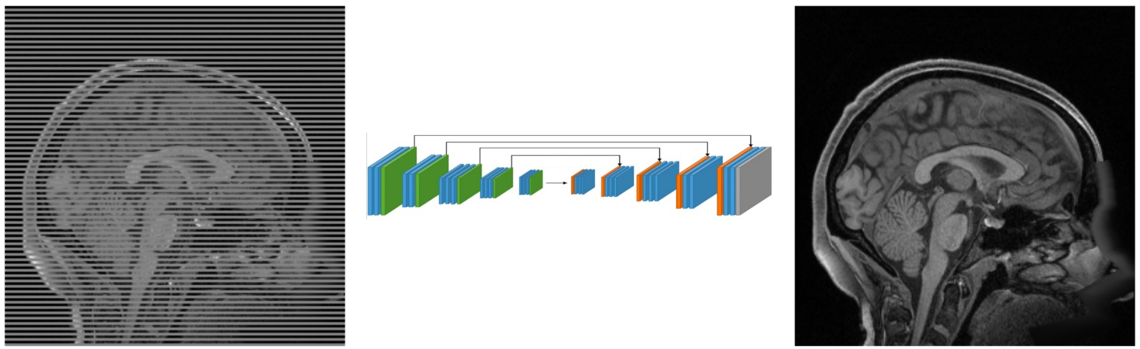
Medical Imaging Workshop @ NeurIPS
16th Dec 2023

Convolve and Conquer:

Data Comparison with Wiener Filters

Deborah Pelacani Cruz*
George Strong*, Oscar Bates, Carlos Cueto, Jiashun Yao, Lluis Guasch



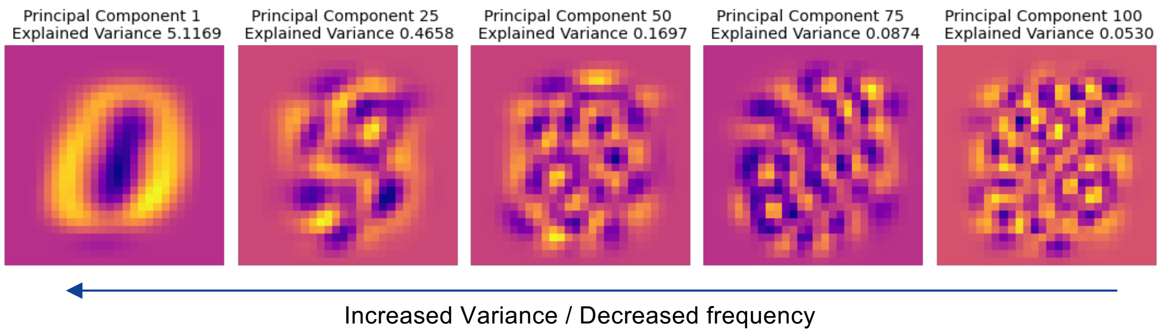
* Equal contributors to this work

Motivation and Aims

- Reconstruction problems often performed under **pixel-wise L2 loss**
  - Inexpensive, differentiable and smooth solution space

$$f = \frac{1}{2}||\boldsymbol{p} - \boldsymbol{x}||^2$$

- Assumption that all data points are **independent and of equal importance → Limited contextual awareness**
- Variance minimisation **focuses on low frequencies**→ often leads to **averaged/blurry results**



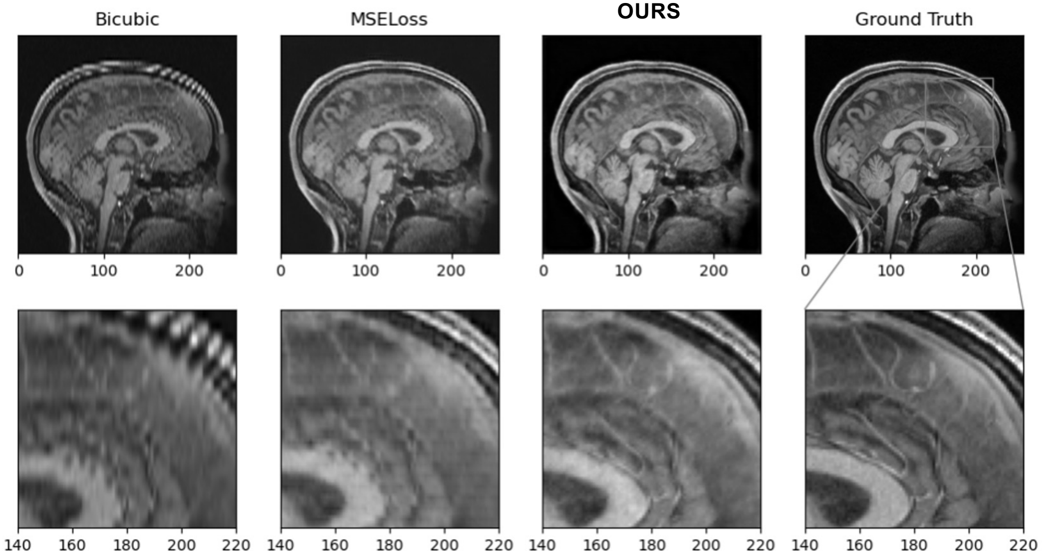
e.g. PCA Analysis of error between MNIST and random noise

