ANALYTICVARNET: COVARIANT WAVELET SCATTERING FOR ROBUST MRI RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unrolled variational networks are the dominant framework for accelerated MRI reconstruction, but their learned regularizers can overfit the sampling pattern seen in training. We replace the learned spatial operator of the regularizer with a fixed covariant wavelet scattering transform: a low-pass filter and oriented Morlet wavelets followed by a modulus, without the averaging that makes standard scattering invariant. The resulting operator tiles the full frequency plane with explicit frame bounds, and we prove it is translation-equivariant, retains every frequency band through a lower frame bound, and is non-expansive and stable to small deformations. The filterbank itself has no trainable parameters. On fastMRI knee and brain data, AnalyticVarNet improves over a neural-operator baseline in-distribution and is substantially more robust to unseen incoherent sampling masks, with $2.47\times$ fewer parameters.

1 INTRODUCTION

In this work, we replace the learned continuous operator in an unrolled MRI reconstruction network with a fixed wavelet-scattering operator. This substitution imposes an analytically defined spectral structure on the regularizer, while subsequent learned layers adapt the resulting features to the reconstruction task.

We build this operator from a wavelet scattering transform (Mallat, 2012; Bruna & Mallat, 2013): a Gaussian low-pass filter (zeroth-order coefficients) and a bank of oriented Morlet wavelets followed by a complex modulus (first-order coefficients). Together these filters tile the entire frequency plane. We keep this complete tiling but omit the averaging and subsampling that make standard scattering translation-invariant. The resulting *covariant* operator retains full spatial resolution, as reconstruction requires, and we show that it retains every frequency band through an explicit lower frame bound and remains non-expansive and stable to small deformations without the averaging step (Proposition 1). By embedding this fixed operator within an already existing variational network cascade, we decouple spectral structure from coefficient learning. This reduces the hypothesis space while retaining sufficient expressive capacity for reconstruction. Whereas neural-operator regularizers (Jatyani et al., 2025) learn the spectral representation and the reconstruction mapping jointly, and nothing constrains the learned kernels to cover the whole spectrum, our regularizer fixes a complete spectral partition analytically, with a lower frame bound that holds independently of the training data. We evaluate this structured inductive bias in-distribution and under three kinds of shift: unseen sampling masks, higher acceleration factors, and zero-shot transfer to a higher spatial resolution. AnalyticVarNet improves over the neural-operator baseline in-distribution on knee and brain data. Under mask shift, it performs comparably on masks close to the training pattern and substantially better on incoherent ones (radial, Gaussian, Poisson), both zero-shot and after finetuning on 100 slices. These results are obtained with $2.47\times$ fewer trainable parameters and 18% faster inference.

In summary, our contributions are:

- We introduce AnalyticVarNet, an analytically structured covariant scattering operator for unrolled MRI reconstruction that replaces learned spectral kernels with a fixed Morlet filterbank followed by learned layers.
- We show that the covariant scattering block tiles the full frequency plane with explicit frame bounds, and that omitting the averaging of standard scattering forgoes translation

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

invariance but not stability, i.e. the block is translation-equivariant, non-expansive, and Lipschitz-stable to small deformations (Proposition 1).

- We conduct extensive experiments on fastMRI knee and brain data showing improved in-distribution reconstruction and substantially higher robustness to unseen sampling masks than a neural-operator baseline, with $2.47\times$ fewer learnable parameters.

2 Related Work

Accelerated MRI reconstruction. Classical accelerated MRI combines parallel imaging, which exploits coil redundancy, with compressed sensing (CS), which uses sparsity priors to reconstruct undersampled k-space data (Pruessmann et al., 1999; Griswold et al., 2002; Lustig et al., 2007). Deep learning instead learns reconstruction priors from data, using architectures ranging from CNNs such as U-Net to unrolled networks that integrate learned regularization with data consistency (Ronneberger et al., 2015; Hammernik et al., 2018; Sriram et al., 2020).

Generative and diffusion-based approaches. Diffusion models formulate MRI reconstruction as posterior sampling from learned image distributions, iteratively recovering realistic images while enforcing consistency with measured k-space data (Chung & Ye, 2022; Jalal et al., 2021; Güngör et al., 2023; Wu et al., 2024).

Discretization-agnostic learning and neural operators. Neural operators learn mappings between function spaces rather than fixed-resolution tensors, enabling generalization across discretizations (Kovachki et al., 2023; Azizzadenesheli et al., 2024). In MRI, Jatyani et al. (2025) replace CNN regularizers in unrolled reconstruction with discretization-consistent operators, allowing transfer across image resolutions without retraining.

