard to solve due to suboptimal prior \( p(x) \)
Models and Priors: ultrasound

\[ y(t) = a(\theta)x(t) + n(t) \]

- **DAS** (matched filter)
- **MVDR** (minimize total (noise) power but retain unity gain)
- **iMAP** (assume \( x(t) \sim \mathcal{N}(0, \sigma_x^2) \); iteratively estimate \( \sigma_x, \sigma_n \) per pixel)
- **ADMIRE** (aperture domain model of signal and noise/interference/clutter; separate using optimization methods)

**General problem:** model \( p(y|x) \) too simple / incomplete / or hard to solve due to suboptimal prior \( p(x) \)
Models and Priors: MRI

\[ p(y|x): \text{Likelihood of measurements } y \text{ given object } x \]

K-space measurement:

\[ y = AFx + n \quad \text{A: Subsampling matrix} \]
\[ F: 2D \text{ Fourier transform} \]

For low-field: \( n \) is dominating higher frequencies
For fast-MRI: \( A \) is sampling below Nyquist (compressed sensing)

Classical reconstruction: \( x = F^{-1}z \) \( (z = \text{zero-filled k-space } y) \)

Inverse (MAP) problem:

\[
\hat{x} = \arg\max_x \quad p(y|x)p(x) \\
\hat{x} = \arg\min_x \quad \frac{1}{2}||y - AFx||_2^2 - \log p(x)
\]

General problem: prior \( p(x) \) not sufficiently expressive
What if we could improve our models using deep learning without ending up just throwing away what we do know about the problem?
How to bring together to get the best of both?

Van Sloun et al. Proceedings of the IEEE, 2020
Luijten, ..., Van Sloun. IEEE trans. med. im., 2020
Solomon, ... van Sloun, Eldar. IEEE trans. med. im., 2019
One step back: Model-based optimization

**General recipe:** construct a model-based optimization algorithm based on a-priori knowledge of the measurement process and statistical priors

Define measurement model with image of interest $x$:

$$y = Ax + n$$

Formulate MAP optimization problem:

$$\hat{x} = \arg\max_x p(y|x) p(x)$$

$$\hat{x} = \arg\min_x -\log (p(y|x) p(x))$$

Likelihood model under Normal distribution:

$$p(y|x) = ce^{-\frac{1}{2}(y-Ax)^T\Sigma^{-1}(y-Ax)}$$

$$\hat{x} = \arg\min_x \frac{1}{2}(y - Ax)^T\Sigma^{-1}(y - Ax) - \log p(x)$$

(assume uncorrelated noise)

$$\hat{x} = \arg\min_x \frac{1}{2}\|y - Ax\|^2 - \log p(x)$$

Many iterative solvers (for particular choices of $p(x)$)

ISTA, ADMM, etc.
One step back: Model-based optimization

**General recipe:** construct a model-based optimization algorithm based on a-priori knowledge of the measurement process and statistical priors

Define measurement model with image of interest $x$:

$$y = Ax + n$$

$$\hat{x} = \arg\min_x \frac{1}{2} \left\| y - Ax \right\|_2^2 - \log p(x)$$

$$\hat{x} = \arg\min_x \frac{1}{2} \left\| y - Ax \right\|_2^2 + g_\theta(x)$$  \hspace{1cm} \text{(regularizer $g_\theta$ parameterized by $\theta$)}

**Iterative proximal gradient solvers:**

$$z = x^k - \mu \left( \nabla_x \left\| y - Ax \right\|_2^2 \right)_{x=x^k} = f_1(x^k)$$

$$x^{k+1} = \arg\min_x \frac{1}{2} \left\| z - x \right\|_2^2 + g_\theta(x) = \text{Prox}(z)$$

2-step factorized optimization

Iterative model-based algorithm with input $y$ and output $x$
One step back: Model-based optimization

**General recipe:** construct a model-based optimization algorithm based on a-priori knowledge of the measurement process and statistical priors

Define measurement model with image of interest \( x \):

\[
y = Ax + n
\]

\[
\hat{x} = \arg\min_x \frac{1}{2} ||y - Ax||^2_2 - \log p(x)
\]
One step back: Model-based optimization

