Physics in Medicine &
Biology

Deep learning for restoring MPI system matrices using simulated training data

IOP Publishing Phys. Med. Biol. 71 (2026) 095029 https://doi.org/10.1088/1361-6560/ae6016

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
12 December 2025

REVISED
19 March 2026

ACCEPTED FOR PUBLICATION
14 April 2026

PUBLISHED
13 May 2026

PAPER

Deep learning for restoring MPI system matrices using simulated training data

Artyom Tsanda[1,2,*], **Sarah Reiss**[1,2], **Konrad Scheffler**[1,2], **Marija Boberg**[1,2] and **Tobias Knopp**[1,2,3]

[1] Institute for Biomedical Imaging, Hamburg University of Technology, Hamburg, Germany
[2] Section for Biomedical Imaging, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3] Fraunhofer Research Institution for Individualized Medical Technology and Engineering IMTE, Lübeck, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: artyom.tsanda@tuhh.de

Keywords: magnetic particle imaging, system matrix recovery, machine learning, image restoration

Abstract

Objective. Magnetic particle imaging reconstructs tracer distributions using a system matrix (SM) obtained through time-consuming, noise-prone calibration measurements. Methods for addressing imperfections in measured system matrices increasingly rely on deep neural networks, yet curated training data remain scarce. This study evaluates whether physics-based simulated system matrices can be used to train deep learning (DL) models for different SM restoration tasks, i.e. denoising, accelerated calibration, upsampling, and inpainting, that generalize to measured data. *Approach.* A large dataset of system matrices was generated using an equilibrium magnetization model extended with uniaxial anisotropy. The dataset spans particle, scanner, and calibration parameters for 2D and 3D trajectories, and includes background noise injected from empty-frame measurements. For each restoration task, DL models were compared with classical non-learning baseline methods. Quantitative performance was evaluated on simulated data using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). For measured data, performance was assessed qualitatively by visual comparison of system matrices and the resulting reconstructions. *Main results.* The models trained solely on simulated system matrices generalized to measured data across all tasks: for denoising, DnCNN/RDN/SwinIR outperformed discrete cosine transform and soft thresholding baseline by >10 dB PSNR and up to $+0.1$ SSIM on simulations and led to perceptually better reconstructions of real data; for 2D upsampling, SMRnet exceeded bicubic by ~ 20 dB PSNR and ~ 0.08 SSIM at $\times 2$–$\times 4$ but these gains did not transfer qualitatively to real measurements. For 3D accelerated calibration, SMRnet matched tricubic in noiseless cases and was more robust under noise, and for 3D inpainting, biharmonic inpainting was superior when noise-free but degraded with noise, while a PConvUNet maintained quality and yielded less blurry reconstructions. *Significance.* The demonstrated transferability of DL models trained on simulations to real measurements mitigates the data-scarcity problem, which intensifies with model scale. This enables the development of new methods beyond current measurement capabilities and supports pre-training of large models on simulated data.

1. Introduction

Magnetic particle imaging (MPI) is a preclinical tomographic imaging modality recovering the spatial distribution of superparamagnetic iron oxide nanoparticles (Gleich and Weizenecker 2005). Due to inherent magnetic properties, the particles produce a detectable signal when exposed to an oscillating magnetic field. Spatial encoding in MPI is achieved through the superposition of a spatially homogeneous oscillating drive field (DF) and a static gradient field, known as the selection field (SF). The SF defines regions of low magnetic field strength, practically either a field-free point (FFP) or a field-free line (FFL), within which particles contribute to the measured signal. The spatial encoding is achieved

by moving this region across the field of view (FOV). MPI provides quantitative, background-free, and real-time imaging capabilities without the use of ionizing radiation (Knopp *et al* 2017). Potential clinical applications include cardiovascular imaging (Weizenecker *et al* 2009, Haegele *et al* 2012, Salamon *et al* 2016, Herz *et al* 2019, Mohtashamdolatshahi *et al* 2020), perfusion assessment (Molwitz *et al* 2019, Ludewig *et al* 2022), and localized magnetic hyperthermia (Tay *et al* 2018, Healy *et al* 2022).

The absence of particle-particle interactions makes the MPI signal linear with respect to particle concentration. For this reason, measuring the response to a small delta sample of known concentration, moved across a predefined grid, known as the calibration process, yields the forward mapping from concentration distributions to the signal, called the system matrix (SM). The SM defines the linear system used for reconstructing unknown concentration maps, thereby playing a decisive role in the quality of the resulting images. Alternatively, the MPI signal can be characterized by a point-spread function, leading to the x-space approach (Goodwill and Conolly 2011).

Acquiring a SM is a time-consuming process, it can take about 32 h for a $37 \times 37 \times 37$ grid. Yet, being dependent on scanner, acquisition, and particle parameters, the SM needs to be re-measured every time one of them changes. For this reason, accelerating the SM measurement is actively being researched. In early work, compressed sensing was applied to recover a full SM acquired on an incoherently downsampled grid (Knopp and Weber 2013, Weber and Knopp 2015, Ilbey *et al* 2019). Super-resolution methods, primarily based on deep learning (DL), have been applied to a regularly downsampled grid (Baltruschat *et al* 2020, Kluth *et al* 2020, Güngör *et al* 2021, 2022, Shi *et al* 2023, Zhang *et al* 2025). Alternatively, when sampled sufficiently, the calibration size of a measured SM can be further increased, resulting in reconstructions with improved perceived image quality. Another downside of the measured SM is the presence of noise; although averaging mitigates it, the effect persists at higher frequencies. The problem has been addressed by frequency-domain filters (Weber *et al* 2015) and data compression (Grosser *et al* 2020). Lastly, measurements for some positions may be corrupted, often becoming evident only retrospectively. The described imperfections form a set of SM restoration problems considered in this paper: denoising, accelerated calibration, upsampling, and inpainting. In the MPI community, the term 'SM recovery' is primarily used for the estimation of missing SM values caused by downsampling. In this work, this task is referred to as 'accelerated calibration', and the term 'restoration' is used to denote all SM-related problems considered here.

Despite the trend towards DL-based SM restoration methods and the dependency of the SM on multiple parameters, the amount of available SM data remains limited. The majority of methods rely on the Open MPI dataset, which provides SMs measured on a preclinical FFP scanner (Bruker, Ettlingen) (Knopp *et al* 2020). It includes a few grid sizes ($19 \times 19 \times 19$, $33 \times 33 \times 27$, and $37 \times 37 \times 37$) and two tracer types (Perimag and Synomag-D, Micromod GmbH, Germany). Given its limited diversity and modest size, the dataset may pose challenges for training DL models that generalize across different systems and acquisition settings. Consequently, high-quality datasets covering the variability of the parameter space are still in demand for training DL models.

Alternatively, the SM can be computed using a physical model (Knopp *et al* 2010c, Kluth 2018). Accurately describing particle dynamics becomes critical for this approach. For example, taking into account relaxation effects leading to anisotropies by solving the Fokker–Planck equation for the probability density function of magnetic moments. In this case, the task becomes computationally expensive. Recently, the equilibrium model has been extended to take anisotropies into account while maintaining reasonable reconstruction quality (Maass *et al* 2024). The primary goal of the described approaches is to substitute the tedious calibration process; however, they are rarely applied in the DL context.

In this work, we address the problem of data scarcity in SM restoration using DL models. Although direct measurement remains the most accurate method for characterizing an MPI system, we argue that simulated SM data are sufficiently accurate to train DL-based restoration models that can subsequently be applied to real observations. To demonstrate this, we generated a training dataset spanning the SM parameter space and considered several restoration problems showing the generalization capability of the trained DL models. This paper has the following major contributions:

1. A method for generating a comprehensive modeled SM dataset spanning particle, scanner, and calibration parameters for 2D/3D Lissajous trajectories.
2. Applications, such as denoising, accelerated calibration, upsampling, and inpainting, showing that DL models trained exclusively on simulated SMs generalize to measured data.
3. New restoration methods for SM denoising and inpainting.

This paper is an extension of the conference abstract (Tsanda *et al* 2025c). Compared to the abstract, we made the following major extensions. First, we expanded the parameter space for SM simulation, including 3D trajectories. Second, we improved the quality of denoising by employing new methods. Third, we considered new applications, namely upsampling and inpainting.

In line with the principles of open science, we are making the source code (Tsanda *et al* 2025a) and data (Tsanda *et al* 2025b) used in this study publicly available to ensure reproducibility and support future research.

2. Background

2.1. Related works

In previous works, simulated data have been primarily used for quantitative evaluations of the proposed methods. SMs simulated for the in-house open-sided FFL MPI scanner (Top and Güngör 2020) were employed to evaluate the proposed SM super-resolution method and to perform ablation studies on individual components of the DL model (Güngör *et al* 2022). The dataset included 100 2D simulated SMs with varied SF gradients and particle diameters. A pre-training strategy using pseudo-labeled low-resolution SM data, which incorporates frequency and coil information as tokens in the transformer model, was proposed (Shi *et al* 2023). The approach allows for more effective fine-tuning on real data later on. The method was validated on a simulated dataset, namely four 3D SMs were simulated using Langevin model with different SF gradients (0.5, 1.0 and 5 T m $^{-1}\mu_0^{-1}$) and particle diameters (25 and 12.5 nm). Unlike previous approaches, we show that models trained on simulated data can generalize to measured SMs.

Concurrently, a DL-based denoising approach for SMs was introduced (Liu *et al* 2025). The authors evaluated several denoising models, including DnCNN (Zhang *et al* 2017) and SwinIR (Liang *et al* 2021) considered in this paper, and proposed a new encoder-decoder architecture with elements of the Swin Transformer (Liu *et al* 2021). Their models were trained on a single 2D simulated SM generated using the Langevin model with added white Gaussian noise, and validated on the Open MPI dataset with synthetic Gaussian noise as well as on an in-house dataset. That study focused on methodological developments for denoising. In contrast, our study investigates the transferability of DL models trained on simulated data in general considering a broad range of applications. We employ a more comprehensive simulation framework, vary simulation parameters, and incorporate real background measurements as the noise source. For denoising specifically, we additionally consider 3D data and observe improved qualitative performance for DnCNN and SwinIR on measured phantoms, with noticeably reduced noise in the reconstructions.