Robustness of learned reconstruction. Learned MRI reconstruction can be unstable under small perturbations and under shifts in the acquisition protocol (Antun et al., 2020). Trained networks degrade when the sampling pattern or acceleration at test time differs from training (Knoll et al., 2019), and the resulting robustness gap has been measured systematically for accelerated MRI (Darestani et al., 2021). Jatyani et al. (2025) address pattern shift by training a neural operator on multiple sampling patterns. We study the complementary setting of a single training pattern and change the inductive bias of the regularizer.

Wavelets and scattering representations. Wavelets provide multi-resolution sparsifying priors widely used in CS-MRI (Lustig et al., 2007). Wavelet scattering extends them through cascaded wavelet convolutions and modulus nonlinearities, yielding multi-scale representations stable to small deformations (Mallat, 2012; Bruna & Mallat, 2013; Andreux et al., 2020). Hybrid scattering networks (Oyallon et al., 2017) place a fixed scattering front-end before learned layers for classification, and structured receptive fields (Jacobsen et al., 2016) restrict CNN filters to learned combinations of a fixed Gaussian-derivative basis. Both use fixed filters for recognition. We use them inside the regularizer of an inverse-problem solver, at full resolution and without the invariance-inducing averaging. Wavelet-based U-Nets (Liu et al., 2018) optimize wavelet-like analysis and synthesis modules inside a standalone restoration network. DISCO convolutions (Ocampo et al., 2023), used by neural-operator approaches, instead learn a discrete-continuous convolutional representation from data. AnalyticVarNet differs from both as we insert a fixed analytic Morlet filterbank into an unrolled variational network with explicit data consistency and learns only the subsequent convolutional layers.

3 Methodology

AnalyticVarNet is an unrolled reconstruction network whose regularizer is built on a fixed covariant scattering operator. Sec. 3.1 describes the reconstruction pipeline, Sec. 3.2 the scattering operator and its properties, and Sec. 3.3 the regularizer architecture. Algorithm 1 summarizes the full forward pass.

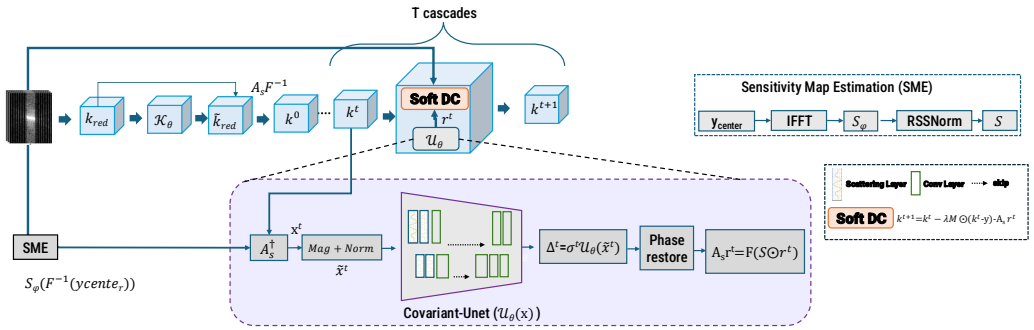


Figure 1: Overview of AnalyticVarNet. Coil-sensitivity maps are estimated from the central calibration region, and a learned k-space module initializes the multicoil estimate. Each of the T cascades applies sensitivity reduction, magnitude normalization, covariant scattering U-Net regularization, phase restoration, sensitivity encoding, and learned soft data consistency. After the final cascade, the reconstructed magnitude image is obtained by inverse Fourier transformation and root-sum-of-squares coil combination.

3.1 Unrolled Reconstruction of Undersampled MRI

Problem formulation. We consider accelerated multi-coil MRI reconstruction from undersampled k-space measurements. Let $y \in \mathbb{C}^{C\times H\times W}$ denote the k-space data acquired by C receiver coils, $M \in \{0,1\}^{H\times W}$ the binary undersampling mask, $x \in \mathbb{C}^{H\times W}$ the unknown image, and $S = \{S_c\}_{c=1}^{C}$, $S_c \in \mathbb{C}^{H\times W}$, the coil sensitivity maps. The forward model is

$$y_c = M \odot \mathcal{F}(S_c \odot x) + \epsilon_c, \qquad c = 1, \dots, C, \tag{1}$$

where $\mathcal{F}$ is the two-dimensional Fourier transform, $\odot$ denotes element-wise multiplication, and ϵ_c is measurement noise. Reconstruction is posed as the inverse problem

$$\min_x \ \|M \odot \mathcal{F}(S \odot x) - y\|_2^2 + \mathcal{R}(x), \tag{2}$$

where the first term enforces consistency with the acquired samples and $\mathcal{R}$ is a regularizer that promotes plausible image structure.