Motivation and Aims

Ongoing work in the machine learning community on perceptual quality	Our **Wiener Metric** (from Wiener-filter theory)
• Feature-wise losses, adversarial losses, focal frequency loss, perceptual losses, etc • Expensive • Unstable • Data Biases	• **Convolution-based** metric aimed at promoting full-spectrum data recovery • Does not assume local element-wise relationships • Stable with differentiable and smooth solution space • Inexpensive • No data biases

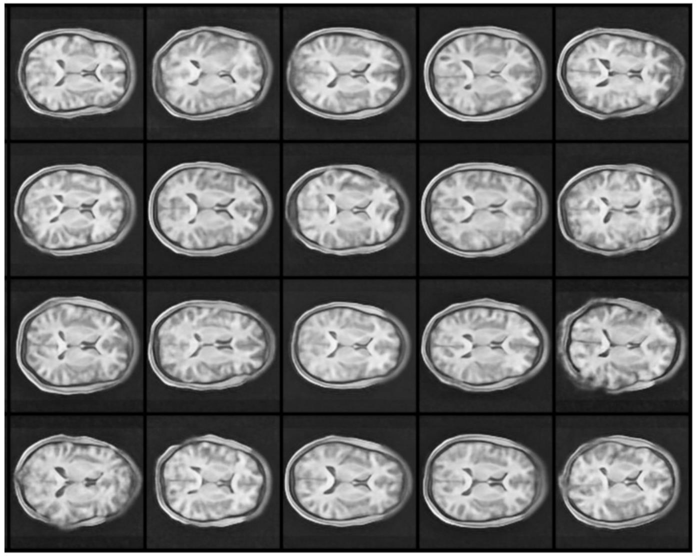
Motivation and Aims

Wiener Loss

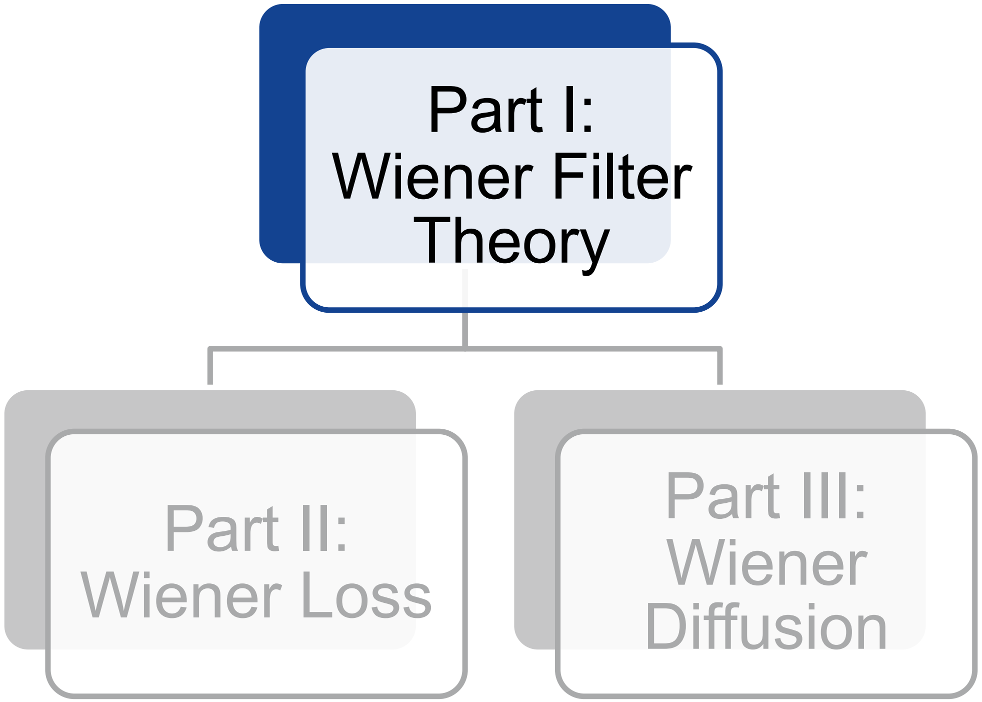


MRI imputation results on heavily corrupted data

Wiener Diffusion



Batch of generated MRI samples using the Wiener Diffusion
(no training required)