2.2. MPI signal model and system function

With negligible particle-particle interactions, the MPI signal in the frequency domain $u_{l,k} \in \mathbb{C}$ measured in the imaging volume $\Omega \subset \mathbb{R}^3$ from receive coil $l \in \{1,\dots,L\}$ at frequency index $k \in \{0,\dots,K-1\}$ is related to the spatial particle distribution $c : \Omega \to \mathbb{R}$ by

$$u_{l,k} = \int_\Omega \underbrace{\left(-a_k \frac{\mu_0}{T} \boldsymbol{p}_l(\boldsymbol{r}) \cdot \int_0^T \mathrm{e}^{-2\pi \mathrm{i} k t/T} \frac{\partial}{\partial t} \overline{\boldsymbol{m}}(\boldsymbol{H}(\boldsymbol{r},t),t)\, \mathrm{d}t \right)}_{=: s_{l,k}(\boldsymbol{r})} c(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \tag{1}$$

where $s_{l,k} : \Omega \to \mathbb{C}$ is the system function, $a_k \in \mathbb{C}$ is the transfer function of the analog filter in the signal acquisition chain, $\boldsymbol{p}_l : \Omega \to \mathbb{R}^3$ is the sensitivity profile of the receive coil l, T is the acquisition time, $\overline{\boldsymbol{m}} : \mathbb{R}^3 \times (0,T) \to \mathbb{R}^3$ is the mean magnetic moment depending on the applied magnetic field $\boldsymbol{H} : \Omega \times (0,T) \to \mathbb{R}^3$, μ_0 is the vacuum permeability (Kluth 2018).

To evaluate the system function in equation (1), the mean magnetic moment must be estimated, which is challenging because the underlying distribution of magnetic moments is unknown. Assuming thermodynamic equilibrium and a static magnetic field, an analytical solution can be derived based on the Langevin function. Albeit being one of the most studied, the Langevin model does not take into account relaxation effects present in real particle systems. A more accurate approach requires solving the Landau–Lifshitz–Gilbert equation using the Fokker–Planck method, which is computationally expensive (Kluth 2018). To avoid these complications, the system function is typically measured using a small

sample $\boldsymbol{\Delta}:\Omega\to\{0,1\}$ at a position $\boldsymbol{r}_n\in\Omega$, $n\in\{1,\ldots,N\}$, filled with a fixed particle concentration c_0:

$$u_{l,k}^{n}=\int_{\Omega}s_{l,k}\left(\boldsymbol{r}\right)c_0\boldsymbol{\Delta}\left(\boldsymbol{r}-\boldsymbol{r}_n\right)\mathrm{d}\boldsymbol{r}\approx c_0 s_{l,k}\left(\boldsymbol{r}_n\right)\underbrace{\int_{\Omega}\boldsymbol{\Delta}\left(\boldsymbol{r}\right)\mathrm{d}\boldsymbol{r}}_{=:V_{\Delta}}. \tag{2}$$

The measured discrete system function, hereafter referred to as the SM, is then formed by moving the sample on a discrete Cartesian grid and repeating the measurement. The resulting matrix $\boldsymbol{S}\in\mathbb{C}^{M\times N}$ maps N spatial locations to $M=KL$ frequencies from receive coils.

Recently, an extension to the equilibrium model has been proposed to account for anisotropies in magnetization dynamics (Maass *et al* 2024). It adds an anisotropy term to the partition function:

$$\mathcal{Z}\left(\beta\boldsymbol{H};\mathbb{O}\right)=\int_{\mathbb{S}^2}\mathrm{e}^{\beta\boldsymbol{H}^{\mathrm{T}}\boldsymbol{m}+\alpha_K(\boldsymbol{r})\left(\boldsymbol{n}(\boldsymbol{r})^{\mathrm{T}}\boldsymbol{m}\right)^2}\mathrm{d}\boldsymbol{m}, \tag{3}$$

with

$$\beta:=\frac{\mu_0 m_0}{k_{\mathrm{B}}T_{\mathrm{P}}}. \tag{4}$$

Here, k_{B} is the Boltzmann constant, T_{P} is the particle temperature, $m_0\in\mathbb{R}^+$ is the magnitude of the magnetic moment of a single particle, $\alpha_K:\mathbb{R}^3\to\mathbb{R}$ and $\boldsymbol{n}:\mathbb{R}^3\to\mathbb{S}^2$ are, respectively, the spatially varying anisotropy constant and the uniaxial easy-axis direction; together they form the observable parameter set $\mathbb{O}$. The model describes immobilized particles, for which both the easy axis and the anisotropy constant are fixed, and fluid particles, for which the easy axis is oriented along the SF and the anisotropy depends on the spatial modulation parameter $q\in\mathbb{R}_+$ (Maass *et al* 2024). The mean magnetic moment can then be evaluated using

$$\overline{\boldsymbol{m}}\left(\boldsymbol{H}\right)=\frac{m_0}{\beta}\nabla_{\boldsymbol{H}}\ln\left(\mathcal{Z}\left(\beta\boldsymbol{H}\right);\mathbb{O}\right). \tag{5}$$

In this case, equation (5) cannot be solved in closed form but can be approximated using truncated series of Bessel functions. The resulting SM provides accurate enough estimations for a SM to reconstruct real data while being computationally less demanding than, for example, the Landau–Lifshitz–Gilbert equation approach. For this reason, we select this model in this paper.

2.3. Reconstruction

Having a SM $\boldsymbol{S}\in\mathbb{C}^{M\times N}$ and a measurement $\boldsymbol{u}_{\mathrm{meas}}\in\mathbb{C}^{M}$, we reconstruct the unknown concentration vector $\boldsymbol{c}\in\mathbb{R}_+^N$ by solving the Tikhonov regularized least squares problem:

$$\underset{\boldsymbol{c}\in\mathbb{R}_+^N}{\operatorname{argmin}}\ \left\|\boldsymbol{W}\tilde{\boldsymbol{P}}_{\Theta}\left(\boldsymbol{S}\right)\boldsymbol{c}-\boldsymbol{W}\boldsymbol{P}_{\Theta}\left(\boldsymbol{u}_{\mathrm{meas}}\right)\right\|_2^2+\lambda\left\|\boldsymbol{c}\right\|_2^2, \tag{6}$$

where $\boldsymbol{P}_{\Theta}:\mathbb{C}^{KL}\to\mathbb{C}^{K_{\Theta}L}$ and $\tilde{\boldsymbol{P}}_{\Theta}:\mathbb{C}^{KL\times N}\to\mathbb{C}^{K_{\Theta}L\times N}$ are the frequency selection based on signal-to-noise ratio (SNR) threshold Θ and resulting in K_{Θ} frequencies for the measurement and the SM, respectively, $\boldsymbol{W}\in\mathbb{R}_+^{K_{\Theta}L\times K_{\Theta}L}$ is the weighting based on the L^2 norm of the SM rows, λ is the regularization parameter. In this work, the regularization parameter is not further scaled but reported as is. The optimization problem equation (6) is solved using the Kaczmarz method with a fixed number of iterations $n_{\mathrm{iter}}\in\mathbb{N}$ (Knopp *et al* 2010b).

3. Method

3.1. Data generation

SMs in MPI depend on many parameters. To structure the simulation process, we divide them into three groups: particle, scanner, and calibration parameters. The summary for each of those can be found in table 1.

As we employ the equilibrium model with anisotropy from equation (3) to calculate the dynamics of the mean magnetic moment, the particle parameters are β, α_K, $\boldsymbol{n}$, and q. The magnetic moment of a single particle m_0 depends on the particle parameter d_{P} via:

$$m_0=\frac{\pi d_{\mathrm{P}}^3 M_{\mathrm{S}}}{6}, \tag{7}$$

Table 1. Simulation parameters used for data generation.

Parameter		Value
	Particle	
Temperature	T_P	293 K
Sat. magnetization	M_S	$474\,000\,\mathrm{J\,m^3\,T^{-1}}$
Particle core diameter	d_P	$\mathcal{U}(15^3, 25^3)^{1/3}$ nm
Anisotropy constant	K^anis	$10^{\mathcal{U}(3,4)}\,\mathrm{J\,m^{-3}}$
Fluid/immobilized		Bernoulli($p = 0.5$)
Spatial modulation (fluid)	q	$\mathcal{U}(0.3, 1.3)$
Easy-axis direction (immob.)	n	$\mathcal{U}(\mathbb{S}^2)$
	Scanner	
DF amplitudes	$[A_x, A_y, A_z]$	$[\mathcal{U}(5,14), \mathcal{U}(5,14), \mathcal{U}(5,14)]\,\mathrm{mT}\,\mu_0^{-1}$
DF frequencies	$[f_x, f_y, f_z]$	[2.5/102, 2.5/96, 2.5/99] MHz
SF gradient	$[G_x, G_y, G_z]$	$[\mathcal{U}(0.1, 1.5), \mathcal{U}(0.1, 1.5), -(G_x + G_y)]\,\mathrm{T\,m^{-1}}\,\mu_0^{-1}$
Trajectory		2D/3D Lissajous
DF FOV per axis	$\mathrm{FOV}_i^\mathrm{DF}$	$2A_i/\lvert G_i\rvert$ mm
	Calibration	
Calibration FOV per axis	$\mathrm{FOV}_i^\mathrm{calib}$	$\mathrm{FOV}_i^\mathrm{DF} \cdot \mathcal{U}(1,2)$ mm
Margins per axis	m_i	$(\mathrm{FOV}_i^\mathrm{calib} - \mathrm{FOV}_i^\mathrm{DF})/2$ mm
Calibration center per axis		$\mathcal{U}(-m_i, m_i)$ mm
Calibration size (voxels/axis)	N_i^calib	$\lfloor \mathrm{FOV}_i^\mathrm{calib}/R_i^\mathrm{FWHM} \cdot \mathcal{U}(6.24, 8.32) \rceil$

where M_S is the saturation magnetization of the particle core, taken fixed at $474\,000\,\mathrm{J\,m^3\,T^{-1}}$ (Deissler *et al* 2014). We vary the particle diameter d_P, which changes the magnetic moment of a single particle m_0 and thus the parameter β (see equation (4)). We select the same range of particle diameters as used by Maass *et al* (2024), namely [15, 25] nm. The cubed diameter is then sampled uniformly between the corresponding cubed values. The temperature is taken fixed at 293 K. The anisotropy constant K^anis is drawn from a log-uniform distribution over the interval $[10^3, 10^4]\,\mathrm{J\,m^{-3}}$. To augment the resulting dataset, we consider both immobilized and fluid particles. Each particle type is selected with probability 0.5 for every simulated SM. When immobilized, the direction of the easy axis is sampled uniformly on the unit sphere. When fluid, the spatial modulation parameter q is sampled uniformly from [0.3, 1.3], where the range is selected heuristically.