Unrolled cascade. We approximate the solution of equation 2 with T unrolled cascades, each consisting of a learned regularization step followed by data consistency (Figure 1). Let $A_S x = \mathcal{F}(S \odot x)$ denote sensitivity encoding and $A_S^\dagger k = \sum_c \overline{S_c} \odot \mathcal{F}^{-1}(k_c)$ sensitivity reduction, where $\overline{S_c}$ is the complex conjugate. With $k^t \in \mathbb{C}^{C\times H\times W}$ the multi-coil k-space estimate at cascade t, each cascade computes

$$k^{t+1} = k^t - \lambda_t M \odot (k^t - y) - A_S r^t, \tag{3}$$

where λ_t is a learned data-consistency weight and r^t is the phase-restored residual predicted by the regularizer.

Magnitude-domain regularization. The regularizer operates on the normalized magnitude of the reduced image $x^t = A_S^\dagger k^t$,

$$\Delta^t = \sigma^t \mathcal{U}_\theta(\tilde{x}^t), \qquad r^t = \Delta^t \odot \frac{x^t}{|x^t| + 10^{-8}}, \qquad \tilde{x}^t = (|x^t| - \mu^t)/\sigma^t, \tag{4}$$

where μ^t and σ^t are the spatial mean and standard deviation of $|x^t|$ and $\mathcal{U}_\theta$ is the regularizer network of Sec. 3.3. This design reflects the assumption that anatomical structure resides primarily in magnitude variations. The real-valued magnitude residual is converted to a complex residual using the phase of the current reduced image and then expanded through A_S. After T cascades, the output is the root-sum-of-squares image $\hat{x} = \left(\sum_c |\mathcal{F}^{-1}(k_c^T)|^2\right)^{1/2}$.

Algorithm 1 AnalyticVarNet: Covariant Wavelet Scattering Reconstruction

Require: Multi-coil k-space y, mask M, cascades T
Ensure: Reconstructed magnitude image $\hat{x}$

$S \leftarrow \mathrm{RSSNorm}\big(\mathcal{S}_\phi\big(\mathcal{F}^{-1}(y_{\mathrm{center}})\big)\big)$
$k_{\mathrm{red}} \leftarrow \mathcal{F}\big(\sum_c \overline{S_c} \odot \mathcal{F}^{-1}(y_c)\big), \quad k_{\mathrm{red}} \leftarrow k_{\mathrm{red}} + \mathcal{K}_\theta(k_{\mathrm{red}})$
$k^0 \leftarrow \mathcal{F}\big(S \odot \mathcal{F}^{-1}(k_{\mathrm{red}})\big)$
for $t = 0$ to $T-1$ **do**
    $x^t \leftarrow \sum_c \overline{S_c} \odot \mathcal{F}^{-1}(k_c^t), \quad \tilde{x}^t \leftarrow (|x^t| - \mu^t)/\sigma^t$
    $\Delta^t \leftarrow \sigma^t \mathcal{U}_\theta(\tilde{x}^t), \quad r^t \leftarrow \Delta^t \odot x^t/(|x^t| + 10^{-8})$
    $k^{t+1} \leftarrow k^t - \lambda_t\, M \odot (k^t - y) - \mathcal{F}(S \odot r^t)$
end for
$\hat{x} \leftarrow \mathrm{RSS}\big(\mathcal{F}^{-1}(k^T)\big)$
return $\hat{x}$

Initialization and sensitivity estimation. Following Jatyani et al. (2025), the cascade is initialized by a learned correction of the reduced k-space $k_{\mathrm{red}} = \mathcal{F}(A_S^\dagger y)$,

$$k^0 = A_S\, \mathcal{F}^{-1}\big(k_{\mathrm{red}} + \mathcal{K}_\theta(k_{\mathrm{red}})\big), \tag{5}$$

where $\mathcal{K}_\theta$ has the architecture of Sec. 3.3. The sensitivity maps are estimated from the calibration region by a separate scattering U-Net $\mathcal{S}_\phi$, which processes real and imaginary channels separately. Its output is normalized by root-sum-of-squares across coils, so that $\sum_c |S_c|^2 \approx 1$ at each location. As in Sriram et al. (2020), ϕ is trained end-to-end through the reconstruction loss without auxiliary sensitivity targets. Details are given in Appendix B.