Part I:
Wiener Filter
Theory
Part II:
Wiener Loss
Part III:
Wiener
Diffusion

Wiener Filters: Principles

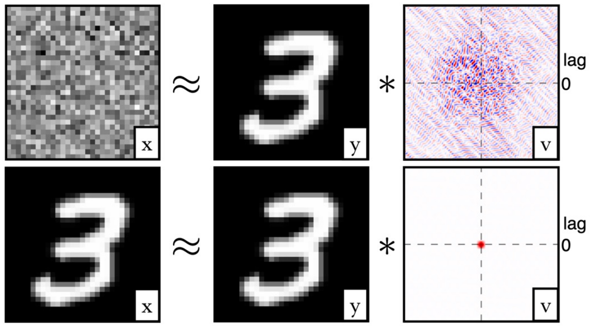
- Any signal x can be constructed through a signal y via a **convolutional matching filter** v :

$$\mathbf{x} = \mathbf{y} * \mathbf{v}$$

- Reformulate y as a Toeplitz matrix Y to achieve convolution in matrix form:

$$\mathbf{x} = \mathbf{Y}\mathbf{v}$$

- A convolutional filter that is a **Dirac delta function at zero lag** (convolutional identity) leaves the input signal y unchanged



Wiener Filters: Principles

$$g = \frac{1}{2}||\mathbf{Yv} - \mathbf{x}||^2$$

$$\frac{dg}{d\mathbf{v}} = \mathbf{Y}^T\mathbf{Y}\mathbf{v} - \mathbf{Y}^T\mathbf{x}$$

Data domain

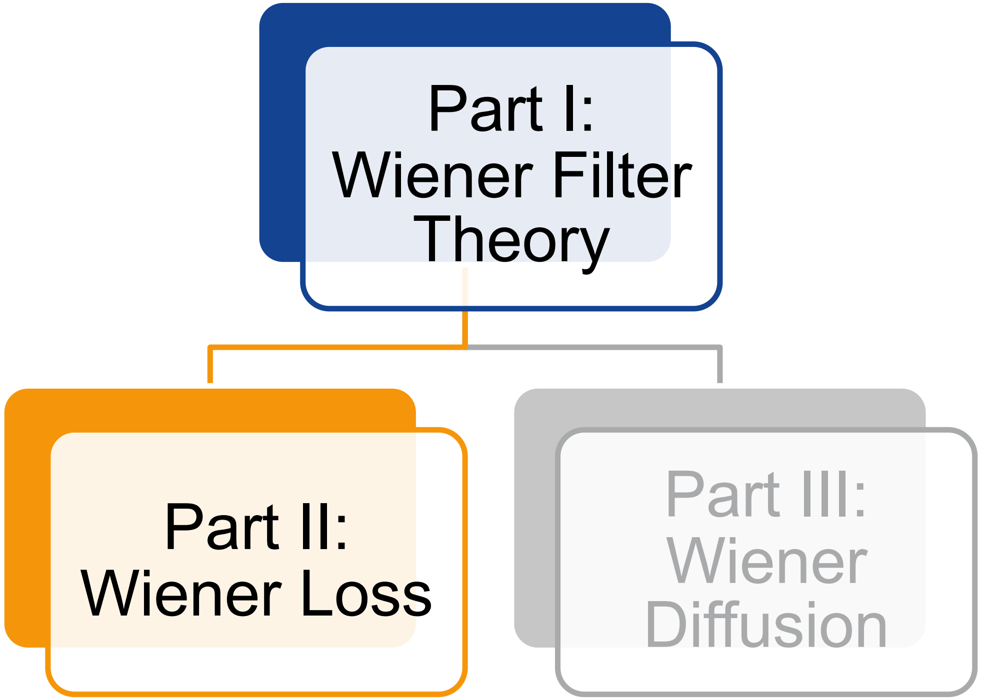
$$\mathbf{v}(\mathbf{x},\mathbf{y}) = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{x}$$

Frequency domain

$$\mathbf{v}(\mathbf{x},\mathbf{y}) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{y})\,\mathcal{F}^*(\mathbf{x})}{\mathcal{F}(\mathbf{y})\,\mathcal{F}(\mathbf{y})}\right)$$

- Wiener filter computes the **convolutional** matching filter $\boldsymbol{v}$ on a statistical approach that **minimises the mean-squared error between the target** $(\boldsymbol{y})$ **and reconstructed signal** $(\boldsymbol{x})$
- The coefficients of $\boldsymbol{v}$ are the ones that minimise the functional g (smooth and differentiable)
- Interpret as: cross-correlation of the target and reconstructed data, deconvolved by the autocorrelation of the target data (Warner and Guasch, 2016)

Michael Warner and Lluís Guasch. Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445, 2016.