In this paper, we limit data generation to FFP scanners with Lissajous trajectories. In this case, the SF is parametrized with three spatial gradients G_x, G_y, G_z. We select the first two components along x and y uniformly from [0.1, 1.5] $\mathrm{T\,m^{-1}}\,\mu_0^{-1}$. The gradient along z is then calculated as the negative sum of the first two to fulfill Gauss's law for magnetism. The range of DF amplitudes $A_i,\ i \in \{x, y, z\}$, is set to [5,14] $\mathrm{mT}.\mu_0^{-1}$. Both ranges for SF gradients and DF amplitudes were selected according to hardware specifications of several existing MPI systems: a preclinical scanner (Bruker, Ettlingen) (Knopp *et al* 2020), a human-sized scanner (Thieben *et al* 2024), and an open-sided scanner (Top and Güngör 2020). We consider 2D and 3D Lissajous trajectories. DF frequencies f_x, f_y, f_z remain fixed at $2.5/102, 2.5/96, 2.5/99$ MHz respectively.

In the final subset of parameters, we specify the calibration. The calibration FOV is linked to the DF FOV: along each axis $i \in \{x, y, z\}$, $\mathrm{FOV}_i^\mathrm{calib}$ is drawn from $\mathrm{FOV}_i^\mathrm{DF} \cdot \mathcal{U}(1,2)$ where $\mathrm{FOV}_i^\mathrm{DF} = 2A_i/\lvert G_i\rvert$. Here, $\mathcal{U}(a,b)$ denotes the uniform distribution on the interval $[a,b]$. To fully cover the resulting SM, the center of the calibration FOV is sampled uniformly within the interval $[-m_i, m_i]$ defined by the margins $m_i = (\mathrm{FOV}_i^\mathrm{calib} - \mathrm{FOV}_i^\mathrm{DF})/2$. We sample the calibration size (pixels per axis) based on the 1D MPI convolution kernel's full width at half maximum (FWHM) (Knopp *et al* 2010a):

$$N_i^\mathrm{calib} = \left\lfloor \left(\frac{\mathrm{FOV}_i^\mathrm{calib}}{R_i^\mathrm{FWHM}} \cdot \mathcal{U}(6.24, 8.32) \right) \right\rceil, \qquad R_i^\mathrm{FWHM} = \frac{4.16}{\beta\,\lvert G_i\rvert}, \tag{8}$$

where G_i is the SF gradient along each axis.

Following the described sampling of simulation parameters we generate 1000/300/300 2D and 50/15/15 3D SMs for the training, validation, and test sets, respectively. A few examples are shown in figure 1.

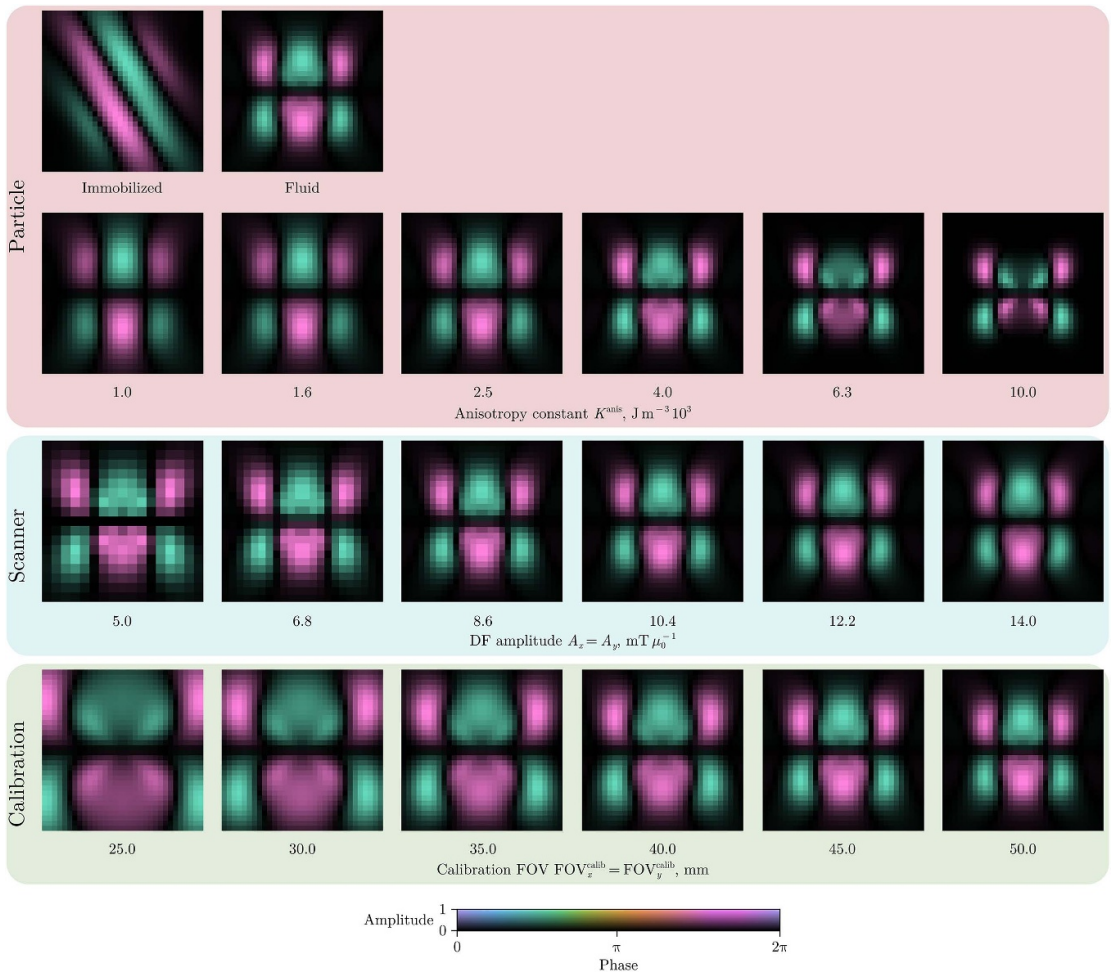

Figure 1. Simulated 2D SMs at the frequency $f = 99.5$ kHz illustrating variations across parameter groups: anisotropy constant (particle), DF amplitude (scanner), and calibration FOV (calibration). Unless stated otherwise, parameters are fixed to $d_{\text{P}} = 20$ nm, $K^{\text{anis}} = 3.2\,\text{J m}^{-3}\,10^3$, $q = 1$, $\boldsymbol{n} = [\cos(-3\pi/4), \sin(-3\pi/4)]$, $A_x = A_y = 10\,\text{mT}\,\mu_0^{-1}$, $G_x = G_y = 0.8\,\text{T m}^{-1}\,\mu_0^{-1}$, $\text{FOV}_x^{\text{calib}} = \text{FOV}_y^{\text{calib}} = 4A_x/G_x$, and $N_x^{\text{calib}} = N_y^{\text{calib}} = 7 \cdot \text{FOV}_x^{\text{calib}}/R_x^{\text{FWHM}}$.

3.2. Restoration problems

We consider four degradation problems: denoising, accelerated calibration, upsampling, and inpainting. Having a simulated ground-truth SM $\boldsymbol{S}_{\text{GT}}$, we corrupt it according to the problem. The corruption model can be generally described by the following equation

$$\boldsymbol{S}^p_{\text{corrupt}} = \mathcal{A}(\boldsymbol{S}_{\text{GT}}) + \boldsymbol{N}, \tag{9}$$

where $\boldsymbol{N} \in \mathbb{C}^{M\times N}$ is a noise matrix and $\mathcal{A} : \mathbb{C}^{M\times N} \to \mathbb{C}^{M\times N}$ is a linear transform specific to each problem: (i) identity for denoising; (ii) regular downsampling, selecting each nth pixel along each dimension, for accelerated calibration and upsampling; (iii) masking pixels corresponding to valid calibration measurements for inpainting.

Practically, in addition to the particle-induced signal $\boldsymbol{u}$ described in equation (1), the measured voltage $\boldsymbol{u}_{\text{meas}}$ contains the following components:

$$\boldsymbol{u}_{\text{meas}} = \boldsymbol{u} + \boldsymbol{u}^{\text{BG}} + \boldsymbol{u}^{\text{N}} + \boldsymbol{u}^{\text{D}}, \tag{10}$$

where $\boldsymbol{u}^{\text{BG}} \in \mathbb{C}^M$ is the systematic background signal, $\boldsymbol{u}^{\text{N}} \in \mathbb{C}^M$ is thermal noise, $\boldsymbol{u}^{\text{D}} \in \mathbb{C}^M$ corresponds to abrupt signal changes or distortions, which may be caused by hardware imperfections or external factors (Paysen *et al* 2020). In this work, we consider all terms except the particle signal to be noise. The drifts inside the background signals and the abrupt component are typically non-stationary in time and therefore affect different calibration measurements to varying degrees. Since a SM is measured consecutively for multiple positions, these temporal fluctuations result in spatially structured artifacts and inconsistent noise levels. As these effects are challenging to model, we take a sequence of background measurements

as a source of noise. In particular, we measured 7801 200 and 500 000 frames for 2D and 3D Lissajous trajectories respectively with the preclinical FFP scanner (Bruker, Ettlingen). To sample noise for a particular SM, we select a sub-sequence of frames and reshape them into the required size selecting the corresponding channel and frequency.