3.2 Covariant Scattering Operator

Wavelet scattering representation. We build our operator on the Morlet wavelet filterbank from Kymatio (Andreux et al., 2020; Oyallon, 2017). For scale $j \in \{0, \dots, J-1\}$ and orientation $\theta \in \{k\pi/L\}_{k=0}^{L-1}$, the complex Morlet wavelet is

$$\psi_{j,\theta}(u) = 2^{-2j}\psi\big(2^{-j} r_{-\theta}(u)\big), \qquad \psi(u) = \big(e^{i\xi\cdot u} - \beta\big)\, e^{-|u|^2/2\sigma^2}, \tag{6}$$

where $r_{-\theta}$ is a rotation by $-\theta$, ξ the center frequency, σ the envelope width, and β a small correction that ensures zero mean. The first-order coefficients $U[j,\theta](x) = |x * \psi_{j,\theta}|$ are the moduli of the wavelet responses. Together with the zeroth-order coefficient $x * \phi_J$, where ϕ_J is the Gaussian low-pass filter at scale 2^J, they form the scattering representation

$$U(x) = \Big\{\, x * \phi_J,\ U[j,\theta](x) \,\Big\}_{j,\theta} \in \mathbb{R}^{B\times(C\cdot K)\times H\times W}, \tag{7}$$

for an input $x \in \mathbb{R}^{B\times C\times H\times W}$, with $K = 1 + JL$ channels per input channel.

Frequency tiling. The low-pass $\hat{\phi}_J$ covers the low frequencies and the oriented wavelets cover the remaining annuli, so the filterbank tiles the frequency plane. For real inputs, Hermitian symmetry makes orientations in $[0, \pi)$ sufficient, and the coverage is measured by the Littlewood–Paley sum

$$\Lambda(\omega) = |\hat{\phi}_J(\omega)|^2 + \tfrac{1}{2}\sum_{j,\theta}\big(|\hat{\psi}_{j,\theta}(\omega)|^2 + |\hat{\psi}_{j,\theta}(-\omega)|^2\big). \tag{8}$$

We use $J = 1$ and $L = 8$, for which $B \le \Lambda(\omega) \le A$ with $A \approx 1.00$ and $B \approx 0.12$ at any image size. The minimum occurs only at the corners of the Nyquist square, and $\Lambda(\omega) \ge 0.68$ for $|\omega| \le \pi/2$, where most of the energy of MR images lies. Increasing J subdivides the low-frequency band into further oriented sub-bands and leaves the frame bounds unchanged for $J = 1, 2, 3$. At the 320×320 resolution of fastMRI this brings no measurable gain (Appendix F), so we use $J = 1$. We expect finer low-frequency partitioning to become more useful at higher native resolutions, where the low band contains more resolvable anatomical structure.

Covariant formulation. Standard scattering networks average the first-order coefficients with ϕ_J and subsample by 2^J, which makes the representation translation-invariant and removes the spatial localization needed for reconstruction (Mallat, 2012). We omit this averaging and all subsampling, and keep the zeroth-order coefficient, so the representation stays at full resolution and retains the tiling of equation 8. We call this formulation covariant, since it commutes with translations and, up to a bounded error, with small deformations. Proposition 1 makes these properties precise, and Appendix G compares the covariant and invariant formulations experimentally.

Proposition 1 (Frequency coverage and deformation covariance)**.** *Let ψ be a Morlet wavelet with zero mean such that ψ and $\nabla\psi$ have fast decay, let U be the representation equation 7, and let $0 < B \le \Lambda(\omega) \le A$ with Λ as in equation 8. Let $L_\tau x(u) = x(u - \tau(u))$ with $\|\nabla\tau\|_\infty \le 1/2$. Then, for real-valued x, x':*

1. *(Equivariance) If $\tau \equiv c$ is constant, $UL_\tau = L_\tau U$ (on the grid, for integer shifts with periodic convolution).*
2. *(Frame bounds) $B\,\|x\|^2 \le \|Ux\|^2 \le A\,\|x\|^2$.*
3. *(Non-expansiveness) $\|Ux - Ux'\| \le \sqrt{A}\,\|x - x'\|$.*
4. *(Deformation stability) There exists $C > 0$, depending only on ψ, J and L, such that $\|UL_\tau x - L_\tau U x\| \le C\,\|\nabla\tau\|_\infty\,\|x\|$.*

The proof is given in Appendix A. Items 1 and 3 hold for any convolutional layer followed by a pointwise nonlinearity with bounded slope, including learned ones. Item 2 is the property specific to the fixed filterbank. The lower bound $B > 0$ is guaranteed analytically and independently of the training data, so no frequency band of the input is suppressed before the learned layers act. A learned kernel bank carries no such lower bound and can attenuate frequency bands that are uninformative under the training sampling pattern. With $A/B \approx 8$, the operator reweights bands within this range and suppresses none of them. All four items concern the operator U. The learned layers of the regularizer, including normalization, are not covered by them.

3.3 Regularizer Architecture

Scattering block. Since the filterbank $\{\phi_J, \psi_{j,\theta}\}$ is fixed, all learning in the block acts on the scattering representation. A learned 1×1 convolution mixes the scattering channels into an intermediate width, and learned 3×3 convolutions with instance normalization and LeakyReLU activations (slope 0.2) follow. The residual scattering block is

$$\mathcal{B}(x) = x + \mathcal{G}_\theta\big(\mathrm{Conv}_{1\times1}(U(x))\big), \tag{9}$$

where $\mathcal{G}_\theta$ denotes these convolution, normalization and activation stages. Each first-order channel carries the local amplitude envelope of one oriented band, so the mixing layer operates on a representation in which frequency content is already explicitly partitioned. By fixing the spectral partition analytically and learning on top of it, we decouple spectral decomposition from coefficient adaptation and reduce the effective hypothesis space of the regularizer.