Part I:
Wiener Filter
Theory
Part II:
Wiener Loss
Part III:
Wiener
Diffusion

Wiener Loss: Formulation

- Inspired by the work of Warner and Guasch (2016) on adaptive waveform inversion
- Construct an objective whose minimisation **maximises the likelihood of the zeroth-lagged Dirac delta function (convolutional identity) under a multivariate Gaussian distribution with mean defined as the data-matching Wiener filters**

$$p(\boldsymbol{\delta}|\mathbf{v}) = \prod_{\mathrm{i}=1} \mathcal{N}\left(\boldsymbol{\delta} \mid \mathbf{v}(\mathbf{x}_{\boldsymbol{\theta}}^{(\mathrm{i})}, \mathbf{y}^{(\mathrm{i})}), \boldsymbol{\Sigma}\right)$$

$$-\log p(\boldsymbol{\delta}|\mathbf{v}) \propto \sum_{\mathrm{i}=1} \frac{1}{2}\left[\left\{\mathbf{v}(\mathbf{x}_{\boldsymbol{\theta}}^{(\mathrm{i})}, \mathbf{y}^{(\mathrm{i})}) - \boldsymbol{\delta}\right\}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left\{\mathbf{v}(\mathbf{x}_{\boldsymbol{\theta}}^{(\mathrm{i})}, \mathbf{y}^{(\mathrm{i})}) - \boldsymbol{\delta}\right\}\right]$$

Michael Warner and Lluís Guasch. Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445, 2016.

Wiener Loss: Formulation

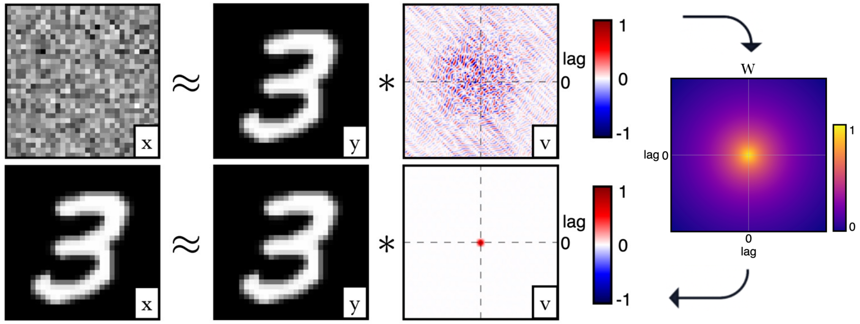
- $\boldsymbol{\Sigma}^{-1}$ is the covariance matrix decomposed as $\boldsymbol{W}^T\boldsymbol{W}$ where $\boldsymbol{W}$ is an hyperparameter monotonic function that penalises non-zero lag energy in the matching filter

$$-\log p(\boldsymbol{\delta}|\mathbf{v}) \propto \sum_{\mathrm{i}=1} \underbrace{\frac{1}{2}\left\|\mathbf{W}\left\{\mathbf{v}\left(\mathbf{x}_{\boldsymbol{\theta}}^{(\mathrm{i})}, \mathbf{y}^{(\mathrm{i})}\right) - \boldsymbol{\delta}\right\}\right\|^2}_{\text{Mahalanobis distance}}$$

$$\mathcal{L}_{\mathrm{W}}(\mathbf{x}_{\boldsymbol{\theta}}, \mathbf{y}) = \frac{1}{2}\|\mathbf{W}\{\mathbf{v}(\mathbf{x}_{\boldsymbol{\theta}}, \mathbf{y}) - \boldsymbol{\delta}\}\|^2$$

Wiener Loss: Formulation

- In other words: train the network to learn the reconstruction of a target by **implicitly driving the corresponding Wiener filter toward a convolutional identity**, i.e. delta function at zero lag

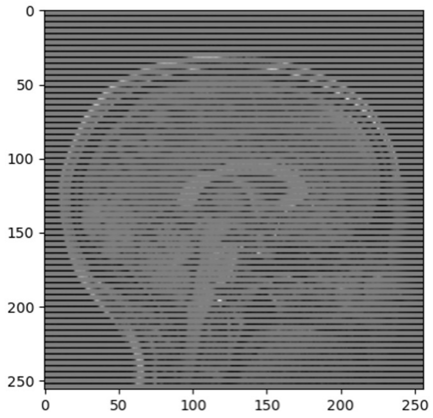


$$\mathcal{L}_W(\mathbf{x}_\theta, \mathbf{y}) = \frac{1}{2}\|\mathbf{W}\{\mathbf{v}(\mathbf{x}_\theta, \mathbf{y}) - \boldsymbol{\delta}\}\|^2$$