**Example: sparse coding**
Many applications: denoising, compressed sensing, image reconstruction, super-resolution, ...

**Sparse coding problem**
\[ y = Ax + n \quad \text{with } x \text{ being sparse} \]

MAP problem for \( x \) :
\[ \hat{x} = \min_x \|Ax - y\|_2^2 + \lambda \|x\|_1 \]

Some intuition:
\( \ell_1 \)-ball

\( \|x\|_1 \)

\( \|Ax - y\|_2^2 \)

\( x \sim \text{Laplace} \)
One step back: Model-based optimization

Example: sparse coding
Many applications: denoising, compressed sensing, image reconstruction, super-resolution, ...

**Sparse coding problem**

\[ y = Ax + n \quad \text{with } x \text{ being sparse} \]

MAP problem for \( x \):

\[
\hat{x} = \min_x \|Ax - y\|_2^2 + \lambda \|x\|_1
\]

\[ (x \sim \text{Laplace}) \]

\[ -\log p(x) \sim \|x\|_1 \]

**Iterative shrinkage and thresholding (ISTA)**

1. Take a gradient step towards

\[
\min_x \|Ax - y\|_2^2 \quad \Rightarrow \quad x - \mu \nabla_x \|Ax - y\|_2^2
\]

2. Move intermediate solution towards prior

\[
\text{Solve:} \quad \min_x \|x - \hat{x}^{(i)}\|_2^2 + \lambda \|x\|_1 \quad S_\lambda(\hat{x}^{(i)})
\]

= soft thresholding function

Some intuition:

- Take a gradient step towards \( \min \|Ax - y\|_2^2 \)
- Move intermediate solution towards prior

\( \|x\|_1 \quad \ell_1 \text{- ball} \)

\( x_2 \quad \|Ax - y\|_2^2 \)

\( x_1 \quad \|x\|_1 \)

\( 0 \)

\( \ell_1 \text{- ball} \)
One step back: Model-based optimization

**Example: sparse coding**
Many applications: denoising, compressed sensing, image reconstruction, super-resolution, ...

**Sparse coding problem**
\[
y = Ax + n \quad \text{with } x \text{ being sparse}
\]

MAP problem for \(x\) :
\[
\hat{x} = \min_x \|Ax - y\|_2^2 + \lambda \|x\|_1
\]

**Iterative shrinkage and thresholding (ISTA)**

**Proximity operator**
Soft thresholding
\[
\min_x \|x - \hat{x}^{(i)}\|_2^2 + \lambda \|x\|_1 = S_\lambda(\hat{x}^{(i)})
\]

**Gradient update step w.r.t.**
Likelihood term
\[
x - \mu \nabla_x \|Ax - y\|_2^2
\]
**Model-based DL: Deep unfolding**

**General recipe:** take a model-based optimization algorithm, unfold it as a fixed-complexity graph, learn parameters.

![Diagram of iterative model-based algorithm](image)

*Iterative model-based algorithm with input $y$ and output $x$*

![Diagram of unfolded model-based algorithm](image)

*Unfolded model-based algorithm with input $y$ and output $x$*

---

**Iterative algorithm**

**Learned iterative algorithm**
Deep unfolding for sparse coding (LISTA)

**General recipe:** take a model-based optimization algorithm, unfold it as a fixed-complexity graph, learn parameters.

### Learned(L)ISTA

- **Deep learning with a model-based signal prior**
  - Learn the weight matrices/convolutions $W_i$
  - Learn the thresholding parameters $\lambda_i$
  - Use the prior (sparsity) and optimization structure

**Note:** Network nonlinearity (activation function) follows directly from the prior!
Optimization with “plug and play priors”

**General recipe:** take a model-based optimization algorithm, .... what if we don’t know the prior, or it is complex to describe/model?