Network. The regularizer $\mathcal{U}_\theta$ is a U-shaped encoder-decoder network. Scattering blocks $\mathcal{B}$ replace standard convolutional layers in the encoder at each resolution level, followed by average pooling for downsampling. The decoder uses learned transposed convolutions for upsampling, skip connections from corresponding encoder levels, and two learned 3×3 convolution–instance-normalization–LeakyReLU stages after each merge, followed by a final 1×1 output head. The pooling hierarchy provides the coarser scales of analysis, and the scattering blocks provide the complete oriented decomposition at each level.

Training objective. Following Sriram et al. (2020) and Jatyani et al. (2025), all networks are trained end-to-end with the loss $\mathcal{L}(\hat{x}, x^*) = -\mathrm{SSIM}(\hat{x}, x^*)$ between the reconstruction $\hat{x}$ and the fully sampled reference x^*.

4 Training Setup

4.1 Datasets

Following Jatyani et al. (2025), we evaluate on the public fastMRI multi-coil knee and brain datasets (Zbontar et al., 2018). The **knee dataset** contains 973 training volumes (34,742 slices) and 199 validation volumes (7,135 slices), with 15 coils. For the **brain dataset**, we use the T2 subset with 6,262 training and 502 evaluation slices, with 16 coils. Complex coil images are center-cropped to 320×320 in image space and transformed back to centered k-space following Jatyani et al. (2025).

4.2 Undersampling Protocol

We train all models using equispaced Cartesian sampling with an acceleration factor of $4\times$. To assess robustness to sampling distribution shifts, we evaluate two settings: in-distribution evaluation, where we train and test on equispaced masks, and cross-mask evaluation, where we train on equispaced masks and test on alternative sampling patterns. We use the same sampling density and calibration region configuration as the baseline for all masks. All masks are generated with the official implementation of Jatyani et al. (2025). Jatyani et al. (2025) train on equispaced $4\times$ data and then fine-tune on a mixed set of all six sampling patterns, or all four acceleration factors in the cross-rate experiments. We adopt a stricter protocol in which the test pattern is unseen during training, which measures generalization to new sampling patterns. Our zero-shot setting uses only equispaced $4\times$ data, and our finetuning setting adds 100 slices of the test pattern. We retrain NO with the official implementation in both settings.

4.3 Architecture and Hyperparameters

We follow the exact configuration reported by Jatyani et al. (2025). We use a variational network with $T = 12$ cascades, each consisting of a data consistency step and a learned regularizer. Coil sensitivity maps are estimated using a U-shaped model with base width 8 and 4 pooling layers. The k-space initialization module (KNO) is implemented as a U-shaped with base width 16 and 4 pooling layers, operating on sensitivity-reduced single-coil k-space. The cascade regularizer uses a U-Net with base width 18 and 4 pooling layers, with two convolutional layers per resolution level followed by instance normalization and nonlinear activation. For AnalyticVarNet, we replace standard convolutions in the encoder blocks with a covariant wavelet scattering transform using one scale ($J = 1$) and $L = 8$ orientations. The resulting scattering coefficients are linearly mixed using a learned convolutional layer. We train all models for 50 epochs using Adam with learning rate 3×10^{-4} and batch size 4. The reported NO comparison uses the official implementation under the same single-mask training condition as AnalyticVarNet.

5 Experimental Results

5.1 In-Distribution Reconstruction

We first evaluate reconstruction quality when training and testing on equispaced undersampling pattern with $4\times$ acceleration. Under identical training conditions, the proposed method consistently improves PSNR and SSIM over the NO baseline on both knee and brain datasets. Table 1 reports reconstruction performance on fastMRI knee and brain datasets under $4\times$ equispaced acceleration, where training and evaluation follow identical protocols. Under identical training settings, our method improves over all baselines on both datasets.

5.2 Cross-Mask Robustness

We evaluate robustness to sampling shift under two regimes. First, we report zero-shot performance, where models trained with equispaced sampling at $4\times$ acceleration are evaluated on unseen masks without retraining. Second, we allow limited adaptation by finetuning on 0.4% (100 slices) of the training data for 10 epochs, simulating a low-data transfer setting.

All evaluations use the same undersampling ratio. Table 2 separates two groups of patterns. On Magic and Random masks, which share the one-dimensional Cartesian structure of the training

Table 1: Reconstruction performance on the fastMRI knee and brain datasets at $4\times$ equispaced acceleration. Results are reported as mean $\pm$ standard deviation.