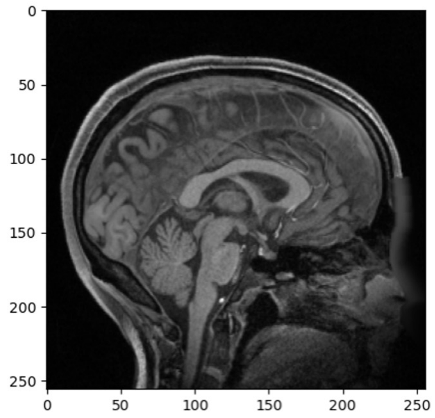
- Convolutional filters incorporates **conceptual awareness** lacking in pixel-wise norms
- Least-squares provides **convex optimisation** spaces
- **General data comparison method** that is perceptually aware

Wiener Loss: MRI Supervised Imputation

- MRI T1-weighted 2D samples in the midsagittal plane
- **553 samples** of size *1x256x256* from HPC Young Adult Database
- **Undersampling mask** applied for input data: 3 pixels in width, 1 pixel in spacing
- **UNet** of 3 channels, 3 residual blocks and *Mish* activation function
- Adam optimiser, learning rate 1e^{-2}, 500 epochs



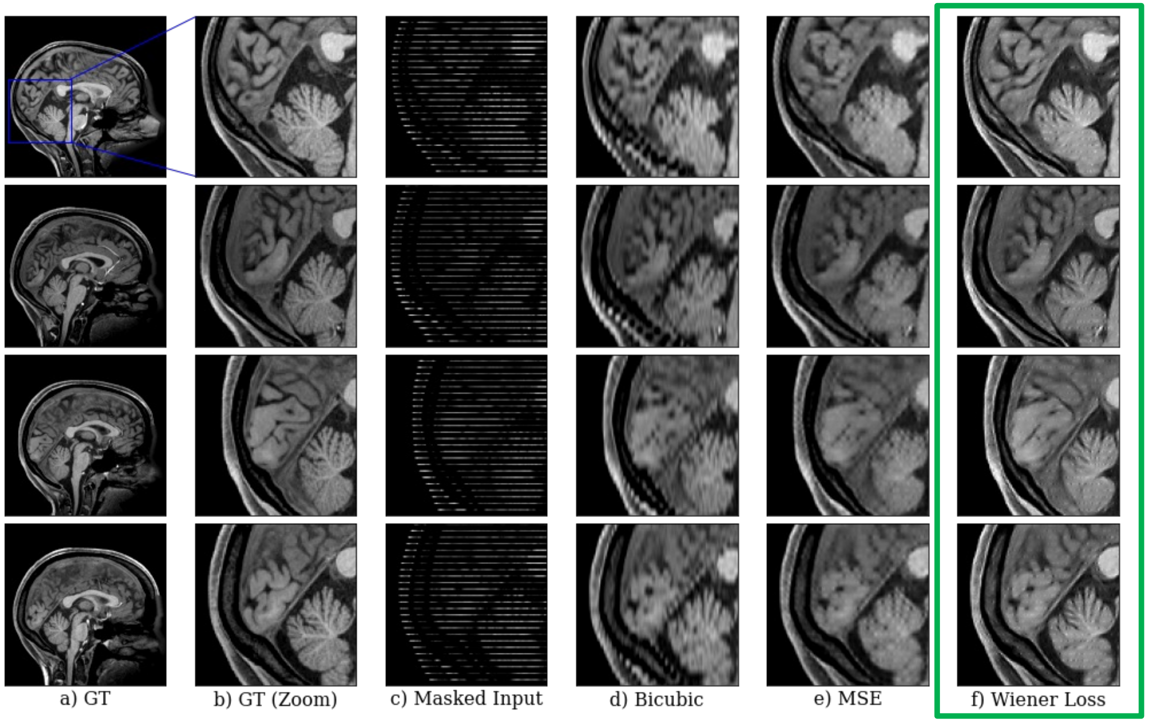
Masked Input



Target

Van Essen et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

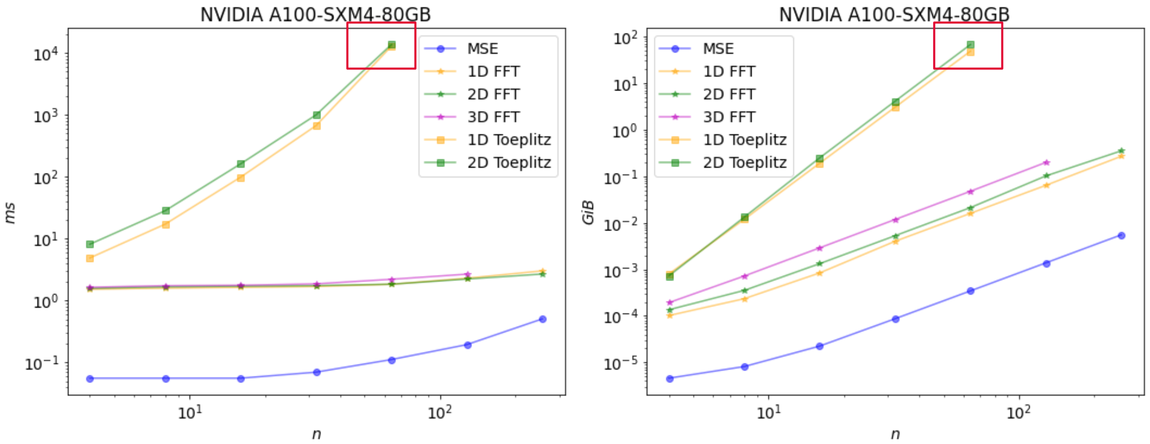
Wiener Loss: MRI Supervised Imputation



	Bicubic	MSE	Wiener Loss
↓ MAE	$4.09\ e^{-4}$	$1.93\ e^{-4}$	$2.22\ e^{-4}$