Denoising: As a baseline method for denoising, we use previously proposed frequency-domain filtering based on multidimensional discrete cosine transform and soft thresholding (DCT-F) (Weber *et al* 2015). For each complex frequency component (per frequency and channel), a DCT is applied along the spatial axes, and soft thresholding is performed by shrinking the magnitude by $\omega\sigma$ while preserving its phase; the denoised image is then obtained via the inverse DCT. We set $\omega = 2.75$. The method is non-blind and requires the noise standard deviation σ. For simulated data, we used the ground-truth values, whereas for real measurements we estimated σ from background noise measurements and heuristically adjusted it by a coefficient (0.3). We consider three DL methods for the denoising problem: DnCNN (Zhang *et al* 2017), RDN (Zhang *et al* 2021), and SwinIR (Liang *et al* 2021). DnCNN represents a sequence of convolution, normalization, and activation layers. In our prior study (Tsanda *et al* 2025c), we provided first denoising results with this method. In this work, we extend the method to 3D trajectories. RDN is composed of residual dense blocks and was used in MPI before for the plug-and-play reconstruction (Askin *et al* 2022, Tsanda *et al* 2024). Finally, SwinIR employs a parameter-effective transformer architecture with windowed self-attention layers and shows state-of-the-art denoising performance for natural images. However, due to hardware limitations, we do not apply SwinIR to 3D problems.

Accelerated calibration: The accelerated calibration task aims to reconstruct a fully sampled SM from a regularly downsampled 3D calibration grid. Regular downsampling induces aliasing and information loss, thereby degrading reconstruction quality. In contrast to spatial super-resolution, accelerated calibration must additionally account for measurement noise. We therefore formulate it as a joint completion-and-denoising problem. As a baseline method, we use tricubic spline interpolation applied independently across channels and frequencies. As a data-driven approach, we employ SMRnet (Baltruschat *et al* 2020), which consists of stacked residual dense blocks and a nearest-neighbor upsampling module followed by convolutional layers.

Upsampling: Although the formal degradation operator is identical to that of accelerated calibration, we treat upsampling separately due to its distinct objective and noise characteristics. In upsampling, an already complete and adequately sampled SM is available and sufficient for reconstruction. The goal is perceptual refinement by increasing the calibration grid resolution beyond the physically achievable optimum. Methods are applied as a post-processing step after denoising. The problem is an example where measured ground truth (GT) is not available for training. We study this problem in 2D and generate an additional dataset with larger calibration grids while keeping all other simulation parameters fixed. We evaluate bicubic interpolation as a baseline method and a 2D version of SMRnet.

Inpainting: During the SM calibration process, individual measurements at different spatial positions can be corrupted by external interference or internal system instabilities. Because a full calibration can last 32 h or more, the risk of corruption is substantial. To date, entire calibration scans are typically discarded and repeated. To mitigate this, we instead propose non-blind inpainting for 3D SMs, using known masks of invalid pixels provided to the method. Because the artifacts correspond to sequences of failed measurements at consecutive robot positions, we generate masks by first flattening the calibration volume into a one-dimensional index space, selecting a single contiguous block of missing indices, and then mapping this pattern back to three dimensions. To emulate different traversal patterns (raster or meandering), the order of elements in the last dimension is reversed with probability 0.5 for each mask. In addition, the spatial dimensions are randomly permuted before mask generation, and the permutation is inverted afterwards. The size of the mask is controlled by the ratio with respect to the number of elements in the original SM. We also combine multiple masks generated independently into one for training. See the original code for further details on mask generation (Tsanda *et al* 2025a). Since we apply methods independently to each frequency component, a new mask is randomly generated for each. As a baseline method, image inpainting is performed by solving biharmonic partial differential equations that extend image values smoothly into missing regions (Chui and Mhaskar 2010, Damelin and Hoang 2018). Alternatively, we employ a U-Net architecture with partial convolutions adapted to 3D (Liu *et al* 2018).

3.3. Training

All models were trained on complex-valued frequency components of SMs from the training set. For each component, the GT image was amplitude-normalized to unit maximum, rotated by a single random angle on the complex plane, and randomly scaled. Having a pre-processed GT, a task-specific degradation operator $\mathcal{A}$ was applied, and measured background noise was added after scaling to a randomly sampled standard deviation. The degraded sample was then re-normalized to the maximum amplitude of the corrupted image. These steps emulate the varying SNR in MPI and improve robustness to global phase shifts introduced by the acquisition chain's transfer function (see equation (1)). To handle varying calibration sizes, each image was zero-padded to a fixed patch size with randomly sampled padding on all sides, resulting in a varying offset within the patch. Finally, complex-valued frequency components were transformed into two-channel real-valued arrays corresponding to the real and imaginary parts.

For denoising, accelerated calibration, and upsampling, the objective was the L1 loss. For inpainting, a weighted L1 loss was used with separate weights for known and missing pixels. In all cases, the loss was computed only on valid (non-padded) pixels of the original sampling grid. Optimization used Adam with a multi-step learning rate schedule (Kingma and Ba 2015). We used mixed precision (Micikevicius *et al* 2018) and applied exponential moving averaging of the model weights during training. All models were trained on a single NVIDIA H100 GPU, except SwinIR, which required two GPUs. Further details can be found in the original code (Tsanda *et al* 2025a).

3.4. Evaluation

We evaluate the trained models on both simulated and real SMs. For simulated data, we fix the SNR by normalizing the magnitude of the ground-truth frequency components to their maximum and scaling the additive noise to a selected standard deviation. At inference, inputs are preprocessed as during training: scaled to unit magnitude, padded to the required patch size, and converted to two-channel arrays (real and imaginary parts). After prediction, these transformations are inverted.

For simulated data, per-frequency errors are quantified using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) (Wang *et al* 2004) and then averaged over the test set. We report mean values with 95% confidence intervals (CIs). The CI is computed using a normal approximation, namely for 95% confidence level $\bar{x} \pm 1.96 \cdot \sigma/\sqrt{n}$, where $\bar{x}$ is the sample mean, σ is the sample standard deviation computed with Bessel's correction, and n is the sample size.

To assess the generalization to real data, we consider several measurements acquired on a preclinical FFP scanner (Bruker, Ettlingen). Unless stated otherwise, all real measurements were acquired using Lissajous DF sequences with DF amplitudes $A_x = A_y = A_z = 12\,\mathrm{mT}\,\mu_0^{-1}$ (with A_z applicable only for 3D trajectories) and DF frequencies $f_x = 2.5/102\,\mathrm{MHz}$, $f_y = 2.5/96\,\mathrm{MHz}$, and, for 3D trajectories, $f_z = 2.5/99\,\mathrm{MHz}$. Data were acquired over three receive channels. For 2D datasets, 1633 frequency components were retained; for 3D datasets—53 857 components. Unless specified, the tracer was Perimag (Micromod GmbH, Germany), and SF gradients were $G_x = G_y = -1\,\mathrm{T\,m^{-1}}\,\mu_0^{-1}$ and $G_z = 2\,\mathrm{T\,m^{-1}}\,\mu_0^{-1}$. The reconstruction parameters were chosen individually for each measurement and then kept fixed across all methods applied to that measurement.

Denoising: To evaluate denoising methods, the snake phantom (see figure 2(a)) was measured using a 2D Lissajous sequence in the *xy*-plane (Maass *et al* 2024). The SM was acquired on a 17×15 Cartesian grid covering a $34 \times 30\,\mathrm{mm}^2$ FOV. At each position the sequence was repeated multiple times. The phantom comprised five rods with $2.5 \times 2.5\,\mathrm{mm}^2$ cross-sections and lengths of 20, 17.5, 15, 8.75 and 5 mm arranged in a meandering ('snake') pattern. The phantom measurement was averaged 1000 times. SM measurements were averaged 10 and 330 times. The 330-average SM serves as a reference for denoising. The reconstruction used the first two receive channels with SNR threshold of 1.5 applied to select frequency components, and the inverse problem was solved using 1000 Kaczmarz iterations with Tikhonov regularization ($\lambda = 0.3$).

Accelerated calibration: For accelerated calibration, the resolution phantom (see figure 2(b)) was measured using a 3D Lissajous sequence (Knopp *et al* 2020). The SM was acquired on a $19 \times 19 \times 19$ Cartesian grid covering a $24 \times 24 \times 12\,\mathrm{mm}^3$ FOV. The phantom consists of five tubes originating from a common point and extending at different angles (two at approximately 20° and 30° in the *xy*-plane, two at approximately 10° and 15° in the *yz*-plane, plus one along the *y*-axis). Both the calibration and the phantom measurements were averaged 1000 times. To evaluate the proposed methods, the acquired SM

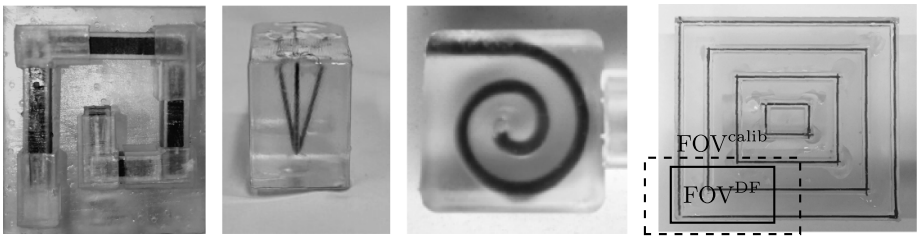

Figure 2. Photographs of the phantoms used for evaluation: (a) snake (denoising); (b) resolution (accelerated calibration); (c) spiral (upsampling); and (d) rectangular (inpainting).

was downsampled along each dimension by the factor of two starting from the first index which resulted in a $10 \times 10 \times 10$ Cartesian grid. Similarly to the original reconstruction, we used frequency selection with SNR threshold of 3.0, and 3 Kaczmarz iterations with Tikhonov regularization ($\lambda = 0.001$).

Upsampling: For upsampling, the spiral phantom (see figure 2(c)) was measured using a 2D Lissajous sequence in the *xy*-plane (Mohn *et al* 2024). The SM was acquired on a 29×29 Cartesian grid covering a $29 \times 29\,\mathrm{mm}^2$ FOV. The phantom comprised a spiral with two full windings. The phantom measurement was averaged 500 times; SM measurements were averaged 1000 times. The data were filtered with an SNR threshold of 1.5 and reconstructed with 1000 Kaczmarz iterations with Tikhonov regularization ($\lambda = 0.4$).