Method	Knee		Brain	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Zero-filled	31.00±3.33	0.7848±0.0616	30.86±1.73	0.8065±0.0376
ℓ_1-Wavelet (Lustig et al., 2007)	25.67±3.91	0.5667±0.2001	28.68±1.31	0.6783±0.0684
CSGM (Jalal et al., 2021)	26.52±3.21	0.6789±0.1220	–	–
ScoreMRI (Chung & Ye, 2022)	25.72±1.80	0.5789±0.0910	–	–
PnP-DM (Wu et al., 2024)	26.52±3.14	0.6383±0.1320	–	–
U-Net (Ronneberger et al., 2015)	37.07±2.47	0.8803±0.0504	37.27±1.76	0.9211±0.0272
E2E-VN (Sriram et al., 2020)	38.33±3.06	0.9048±0.0732	38.06±2.70	0.9620±0.0107
NO† (Jatyani et al., 2025)	39.14±2.93	0.9219±0.0724	38.82±2.77	0.9621±0.0086
AnalyticVarNet	**39.27±2.57**	**0.9270±0.0579**	**39.10±1.87**	**0.9655±0.0227**

† Reproduced using the official architecture and training configuration.
Results for Zero-filled, ℓ_1-Wavelet, CSGM, ScoreMRI, PnP-DM, U-Net, and E2E-VN are those reported by Jatyani et al. (2025) on the same data splits and evaluation protocol.

Table 2: Cross-mask generalization results on the fastMRI knee dataset comparing NO and the proposed method under different undersampling patterns. Metrics are reported as mean $\pm$ standard deviation. FT denotes finetuning on a smaller dataset.

		PSNR (dB)↑		SSIM↑	
Pattern	Regime	NO	AnalyticVarNet	NO	AnalyticVarNet
Magic	Zero-shot	**38.56±2.32**	38.29±2.32	**0.921±0.055**	0.919±0.053
	FT	39.19±3.31	**39.35±2.88**	0.918±0.055	**0.923±0.054**
Random	Zero-shot	37.08±2.17	**37.40±2.31**	0.902±0.036	**0.902±0.055**
	FT	37.55±2.31	**37.77±2.42**	0.907±0.056	0.907±0.055
Radial	Zero-shot	22.18±2.72	**31.97±2.58**	0.690±0.061	**0.846±0.051**
	FT	34.97±2.37	**36.44±2.32**	0.895±0.054	**0.903±0.051**
Gaussian	Zero-shot	11.85±2.05	**26.67±3.02**	0.345±0.080	**0.786±0.059**
	FT	29.74±1.94	**31.01±2.39**	0.800±0.103	**0.869±0.051**
Poisson	Zero-shot	13.16±2.82	**27.67±3.23**	0.344±0.085	**0.781±0.062**
	FT	31.71±2.49	**32.79±2.42**	0.863±0.051	**0.864±0.051**

mask, the two models are within 0.3 dB PSNR. On the two-dimensional incoherent patterns, AnalyticVarNet improves PSNR zero-shot by 9.8 dB (Radial), 14.8 dB (Gaussian), and 14.5 dB (Poisson), and it remains ahead by 1.1 to 1.5 dB after finetuning. The two groups differ in which regions of k-space are missing, and the gains appear where these regions differ most from training. This is consistent with the lower frame bound of Sec. 3.2, which keeps every frequency band represented in the regularizer independently of the training pattern. Figure 2 presents representative reconstructions under different masks.

5.3 Performance Across Acceleration Factors

We evaluate zero-shot reconstruction performance across multiple equispaced acceleration factors. Both models are trained once at $4\times$ acceleration and the same checkpoints are evaluated at $8\times$, $16\times$, and $32\times$, without finetuning. For each acceleration level, models are trained and evaluated under identical sampling conditions to isolate the effect of undersampling severity. As acceleration increases, fewer k-space measurements are available, making the reconstruction problem increasingly ill-posed and requiring stronger priors. We compare AnalyticVarNet with the baseline using identi-