↓ MSE	$2.60\ e^{-4}$	$1.35\ e^{-4}$	$1.71\ e^{-4}$
↓ Wiener Loss	0.21	0.14	$9.16\ e^{-2}$
↑ SSIM	0.80	0.91	0.90
↓ **LPIPS**	**0.32**	**0.11**	**$7.93\ e^{-2}$**
↓ **FID**	**4.12**	**1.66**	**0.23**

- Superior recovery of finer structures of the scan, aliased or blurred in other methods
- Improved statistical representation

Computation and Memory Complexity



Complexity graphs of (a) computational time and (b) memory usage; tested using torch.profile with images of size 3 x n x n

- Wiener Loss 1D: flattening data in all dimensions, filter size ***C*H*W***
- Wiener Loss 2D: one filter per channel, filter of size ***1 x H x W***
- Wiener Loss 3D: filter size ***C x H x W***
- **Complexity of FFT implementation comparable to MSE**, ~ one order of magnitude more expensive

Toeplitz

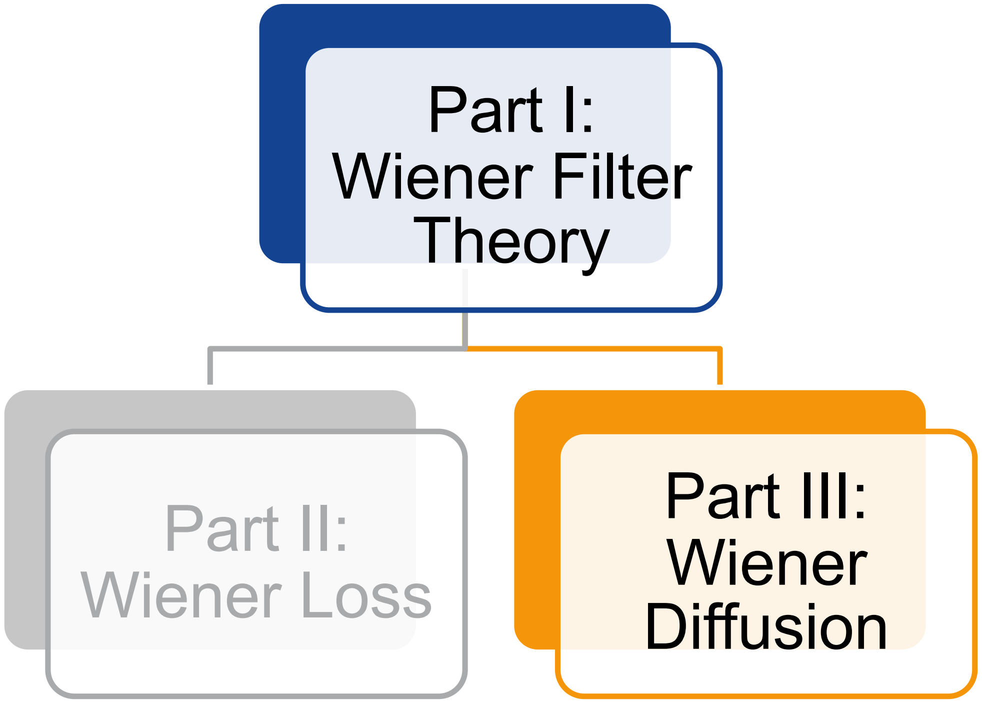
$$\mathbf{v} = (\mathbf{D}^{\mathbf{T}}\mathbf{D})^{-1}\mathbf{D}^{\mathbf{T}}\mathbf{x}$$