Inpainting: For inpainting, one SM corresponding to a corrupted patch from 3D multi-patch measurements of a nested rectangular phantom (see figure 2(d)) was used (Szwargulski *et al* 2019). This experiment employed $G_x = G_y = -0.75\,\mathrm{T\,m^{-1}}\,\mu_0^{-1}$ and $G_z = 1.5\,\mathrm{T\,m^{-1}}\,\mu_0^{-1}$ and used ferucarbotran (Resovist, I'rom Pharmaceuticals, Tokyo, Japan) as tracer. The SM was acquired on a $25 \times 21 \times 27$ Cartesian grid covering a $50 \times 42 \times 27\,\mathrm{mm}^3$ FOV. The phantom consists of four square-shaped nested tubes in the *xz*-plane, the lower left part of which lies within the FOV. The inpainting mask was generated manually to cover corrupted calibration positions. The SM and phantom measurements were averaged 50 and 100 times, respectively. We reconstructed the selected patch according to equation (6) with an SNR threshold of 10 and 10 iterations of the Kaczmarz method with Tikhonov regularization ($\lambda = 10^{-5}$).

3.5. Ablation study: impact of simulation model on transferability to real data

Generally, the accuracy of the simulation model should directly impact the transferability of the trained DL models to real data. In this study, we select the equilibrium model with anisotropy to balance between the computational complexity and the accuracy of the resulting SM. To demonstrate the importance of the simulation model, we perform an ablation study for the denoising problem in 2D. Using the same parameters for particle, scanner, and calibration groups, we generate a dataset using the equilibrium Langevin model. In this case, the parameters related to anisotropy are ignored. With the same hyperparameters, we train RDN on the dataset generated with the Langevin model, referred to as RDN-Langevin. For quantitative analysis we employ the test set of SMs generated using the equilibrium model with anisotropy excluding immobilized particles. For qualitative analysis, we apply the trained model to the measured SM corresponding to the snake phantom.

4. Results

4.1. Denoising

Table 2 summarizes the quantitative results for the denoising task on simulated data with varying noise levels. Compared to the classical DCT-F method, the learning-based approaches achieve substantially higher performance across all noise levels, with PSNR improvements exceeding 10 dB and SSIM gains of up to 0.15 in most cases. For 2D Lissajous trajectories, the SwinIR model consistently achieves the highest PSNR and SSIM values at every noise level, followed by RDN and DnCNN. This ranking mirrors observations reported on natural-image benchmarks, indicating the transferability of these models to the MPI domain. For 3D Lissajous trajectories, RDN performs best in the absence of a 3D SwinIR counterpart, achieving quality levels comparable to those of the 2D setting.

Table 2. Denoising results on simulated SM data with varying noise levels (σ). Mean ± 95% CI of PSNR/SSIM are reported for DCT-F, DnCNN, RDN, and SwinIR methods on 2D and 3D Lissajous trajectories.

σ	Metric	DCT-F	DnCNN	RDN	SwinIR
			2D		
0.06	PSNR	25.0 ± 0.1	38.9 ± 0.3	40. 6± 0.3	41.1 ± 0.3
	SSIM	0.841 ± 0.009	0.989 ± 0.001	0.992 ± 0.001	0.993 ± 0.001
0.10	PSNR	21.9 ± 0.1	35.7 ± 0.3	37.5 ± 0.3	37.8 ± 0.3
	SSIM	0.76 ± 0.01	0.978 ± 0.003	0.986 ± 0.002	0.986 ± 0.002
0.20	PSNR	18.1 ± 0.1	31.1 ± 0.3	33.1 ± 0.3	33.3 ± 0.3
	SSIM	0.57 ± 0.01	0.947 ± 0.005	0.966 ± 0.004	0.968 ± 0.003
0.30	PSNR	16.0 ± 0.1	28.1 ± 0.3	30.4 ± 0.3	30.6 ± 0.3
	SSIM	0.42 ± 0.01	0.904 ± 0.007	0.942 ± 0.005	0.946 ± 0.004
			3D		
0.06	PSNR	28.01 ± 0.08	37.23 ± 0.09	38.30 ± 0.09	—
	SSIM	0.914 ± 0.003	0.9876 ± 0.0008	0.9898 ± 0.0008	—
0.10	PSNR	25.23 ± 0.08	34.52 ± 0.09	35.79 ± 0.09	—
	SSIM	0.845 ± 0.004	0.978 ± 0.001	0.9827 ± 0.0010	—
0.20	PSNR	21.74 ± 0.09	30.49 ± 0.09	32.10 ± 0.09	—
	SSIM	0.692 ± 0.006	0.947 ± 0.002	0.961 ± 0.002	—
0.30	PSNR	19.77 ± 0.09	27.83 ± 0.09	29.58 ± 0.10	—
	SSIM	0.558 ± 0.007	0.906 ± 0.003	0.930 ± 0.003	—

The results on real data are presented in figure 3. In contrast to the simulated test data, the SNR of real SMs varies across frequencies. For frequency components with higher SNR, all methods, including the classical DCT-F approach, produce visually similar results. For decreased SNR, the DL-based methods achieve superior denoising performance but may introduce artifacts for extremely noisy components. However, since the absolute signal energy in these components is close to zero, their influence on the final reconstruction is negligible. In these cases, the DCT-F method tends to retain the noise present in the input data. In the reconstructed images, all three learning-based models provide comparable noise suppression, with only minor differences visible in the upper region of the snake phantom. Overall, reconstructions obtained from denoised SMs appear slightly more blurred than the ground-truth reference reconstructed from high-SNR data.

4.2. Accelerated calibration

The quantitative comparison of SMRnet and tricubic interpolation for the accelerated calibration problem is presented in table 3. The methods show comparable performance, with overall PSNR and SSIM values remaining relatively low (SSIM barely exceeds 0.6, while PSNR stays below 22 dB). Notably, the performance of both methods degrades for SMs with lower average sizes. SMRnet, which was also trained to suppress noise, shows higher robustness to it.

Both methods were applied to the downsampled SM of the resolution phantom from the Open MPI dataset. Figure 4 shows selected frequency components and central slices of the corresponding reconstructions along each spatial axis. Compared with tricubic interpolation, SMRnet reduces noise and alters the phase in the frequency components. The reconstructions are of comparable quality.

4.3. Upsampling

Table 4 presents the evaluation results on the simulated test set, both with and without added noise, comparing bicubic interpolation and the SMRnet model. For both scaling factors ($\times 2$ and $\times 4$) in the absence of noise, SMRnet consistently outperforms bicubic interpolation by a large margin (approximately 25 dB higher PSNR and 0.08 higher SSIM). Although overall performance decreases in the presence of noise, the relative ranking between the methods remains unchanged.

We evaluate the trained models on measurements of the spiral phantom. The original SM with a spatial size of 29×29 is first denoised using the SwinIR method and subsequently upsampled to higher resolutions using both bicubic interpolation and the SMRnet model. Selected frequency components and the corresponding reconstructions are shown in figure 5. A noticeable denoising effect is observed in both the frequency components and the reconstructed images. SMRnet demonstrates good generalization to real data, resulting in improved perceived image quality. However, SMRnet and bicubic interpolation

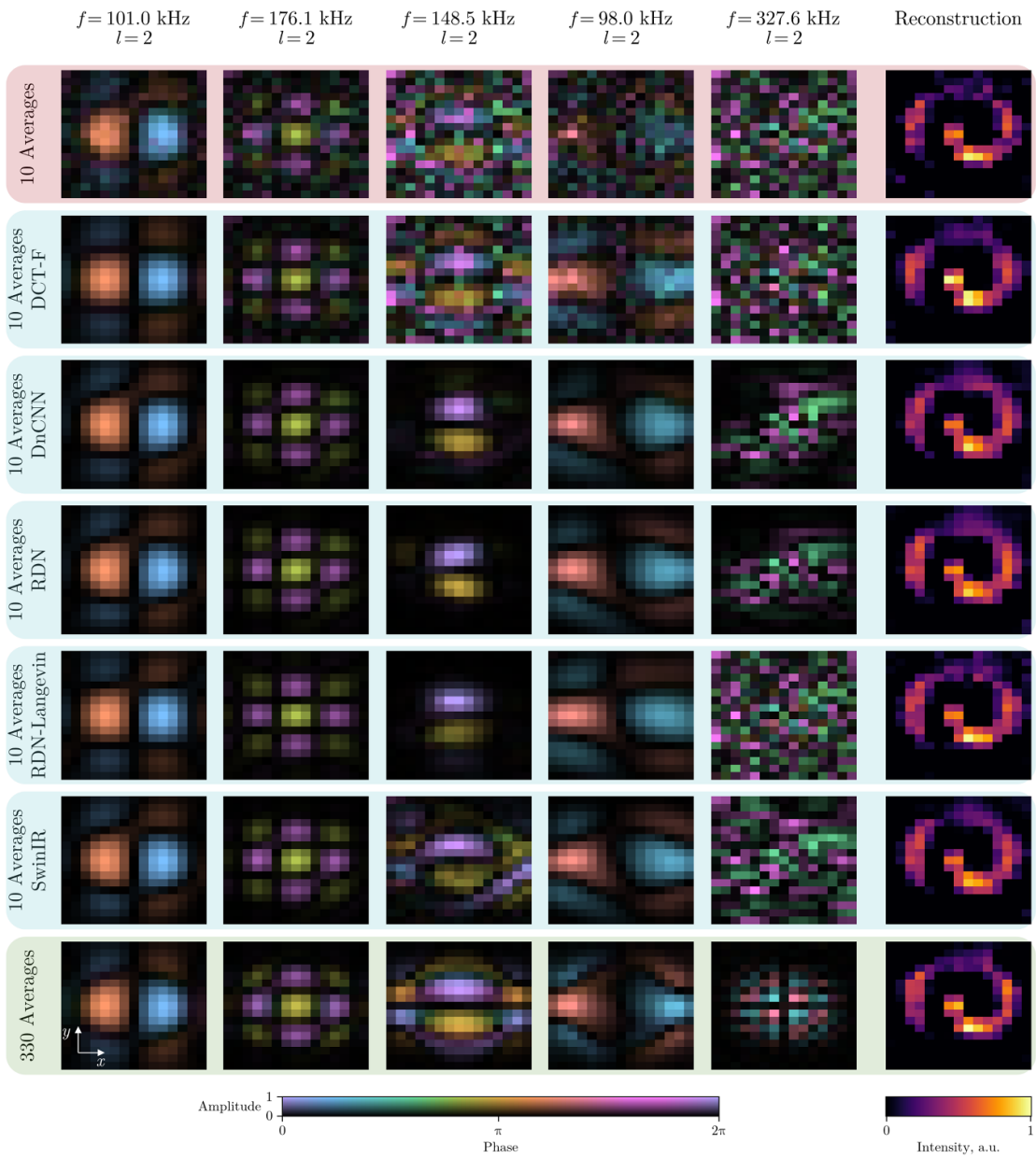

Figure 3. Comparison of denoising methods (DCT-F, DnCNN, RDN, RDN-Langevin, and SwinIR) on a measured 2D SM corresponding to the snake phantom. Selected frequency components from receive coil $l = 2$ and corresponding reconstructions are shown in the *xy*-plane.