Define measurement model with image of interest $x$:

$$y = Ax + n$$

$$\hat{x} = \text{argmin}_x \frac{1}{2} ||y - Ax||^2_2 - \log p(x)$$

$$\hat{x} = \text{argmin}_x \frac{1}{2} ||y - Ax||^2_2 + f_\theta(x)$$ (regularizer $f_\theta$ parameterized by $\theta$)

**Iterative proximal gradient methods:**

$$z = x^k - \mu \left( \nabla_x ||y - Ax||^2_2 \right)_{x=x^k} = f_1(x^k)$$

$$x^{k+1} = \text{argmin}_x \frac{1}{2} ||z - x||^2_2 + f_\theta(x) = \text{Prox}(z)$$

**Question:** what is a good general choice for $\text{Prox}(z)$ in structured signals if you don’t know $p(x)$?

**Typical structure optimization algorithm**

**A high-performant denoiser**

Meinhardt et al. ICCV 2017
Optimization with “plug and play priors”

**General recipe:** take a model-based optimization algorithm, “plug in” a high-performant denoiser (e.g. trained deep neural network) as the Prox

Define measurement model with image of interest $x$:

$$ y = Ax + n $$

$$ \hat{x} = \arg\min_x \frac{1}{2} ||y - Ax||^2_2 - \log p(x) $$

Typical structure optimization algorithm

- Factorized into prior step and data consistency step (=data likelihood)
- Any properly trained denoiser “Plugged in” (denoising priors)

* Properly trained: under a certain Lipschitz condition (Ryu et al., 2019)

Meinhardt et al. ICCV 2017
Deep unfolding & end-to-end training

**General recipe:** take a model-based optimization algorithm, unfold it as a fixed-complexity graph, learn proximal parameters.

Define measurement model with image of interest $x$:

\[ y = Ax + n \]

\[ \hat{x} = \arg\min_x \frac{1}{2} \|y - Ax\|_2^2 - \log p(x) \]

Typical structure optimization algorithm

"Unfold" K iterations

Networks can have independent parameters

Mardani et al. NeurIPS, 2018
Deep unfolding & end-to-end training

**General recipe:** take a model-based optimization algorithm, unfold it as a fixed-complexity graph, learn all parameters.

Define measurement model with image of interest $x$:

$$y = Ax + n$$

$$\hat{x} = \arg\min_x \frac{1}{2} ||y - Ax||_2^2 - \log p(x)$$

![Diagram](image)

“Unfold” $K$ iterations

Typical structure optimization algorithm

Networks can have independent parameters

Mardani et al. NeurIPS, 2018

Gradient update matrices $W_1, W_2$ also learned
Combining models, priors and deep learning for image reconstruction: **use-cases**

- **Magnetic Resonance Imaging** (faster, low-field etc.)
- **Computed Tomography** (low-dose, sparse-view)
- **Ultrasound** (high-quality, reduced data-rates)
Combining models, priors and deep learning for image reconstruction: use-cases

Ultrasound (high-quality, reduced data-rates)
Opportunities for ultrasound

Transmission

Channel Data

Hundreds to thousands of channels/firing

High rate ADC (~20-50MHz / element)

Digital Beamforming

High-data rates

Spatial filtering

Ultrasonic localization microscopy
Shear wave imaging
Doppler/strain
Contrast agents

Advanced signal processing applications

Deep Learning in Ultrasound Imaging

Deep learning is taking an ever more prominent role in medical imaging. This article discusses applications of this powerful approach in ultrasound imaging systems along with domain-specific opportunities and challenges.

By Ruud J. G. van Sloun, Member IEEE, Reuven Cohen, Graduate Student Member IEEE, and Yossi C. Eldar, Fellow IEEE

Van Sloun, Cohen, Eldar, Proceedings of the IEEE, 2019
Adaptive beamforming by MVDR

Geometry-based time-space migration + model-based adaptive apodization

Slow ($O(N^3)$), instable matrix inversions, relies on accurate estimates of statistics (model knowledge)

Mathematical expression: $\frac{R^{-1}a}{a^H R^{-1}a}$
Adaptive beamforming by deep learning (ABLE)

Hybrid inference: Model-based computational graph with integrated NN

Fast & stable, learns actionable statistics & function from data
Adaptive beamforming by deep learning (ABLE)

Delay-and-sum (standard)  Deep learning (ABLE)  Target (MVDR)

Note: without post-processing and s-curve
60 dB, log-scale
Adaptive beamforming by deep learning (ABLE)

- Less clutter
- Higher resolution
- High processing rates, and robustness

Note: without post-processing and s-curve
60 dB, log-scale
High-resolution plane-wave compounding