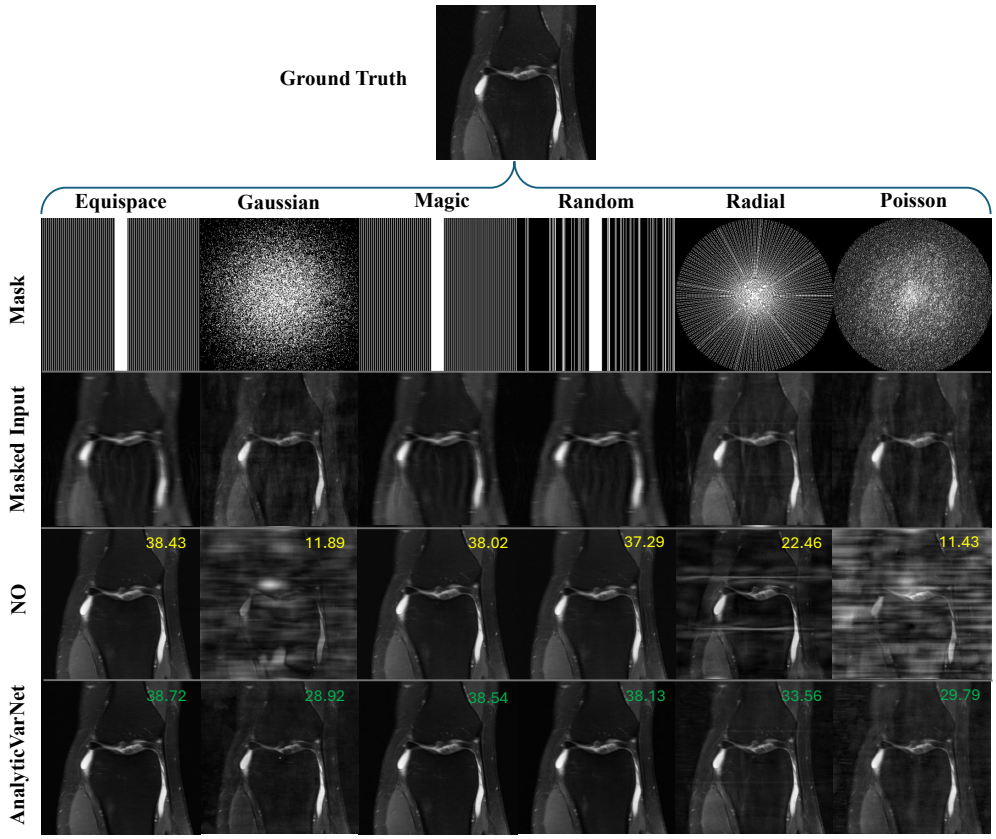


Figure 2: Qualitative comparison under sampling mask distribution shift. Models are trained on the knee dataset with equispaced $4\times$ sampling and evaluated on both in-distribution (equispaced) and unseen masks (Gaussian, Magic, Random, Radial, and Poisson). From top to bottom: ground truth and sampling masks, zero-filled input, NO (Jatyani et al., 2025), and the proposed method. PSNR (dB) is shown below each reconstruction. The proposed approach yields more stable reconstructions across unseen masks.

Table 3: Zero-shot performance across equispaced acceleration factors. Models are trained at $4\times$ and evaluated without finetuning. Results are reported as mean $\pm$ standard deviation.

	PSNR (dB)↑		SSIM↑	
Acceleration	NO	AnalyticVarNet	NO	AnalyticVarNet
$8\times$	29.00±2.42	**29.53±2.43**	0.7960±0.0520	**0.8810±0.0531**
$16\times$	22.10±3.25	**25.02±2.93**	0.6134±0.0882	**0.6936±0.0610**
$32\times$	12.94±3.01	**22.43±2.91**	0.4246±0.0102	**0.6289±0.0687**

cal architectural capacity and training protocols, without additional fine-tuning. Table 3 shows that AnalyticVarNet maintains competitive or improved reconstruction quality as acceleration increases.

5.4 Zero-Shot Cross-Resolution Generalization

We evaluate whether the learned reconstruction operator generalizes across spatial resolutions without retraining. Models are trained on 320×320 images and evaluated at 640×640 resolution under two zero-shot transfer settings: In the first setting, following the evaluation protocol of Jatyani et al. (2025), the input is upsampled to the target resolution using bilinear interpolation before reconstruction. This interpolation adds no measurement information. It provides a consistent discretization on which the learned reconstruction dynamics are evaluated.

Table 4: Zero-shot cross-resolution reconstruction. Models are trained at 320×320 and evaluated at 640×640 under $4\times$ equispaced undersampling without retraining. Results are reported as mean $\pm$ standard deviation.

Setting	Method	PSNR (dB)↑	SSIM↑
Interpolation Resolution	NO (Jatyani et al., 2025)	31.69 $\pm$ 2.13	0.869 $\pm$ 0.037
	AnalyticVarNet	**38.28** $\pm$ 2.57	**0.910** $\pm$ 0.057
Native Resolution	NO (Jatyani et al., 2025)	31.13 $\pm$ 2.36	0.8106 $\pm$ 0.0549
	AnalyticVarNet	**31.36** $\pm$ 2.24	**0.8746** $\pm$ 0.0556

In the second setting, models are directly evaluated on native high-resolution 640×640 k-space measurements without interpolation, resizing, or discretization alignment. This setting evaluates whether the reconstruction operator generalizes across acquisition grids that were not observed during training.