Table 3. The results for accelerated calibration on simulated 3D data for tricubic interpolation and SMRnet methods. Mean ± 95% CI of PSNR and SSIM are reported across different average size ranges and noise levels (σ).

		Tricubic		SMRnet	
σ	Avg. size range	PSNR	SSIM	PSNR	SSIM
0.0	(12, 18]	17.1 ± 0.1	0.353 ± 0.009	16.6 ± 0.1	0.24 ± 0.01
	(18, 24]	19.1 ± 0.2	0.44 ± 0.01	19.4 ± 0.2	0.40 ± 0.01
	(24, 30]	21.6 ± 0.1	0.601 ± 0.010	21.5 ± 0.1	0.51 ± 0.01
	all	18.75 ± 0.09	0.452 ± 0.006	18.74 ± 0.10	0.383 ± 0.007
0.1	(12, 18]	16.5 ± 0.1	0.326 ± 0.008	16.5 ± 0.1	0.24 ± 0.01
	(18, 24]	18.2 ± 0.1	0.38 ± 0.01	19.2 ± 0.2	0.39 ± 0.02
	(24, 30]	19.83 ± 0.10	0.461 ± 0.010	21.7 ± 0.1	0.53 ± 0.01
	all	17.86 ± 0.07	0.392 ± 0.005	18.69 ± 0.10	0.378 ± 0.007

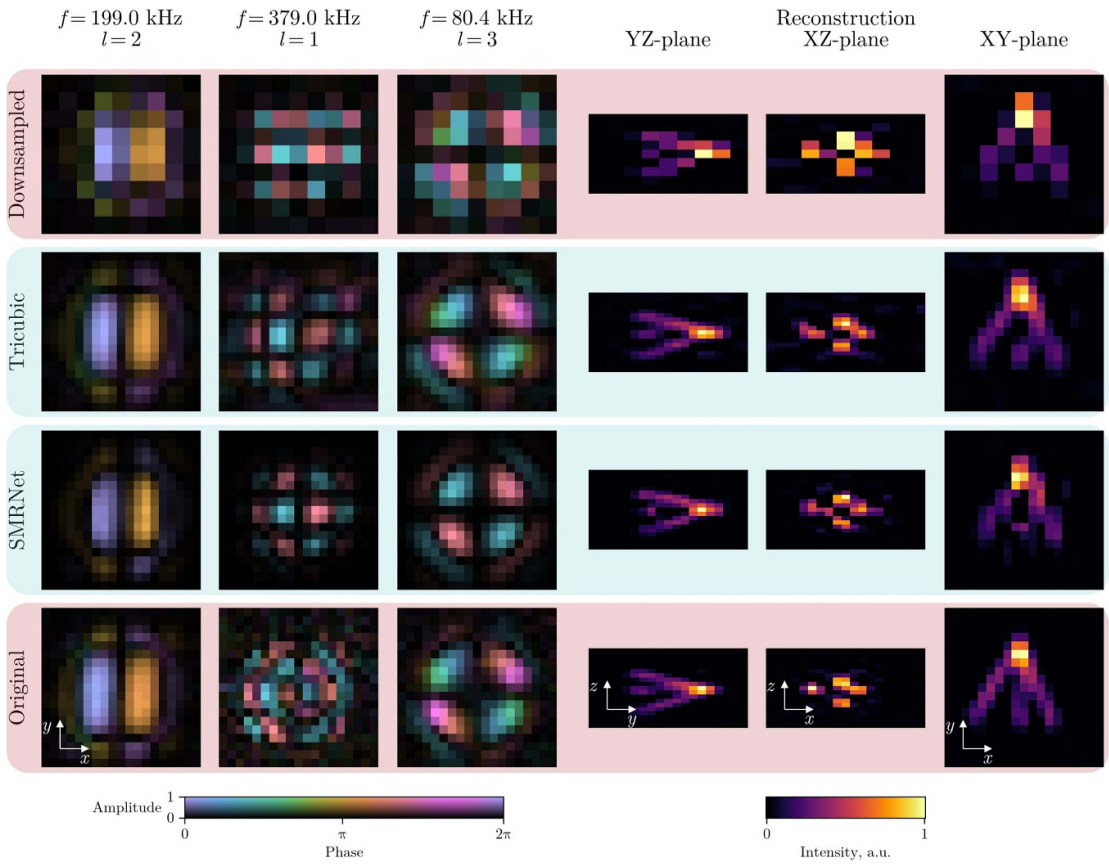

Figure 4. Comparison of tricubic interpolation and SMRnet for restoring a downsampled measured 3D SM corresponding to the resolution phantom. Selected frequency components from receive coils $l \in \{1, 2, 3\}$ in *xy*-plane and central slices of the reconstructions in the *xy*-, *xz*-, and *yz*-planes are shown.

Table 4. Upsampling results on simulated 2D data for bicubic interpolation and SMRnet methods. Mean ± 95% CI of PSNR/SSIM are reported across scaling factors and noise levels (σ).

		Bicubic		SMRnet	
σ	Scale	PSNR	SSIM	PSNR	SSIM
0.0	×2	28.4 ± 0.4	0.911 ± 0.010	54.0 ± 0.6	0.993 ± 0.002
	×4	26.2 ± 0.4	0.87 ± 0.01	52.4 ± 0.7	0.992 ± 0.003
0.1	×2	21.9 ± 0.2	0.58 ± 0.01	38.9 ± 0.3	0.976 ± 0.004
	×4	21.1 ± 0.2	0.47 ± 0.01	38.1 ± 0.4	0.963 ± 0.005

exhibit only minor visual differences in both the frequency components and the reconstructions compared to the overall upsampling effect.

4.4. Inpainting

Table 5 summarizes the quantitative evaluation of inpainting performance on simulated SM data with a fixed 10% mask ratio across average SM size ranges and noise levels (σ). For noiseless SMs ($\sigma = 0.0$), biharmonic inpainting achieves considerably higher PSNR and SSIM values (up to 34.1 dB and 0.94, respectively). However, its performance degrades substantially with increasing noise (PSNR $\approx$ 21.6 dB at $\sigma = 0.1$). In contrast, PConvUNet, trained jointly for inpainting and denoising, exhibits greater robustness, maintaining relatively stable performance across noise levels. Furthermore, PConvUNet demonstrates improved results for larger SMs, suggesting a sensitivity to the spatial scale.

In addition to the simulated evaluation, both methods were applied to real calibration data. Figure 6 shows two representative frequency components and the corresponding reconstructions for selected planes of the rectangular phantom. The corrupted calibration data exhibit planar artifacts that result in missing pixels in the reconstructed volume. After masking the invalid pixels and performing inpainting, both methods successfully restore the missing information and improve the reconstruction quality.

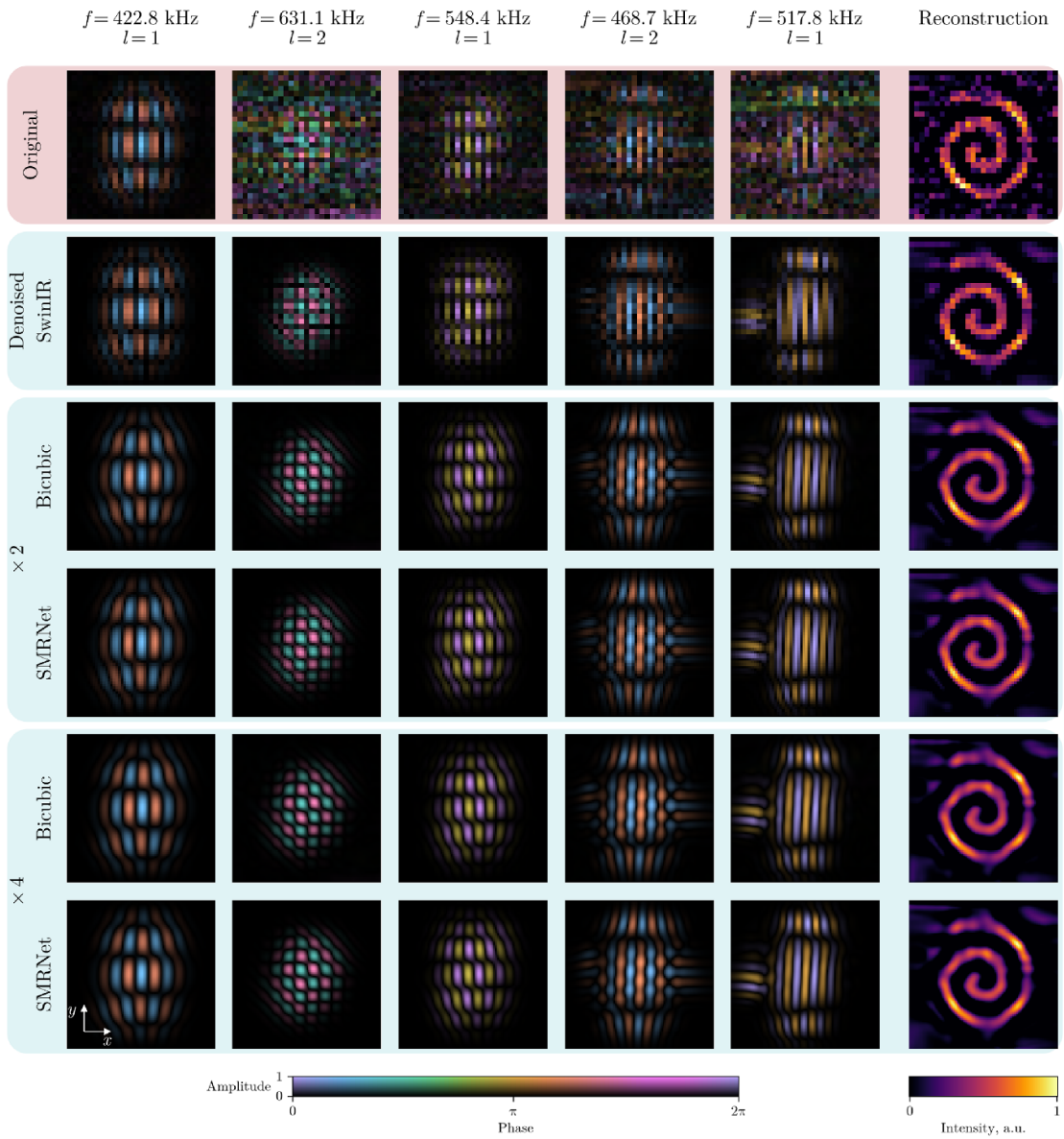

Figure 5. Comparison of bicubic interpolation and SMRnet for upsampling a measured 2D SM of the spiral phantom after SwinIR denoising. Selected frequency components from receive coils $l \in \{1, 2, 3\}$ and reconstructions are shown for both methods at factors 2 and 4.