FFT

$$\mathbf{v}(\mathbf{x},\mathbf{y}) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{y})\,\mathcal{F}^{*}(\mathbf{x})}{\mathcal{F}(\mathbf{y})\,\mathcal{F}(\mathbf{y})}\right)$$

$\boldsymbol{O}(\log \boldsymbol{n})$ - computational

$\boldsymbol{O}(\boldsymbol{n}^2)$ - memory

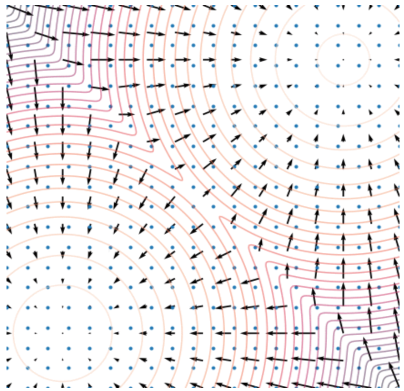

Part I:
Wiener Filter
Theory
Part II:
Wiener Loss
Part III:
Wiener
Diffusion

Wiener Diffusion: Energy-based models recap

$$p(\mathbf{x}) = \frac{\exp(-\mathrm{E}(\mathbf{x}))}{\int \exp\big(-\mathrm{E}(\mathbf{x})\big)\,\mathrm{d}\mathbf{x}}$$

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \frac{\alpha_t}{2} \nabla_{\mathbf{x}} \mathrm{E}(\mathbf{x}_t) + \mathbf{z}_t, \text{ for } t = 0, 1, \ldots, T-1$$

- The probability density of an energy-based model (EBM) is given by the **Boltzman distribution**
- Probability likelihood **computationally intractable** due to normalisation term in the denominator
- **Langevin dynamics** can sample $p(\boldsymbol{x})$ through the **gradient of the energy function $\boldsymbol{E}$** (score), where the normalisation term vanishes



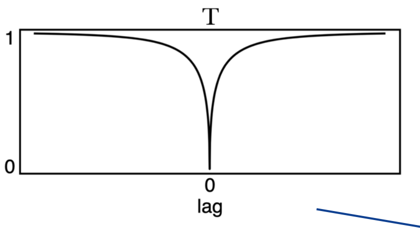
Using Langevin dynamics to sample from a mixture of two Gaussians. Score function ∇E (the vector field) and density function p (contours)

Yang Song and Diederik P Kingma. How to train your energy-based models. *arXiv preprint arXiv:2101.03288*, 2021.
Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." *arXiv preprint arXiv:2011.13456* (2020).

Wiener Diffusion: Formulation

- While training an EBM usually requires the estimation of the energy or score functions, we define an energy function that **non-parametrically** drives the matching Wiener filters of the sample $\boldsymbol{x}$ undergoing diffusion to all samples $\boldsymbol{y}$ of the dataset towards a delta function at zero lag:

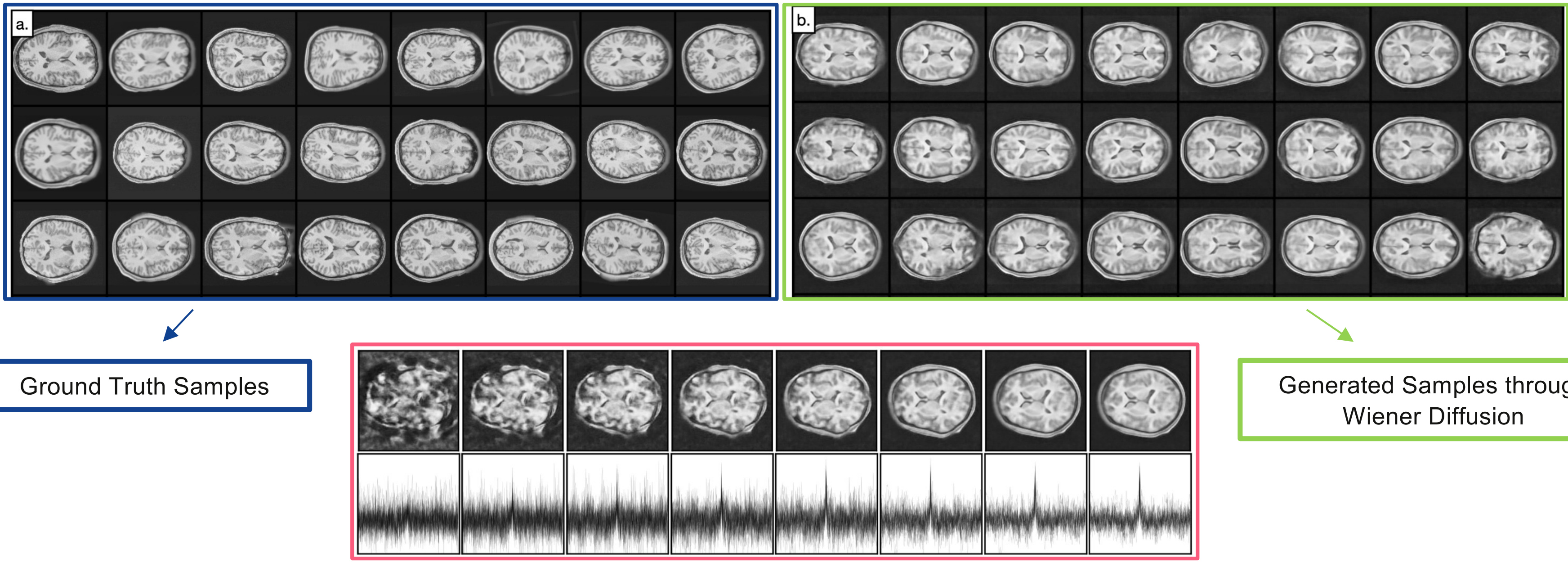
$$\mathrm{E}(\mathbf{x}) = \sum_{\mathrm{i}=1} \frac{1}{2} \frac{||\mathbf{T}\mathbf{v}(\mathbf{x}, \mathbf{y}_{\mathrm{i}})||^2}{||\mathbf{v}(\mathbf{x}, \mathbf{y}_{\mathrm{i}})||^2} + \frac{\gamma}{2} ||\boldsymbol{\delta} \otimes (\mathbf{v}(\mathbf{x}, \mathbf{y}_{\mathrm{i}}) - \boldsymbol{\delta})||^2$$