All experiments are performed strictly zero-shot, without retraining, finetuning, or architectural modification at the target resolution. The neural operator baseline (Jatyani et al., 2025) is implemented using the official publicly released codebase and evaluated under the same protocol. Table 4 reports the results. With interpolated inputs, AnalyticVarNet improves over NO by 6.6 dB PSNR and 0.041 SSIM. On native 640×640 measurements the two methods reach comparable PSNR (31.36 and 31.13 dB), and AnalyticVarNet attains higher SSIM (0.875 and 0.811). Appendix C shows qualitative examples.

Table 5: Comparison of model complexity and inference efficiency. Inference time is measured per slice.

Model	Trainable Parameters	Inference Time (ms) ↓
Neural Operator (NO) (Jatyani et al., 2025)	141,810,568	177.38
AnalyticVarNet	57,366,152	**145.34**

5.5 Parameter Efficiency

The scattering transform has no trainable weights, and AnalyticVarNet has $2.47\times$ fewer trainable parameters than the Neural Operator (NO) baseline. On a single NVIDIA RTX 4090 GPU, it also reduces inference time from 177.38 ms to 145.34 ms per slice, an 18.1% speedup. The GPU configuration was kept fixed for both models. Table 5 summarizes these results.

6 Conclusion

We introduced AnalyticVarNet, a variational network for accelerated multi-coil MRI reconstruction that replaces learned spectral operators with a fixed covariant wavelet scattering transform. By decoupling spectral decomposition from coefficient adaptation, the method provides a complete oriented representation with explicit frame bounds, on which the learned layers operate. Experiments demonstrate improved reconstruction quality and stronger robustness to sampling mask distribution shifts under single-pattern training, and higher SSIM when transferred to a finer grid without retraining.

Limitations. Data consistency weights the acquired samples by a single learned scalar per cascade and does not model the measurement noise explicitly. All experiments use retrospective undersampling of fully sampled fastMRI data, which does not capture every aspect of the acquisition physics, and prospective evaluation is left for future work. All models are trained on a single sampling pattern. Multi-pattern training (Jatyani et al., 2025) is a complementary route to robustness that we do not combine with our operator here. The guarantees of Proposition 1 apply to the fixed operator and do not extend to the learned layers of the regularizer. Finally, the filterbank is defined on the pixel grid, so its physical scale changes with the acquisition resolution.

REFERENCES

Mathieu Andreux, Tomás Angles, Georgios Exarchakis, Roberto Leonarduzzi, Gaspar Rochette, Louis Thiry, John Zarka, Stéphane Mallat, Joakim Andén, Eugene Belilovsky, et al. Kymatio: Scattering transforms in python. *Journal of Machine Learning Research*, 21(60):1–6, 2020.

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Mohammad Zalbagi Darestani, Akshay S. Chaudhari, and Reinhard Heckel. Measuring robustness in deep learning based compressive sensing. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 2433–2444, 2021.

Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

Alper Güngör, Salman UH Dar, Şaban Öztürk, Yilmaz Korkmaz, Hasan A Bedel, Gokberk Elmas, Muzaffer Ozbey, and Tolga Çukur. Adaptive diffusion priors for accelerated mri reconstruction. *Medical image analysis*, 88:102872, 2023.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

Jörn-Henrik Jacobsen, Jan van Gemert, Zhongyu Lou, and Arnold W. M. Smeulders. Structured receptive fields in CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2610–2619, 2016.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in neural information processing systems*, 34:14938–14954, 2021.

Armeet Singh Jatyani, Jiayun Wang, Aditi Chandrashekar, Zihui Wu, Miguel Liu-Schiaffini, Bahareh Tolooshams, and Anima Anandkumar. A unified model for compressed sensing mri across undersampling patterns. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26004–26013, 2025.

Florian Knoll, Kerstin Hammernik, Erich Kobler, Thomas Pock, Michael P. Recht, and Daniel K. Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, 81(1):116–128, 2019.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 773–782, 2018.

Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

Jeremy Ocampo, Matthew A. Price, and Jason D. McEwen. Scalable and equivariant spherical CNNs by discrete-continuous (DISCO) convolutions. In *International Conference on Learning Representations (ICLR)*, 2023.

Edouard Oyallon. *Analyzing and introducing structures in deep convolutional neural networks*. PhD thesis, Université Paris sciences et lettres, 2017.

Edouard Oyallon, Eugene Belilovsky, and Sergey Zagoruyko. Scaling the scattering transform: Deep hybrid networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5618–5627, 2017.

Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated mri reconstruction. In *International conference on medical image computing and computer-assisted intervention*, pp. 64–73. Springer, 2020.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. *Advances in Neural Information Processing Systems*, 37:118389–118427, 2024.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.