Table 5. Inpainting results on simulated SM data with 10% mask ratio. Mean ± 95% CI of PSNR/SSIM are reported for biharmonic and PConvUNet methods across SM size ranges and noise levels (σ).

		Biharmonic		PConvUNet	
σ	Avg. size range	PSNR	SSIM	PSNR	SSIM
0.0	(12, 18]	30.3 ± 0.5	0.930 ± 0.004	18.3 ± 0.2	0.45 ± 0.01
	(18, 24]	31.4 ± 0.5	0.936 ± 0.004	22.1 ± 0.2	0.63 ± 0.02
	(24, 30]	37.7 ± 0.7	0.959 ± 0.003	27.7 ± 0.3	0.81 ± 0.01
	all	34.1 ± 0.3	0.943 ± 0.001	22.7 ± 0.1	0.645 ± 0.007
0.1	(12, 18]	21.40 ± 0.09	0.847 ± 0.005	18.3 ± 0.2	0.44 ± 0.01
	(18, 24]	21.9 ± 0.1	0.795 ± 0.008	22.0 ± 0.2	0.62 ± 0.02
	(24, 30]	22.1 ± 0.1	0.627 ± 0.008	27.4 ± 0.3	0.80 ± 0.01
	all	21.60 ± 0.04	0.653 ± 0.005	22.7 ± 0.1	0.641 ± 0.006

Trained jointly for denoising, the PConvUNet also modifies pixels outside the masked regions, which in some cases introduces minor artifacts. Nevertheless, the PConvUNet reconstruction appears less blurry than that of the biharmonic interpolation, likely due to its additional denoising capability.

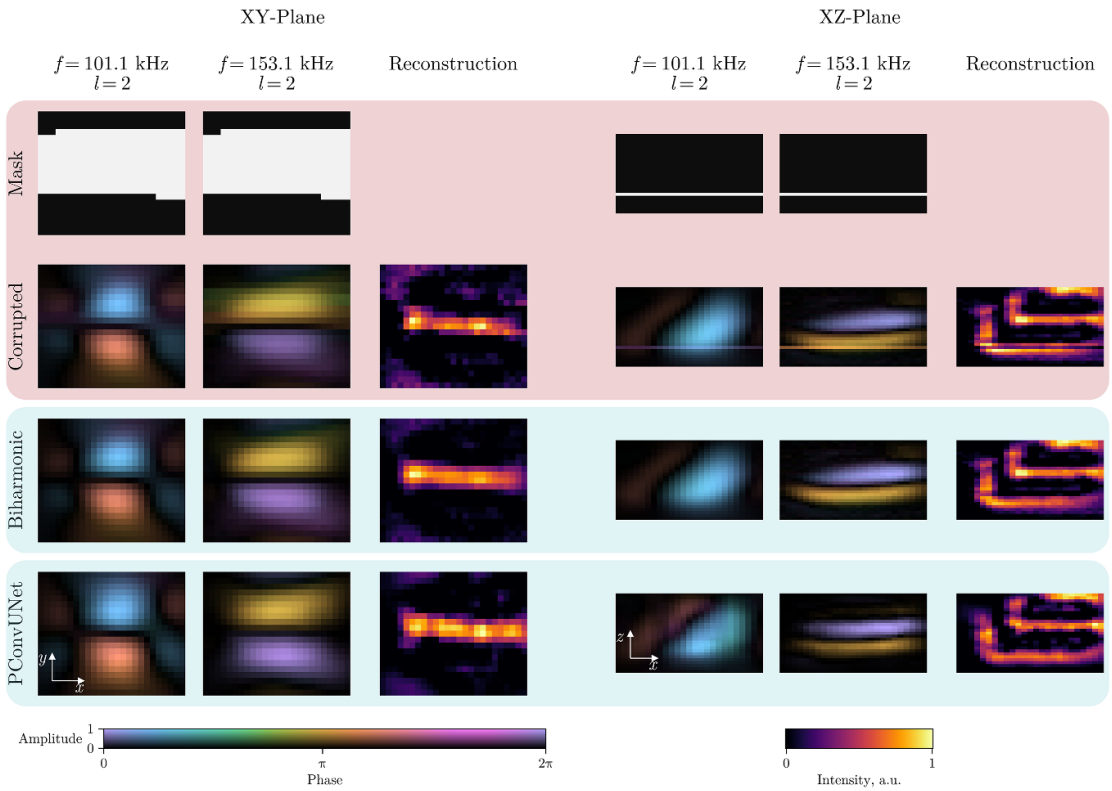

Figure 6. Comparison of biharmonic interpolation and PConvUNet for inpainting a corrupted 3D SM. Selected frequency components from receive coil $l = 2$ and reconstructions are shown in the *xy*- and *xz*-planes for $z = 7$ and $y = 12$ respectively.

Table 6. Comparison of 2D system-matrix denoising performance for the RDN and RDN-Langevin models. The models are trained on simulated data generated with the equilibrium model with anisotropy and the Langevin equilibrium model, respectively. Reported values are mean ± 95% CI of PSNR/SSIM over noise levels (σ) on the test set described in section 3.1, excluding immobilized cases.

σ	Metric	RDN	RDN-Langevin
0.06	PSNR	39.67 ± 0.4	31.3 ± 0.5
	SSIM	0.992 ± 0.001	0.94 ± 0.01
0.10	PSNR	36.8 ± 0.4	29.3 ± 0.5
	SSIM	0.983 ± 0.003	0.91 ± 0.01
0.20	PSNR	32.2 ± 0.4	26.2 ± 0.4
	SSIM	0.959 ± 0.006	0.86 ± 0.02
0.30	PSNR	29.5 ± 0.4	24.4 ± 0.4
	SSIM	0.929 ± 0.008	0.81 ± 0.02

4.5. Ablation study: impact of simulation model on transferability to real data

Table 6 presents the results of the ablation study assessing the impact of the simulation model. The RDN-Langevin model trained on the dataset generated with the Langevin model does not generalize to the test set containing cases with anisotropy, resulting in lower performance compared to the RDN model trained on the dataset generated with the equilibrium model with anisotropy. The gap in SSIM values is more pronounced for higher noise levels, where more prior knowledge about the data distribution is required for effective denoising.

When applied to the real SM of the snake phantom (see figure 3), the RDN-Langevin model shows similar behavior: the RDN-Langevin model tends to produce slightly more symmetric frequency components with the absence of anisotropies, which are present in the RDN model and the original SM. The resulting reconstruction using SM obtained with the RDN-Langevin model appears more noisy.

5. Discussion

This study shows that simulated SMs are suitable for training DL models for multiple SM restoration tasks and that these models qualitatively generalize to measured data. The largest and most robust gains over non-learning baseline methods were observed for denoising, where learning-based methods improved both frequency components and reconstructions. Upsampling yielded modest improvements on measured data, whereas accelerated calibration and inpainting benefited most when combined with denoising, which also provided phase correction and noise suppression.

Quantitative evaluation on measured SMs is limited by the lack of GT and the small number of available datasets. Consequently, quantitative analyses were performed on simulations, acknowledging a potential domain shift between simulated and measured distributions. On simulations, learning-based methods clearly outperformed classical baseline methods for denoising and upsampling; however, for upsampling the numerical advantage did not consistently translate into perceptual gains on measurements. Two factors likely contribute: (i) de-aliasing patterns learned by the model are partially obscured by measurement noise, and (ii) a preceding denoising step can shift the data distribution. For accelerated calibration, performance was strongly dependent on the calibration grid size and noise level. Tricubic interpolation showed minor improvements over SMRnet in the absence of noise, but this trend reversed for noisy SMs, where SMRnet proved more robust. The performance gap between the methods under noisy conditions increased with calibration grid size. Since measured SMs always contain noise and accelerated calibration is particularly relevant for larger calibration grids, the DL-based method may offer greater practical value. Despite low PSNR and SSIM values, the reconstructions of the resolution phantom in figure 4 sufficiently captured the phantom structure. This suggests that pixel-wise performance metrics measured for frequency components do not fully reflect the quality of the reconstructed images, as the reconstruction process itself can compensate for imperfections in the SM. For inpainting, a similar behavior was observed: the learning-based method outperformed the classical approach for noisy SMs and larger calibration grids. The results indicate that classical methods may still be effective for frequency components with high SNR. However, for the entire SM, which typically exhibits varying SNR across frequencies, the learning-based method can provide more consistent performance across frequencies as it is more robust to noise. For both accelerated calibration and inpainting, future works may explore size-aware conditioning or stratified training to further improve performance across varying calibration grid sizes.

Reconstruction can compensate for imperfections in the SM through priors (e.g. Tikhonov regularization) and SNR-based frequency selection, which can mask the effects of SM restoration. For this reason, the reconstruction parameters were fixed across methods for each measurement. Otherwise, any improvement cannot be attributed specifically to the SM restoration method, because the reconstruction also affects the result. Nonetheless, restored SMs enable lowering the SNR threshold and regularization strength while maintaining image quality, thereby reducing bias and leveraging more frequency components. Additionally, for measurements with low SNR, recovering the SM components at frequencies with low signal may not improve reconstruction quality. However, if the measurements have high SNR, the method gives the ability to increase SNR of a SM without performing additional averaging or sampling more calibration points, thus reducing the SM calibration time.