Drives the data-matching Wiener filters towards zero-lag spikes by **minimising the Rayleigh quotient** of a hyper-parameter penalty matrix **T**

Ensures the energy-based model is **sensitive to amplitude** information by encouraging the zeroth-lag of the filters to have an amplitude of 1.

Wiener Diffusion: MRI Generative Modelling



Ground Truth Samples

Generated Samples through Wiener Diffusion

Example of the sampling process through Langevin Dynamics (top) and 30 closest Wiener filters to the sample (bottom)

Wiener Diffusion: Considerations

- Well suited to scenarios with limited data → possibilities for **inexpensive expansion of medical datasets**
- Can also serve as a **prior generator or as a regularisation term** in imaging workflows
- Diffusion process can be **conditioned at inference time** by specifying the data that defines the energy function → **reduces biases (e.g. ethnicity, age)**
- **Sampling in latent space** for stability
- **Monotonically increasing gradient** towards zero-lag:
  - Prevents collapse to global barycenters
  - Required for sampling different distribution modes

Summary

- New method to measure (dis)similarities between paired samples inspired by Wiener-filter theory.
- Convolution-base: promote preservation of contextual information
- Inexpensive, scalable and differentiable
- Novel objective function: **Wiener Loss**
- Novel non-parametric energy-based model: **Wiener Diffusion**
- Readily suited for a wide a range of machine learning and inverse problems, inc. imputation, regularisation, dataset expansion…

$$\mathcal{L}_{\mathrm{w}}(\mathbf{x}_{\theta}, \mathbf{y}) = \frac{1}{2} \|\mathbf{W}\{\mathbf{v}(\mathbf{x}_{\theta}, \mathbf{y}) - \boldsymbol{\delta}\}\|^2$$

$$\mathrm{E}(\mathbf{x}) = \sum_{\mathrm{i}=1} \frac{1}{2} \frac{\left|\left|\mathbf{T}\mathbf{v}(\mathbf{x}_{\theta}, \mathbf{y}_{\mathrm{i}})\right|\right|^2}{\left|\left|\mathbf{v}(\mathbf{x}_{\theta}, \mathbf{y}_{\mathrm{i}})\right|\right|^2} + \frac{\gamma}{2} \left|\left|\boldsymbol{\delta} \otimes (\mathbf{v}(\mathbf{x}_{\theta}, \mathbf{y}_{\mathrm{i}}) - \boldsymbol{\delta})\right|\right|^{\mathbf{2}}$$

Ethical Statement

Our research demonstrates that our new methodology for data comparison can statistically enhance images from limited-resolution scans and generate new samples from the same distribution. Despite these advantages, we highlight an ethical concern about potential biases stemming mainly from exclusive statistical representation of the training data and network architecture. These biases pose a significant risk in identifying misrepresented pathologies and can mislead clinical analyses. In this work, we used an MRI dataset of healthy adult brains to prove and validate our concept, but we emphasise the need for further research with diverse datasets to substantiate our findings.

Acknowledgments

George for his support, proactivity and knowledge

My supervisors Lluis and Mike

WTRUST team: Oscar, Carlos, Yao

Jo and Adriana for ongoing support with my PhD