Unobserved High-res object

Acquisition model

Measured input

Model-based layers

\[ [A_1, A_2, A_3]^T \]

Output

Solve inverse problem:
Deep unfolding with end-to-end training
High-resolution plane-wave compounding

Unobserved High-res object

Acquisition model

Networks have independent parameters

Acquisition model

Gradient update matrices $W_1, W_2$ also learned
High-resolution plane-wave compounding
Spatiotemporal source-extraction/dehazing

Acquisition model

Independent signals

Measured mixed input

Model-based layers

Output

Solve inverse problem:
Deep unfolding with end-to-end training
Model-based layers

Optimization problem:
\[
\min_{L, S} \frac{1}{2} \|D - L - S\|_F^2 + \lambda_1 \|L\|_* + \lambda_2 \|S\|_{1,2}
\]

Prox-grad solution: (iterative)

\[
L^{k+1} = \Sigma \Lambda_{1/2} \left( \frac{1}{2} L^k - S^k + D \right)
\]

\[
S^{k+1} = T_{\lambda/2} \left( \frac{1}{2} S^k - L^k + D \right)
\]
Spatiotemporal source-extraction/dehazing

**Optimization problem:**
\[
\min_{L,S} \frac{1}{2} \|D - L - S\|_F^2 + \lambda_1 \|L\|_1 + \lambda_2 \|S\|_{1,2}
\]

**Prox-grad solution:** (iterative)
\[
L^{k+1} = S \nabla T_{\lambda_1/2} \left( \frac{1}{2} L^k - S^k + D \right)
\]
\[
S^{k+1} = T_{\lambda_2/2} \left( \frac{1}{2} S^k - L^k + D \right)
\]

**ISTA for RPCA**

**Deep convolutional Robust PCA**
Standard contrast-ultrasound
Super-resolution contrast ultrasound by LISTA

Van Sloun et al., Proceedings of the IEEE, 2020
Combining models, priors and deep learning for image reconstruction: use-cases

Magnetic Resonance Imaging (faster, low-field etc.)
MRI: learning acquisition

Van Gorp et al., in review

Huijben et al., ICASSP 2020

Huijben et al., ICLR 2020
Importance of model-based DL for recon:

- Factorizes knowledge
- Sampling and recon update directions ‘decoupled’ during learning

\[ f(\cdot) \]

Acquisition/sampling changes directly update \( f(\cdot) \)
Huijben et al., ICASSP 2020

Huijben et al., ICLR 2020

**MRI: learning sampling density mask**

- **Full k-space**
- **Learned mask**
- **Full sampling**
- **Sampling Low frequencies**
- **Learned sampling**

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full k-space</td>
<td>36.2</td>
</tr>
<tr>
<td>Learned mask</td>
<td></td>
</tr>
<tr>
<td>Sampling Low frequencies</td>
<td>35.8</td>
</tr>
<tr>
<td>Learned sampling</td>
<td>36.2</td>
</tr>
</tbody>
</table>
MRI: learning active acquisition

### Table

<table>
<thead>
<tr>
<th>Method</th>
<th>NMSE</th>
<th>PSNR [dB]</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al., 2019 (active)</td>
<td>0.0398</td>
<td>28.8</td>
<td>0.610</td>
</tr>
<tr>
<td>Pineda et al., 2020 (active)</td>
<td>0.0371</td>
<td>29.2</td>
<td>0.623</td>
</tr>
<tr>
<td>Fixed learned sampling (ours)</td>
<td>0.0360</td>
<td>30.1</td>
<td>0.650</td>
</tr>
<tr>
<td>Active acquisition (ours)</td>
<td>0.0342</td>
<td>30.2</td>
<td>0.654</td>
</tr>
</tbody>
</table>

Factor 8 undersampling
Conclusions

- Acquisition model assumptions incorrect/imprecise
- Statistical image priors not sufficiently expressive/accurate
- Slow reconstruction with high complexity

Deep Learning
- Complex statistical models
- Learns from data

Model-based algorithms
- Domain knowledge
- Sensor physics
- Inductive biases
- Theoretical guarantees

Networks have independent parameters

Gradient update matrices $W_1, W_2$ also learned
Thanks!