As MPI remains a preclinical imaging modality, available measurements are limited and do not permit a rigorous quantitative evaluation of restoration-induced improvements in image quality. Nevertheless, all cases considered in this work offer potential reductions in the calibration time when restoration methods are applied. Hereafter, the calibration time is estimated by neglecting robot motion and computing the product of the number of grid positions, the repetition time, and the number of averages. In this study, the repetition times are 652.8 µs for the 2D sequence and 21.54 ms for the 3D sequence. For the snake phantom (denoising), the calibration time could be reduced from 55 s to 2 s, because it scales linearly with the number of averages. For the resolution phantom (accelerated calibration), the calibration time could be reduced from 4.1 h to 31 min. For the rectangular phantom (inpainting), recovery of the corrupted SM would save 4.2 h, since such measurements would otherwise typically be discarded.

In this study, the equilibrium model with uniaxial anisotropy was selected as it offered a favorable trade-off between computational cost and applicability to real data. Other simulation methods, such as Langevin or Fokker–Planck, can also be used for data generation (Albers *et al* 2022, Knopp *et al* 2023, Kayapinar *et al* 2024, Faldum *et al* 2025). However, under extreme corruption levels, models may behave generatively and introduce artifacts. If relevant physical effects are not included in the simulation, generalization to measurements exhibiting those effects is impaired, as shown by the ablation study.

The method was restricted to 2D and 3D Lissajous trajectories on an FFP scanner with specific DF frequencies, but the approach can be extended to other sequences or systems, such as Cartesian trajectories or FFL systems. While background noise is scanner-specific, it can be re-measured via empty-frame acquisitions, making the procedure practical across systems; the simulated GT SMs remain transferable. Although measured noise characterizes the system most accurately, simplified parametric noise models may be sufficient for model generalization and remain a topic for future work.

Data-efficient models trained on small amounts of real data may outperform simulation-trained models in narrow domains. The proposed approach is most valuable when self-supervised learning is infeasible or GT is unavailable (e.g. denoising and upsampling). Additionally, large simulated datasets enable pre-training of high-capacity models that can later be fine-tuned on limited real data.

Future work may address the need for a comprehensive collection of MPI SM measurements to enable rigorous quantitative evaluation of data-driven methods, as well as extend the proposed approach to additional acquisition trajectories and scanner types.

6. Conclusion

We have found that simulated SMs can be a valid source of training data for DL methods solving SM restoration problems. The proposed data generation method has the potential to improve the quality of existing DL methods that suffer from data scarcity or develop new methods for the problems where GT data were not available before.

Acknowledgment

We thank Martin Möddel for his valuable comments and feedback during the preparation of this manuscript.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.15480/882.16265.

Funding

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1615 - 503850735.

ORCID iDs

Artyom Tsanda 0009-0009-7765-4604
Sarah Reiss 0009-0006-4015-1200
Konrad Scheffler 0000-0002-6842-9204
Marija Boberg 0000-0003-3419-7481
Tobias Knopp 0000-0002-1589-8517

References

Albers H, Kluth T and Knopp T 2022 Simulating magnetization dynamics of large ensembles of single domain nanoparticles: numerical study of Brown/Néel dynamics and parameter identification problems in magnetic particle imaging *J. Magn. Magn. Mater.* **541** 168508

Askin B, Güngör A, Soydan D A, Saritas E U, Bariş Top C and Cukur T 2022 PP-MPI: a deep plug-and-play prior for magnetic particle imaging reconstruction *Machine Learning for Medical Image Reconstruction* (*Lecture Notes in Computer Science*) (Springer) pp 105–14

Baltruschat I M, Szwargulski P, Griese F, Grosser M, Werner R and Knopp T 2020 3d-SMRnet: achieving a new quality of MPI system matrix recovery by deep learning *Medical Image Computing and Computer Assisted Intervention – Miccai 2020* (Springer) pp 74–82

Bariş Top C and Güngör A 2020 Tomographic field free line magnetic particle imaging with an open-sided scanner configuration *IEEE Trans. Med. Imaging* **39** 4164–73

Chui C K and Mhaskar H N 2010 MRA contextual-recovery extension of smooth functions on manifolds *Appl. Comput. Harmon. Anal.* **28** 104–13
Damelin S B and Hoang N S 2018 On surface completion and image inpainting by biharmonic functions: numerical aspects *Int. J. Math. Math. Sci.* **2018** 3950312
Deissler R J, Wu Y and Martens M A 2014 Dependence of Brownian and Néel relaxation times on magnetic field strength *Med. Phys.* **41** 012301
Faldum M, Bachmayr M, Schulz V and Schrank F 2025 Efficient solvers for coupled Brown-Néel Fokker-Planck equations *Int. J. Magn. Part. Imaging* **11** 11
Gleich B and Weizenecker J 2005 Tomographic imaging using the nonlinear response of magnetic particles *Nature* **435** 1214–7
Goodwill P W and Conolly S M 2011 Multidimensional X-space magnetic particle imaging *IEEE Trans. Med. Imaging* **30** 1581–90
Grosser M, Möddel M and Knopp T 2020 Using low-rank tensors for the recovery of MPI system matrices *IEEE Trans. Comput. Imaging* **6** 1389–402
Güngör A, Askin B, Soydan D A, Bariş Top C and Cukur T 2021 Deep learned super resolution of system matrices for magnetic particle imaging *2021 43rd Annual Int. Conf. of the IEEE Engineering in Medicine & Biology Society (EMBC)* pp 3749–52
Güngör A, Askin B, Soydan D A, Saritas E U, Baris Top C and Cukur T 2022 TranSMS: transformers for super-resolution calibration in magnetic particle imaging *IEEE Trans. Med. Imaging* **41** 3562–74
Haegele J, Rahmer J, Gleich B, Borgert J, Wojtczyk H, Panagiotopoulos N, Buzug T M, Barkhausen J and Vogt F M 2012 Magnetic particle imaging: visualization of instruments for cardiovascular intervention *Radiology* **265** 933–8
Healy S, Bakuzis A F, Goodwill P W, Attaluri A, Bulte J W M and Ivkov R 2022 Clinical magnetic hyperthermia requires integrated magnetic particle imaging *WIREs Nanomed. Nanobiotechnol.* **14** e1779
Herz S, Vogel P, Kampf T, Dietrich P, Veldhoen S, Rückert M A, Kickuth R, Behr V C and Bley T A 2019 Magnetic particle imaging-guided stenting *J. Endovasc. Therapy* **26** 512–9
Ilbey S, Baris Top C, Güngör A, Cukur T, Saritas E U and Güven H E 2019 Fast system calibration with coded calibration scenes for magnetic particle imaging *IEEE Trans. Med. Imaging* **38** 2070–80
Kayapinar M H, Alpman A and Saritas E U 2024 Fourier neural operator for coupled brown-Neel rotation model *Int. J. Magn. Part. Imaging* **10** 758
Kingma D P and Ba. J 2015 Adam: a method for stochastic optimization *3rd Int. Conf. on Learning Representations*
Kluth T 2018 Mathematical models for magnetic particle imaging *Inverse Problems* **34** 083001
Kluth T, Bathke C, Jiang M and Maass P 2020 Joint super-resolution image reconstruction and parameter identification in imaging operator: analysis of bilinear operator equations, numerical solution and application to magnetic particle imaging *Inverse Problems* **36** 124006
Knopp T, Albers H, Grosser M, Möddel M and Kluth T 2023 Exploiting the Fourier neural operator for faster magnetization model evaluations based on the Fokker-Planck equation *Int. J. Magn. Part. Imaging* **9** 597
Knopp T, Gdaniec N and Möddel M 2017 Magnetic particle imaging: from proof of principle to preclinical applications *Phys. Med. Biol.* **62** R124
Knopp T, Rahmer J, Sattel T F, Biederer S, Weizenecker J, Gleich B, Borgert J and Buzug T M 2010b Weighted iterative reconstruction for magnetic particle imaging *Phys. Med. Biol.* **55** 1577
Knopp T, Sattel T F, Biederer S and Buzug T M 2010a Limitations of measurement-based system functions in magnetic particle imaging *Medical Imaging 2010: Biomedical Applications in Molecular, Structural and Functional Imaging* vol 7626 (SPIE) pp 427–34
Knopp T, Sattel T F, Biederer S, Rahmer J, Weizenecker J, Gleich B, Borgert J and Buzug T M 2010c Model-based reconstruction for magnetic particle imaging *IEEE Trans. Med. Imaging* **29** 12–18
Knopp T, Szwargulski P, Griese F and Gräser M 2020 OpenMPIData: an initiative for freely accessible magnetic particle imaging data *Data in Brief* **28** 104971
Knopp T and Weber A 2013 Sparse reconstruction of the magnetic particle imaging system matrix *IEEE Trans. Med. Imaging* **32** 1473–80
Liang J, Cao J, Sun G, Zhang K, Van Gool L and Timofte R 2021 SwinIR: image restoration using Swin transformer *Proc. IEEE/CVF International Conference on Computer Vision* vol 2021 pp 1833–44
Liu G, Reda F A, Shih K J, Wang T-C, Tao A and Catanzaro B 2018 Image inpainting for irregular holes using partial convolutions *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol 11215 pp 89–105
Liu S, Bian Z, Tian J, Yang G and Hui H 2025 Deep learning-based noise reduction method for the system matrix in magnetic particle imaging *Phys. Med. Biol.* **70** 235007
Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S and Guo B 2021 Swin transformer: hierarchical vision transformer using shifted windows *2021 IEEE/CVF Int. Conf. on Computer Vision (ICCV)* (IEEE) pp 9992–10002
Ludewig P *et al* 2022 Magnetic particle imaging for assessment of cerebral perfusion and ischemia *WIREs Nanomed. Nanobiotechnol.* **14** e1757
Maass M, Kluth T, Droigk C, Albers H, Scheffler K, Mertins A and Knopp T 2024 Equilibrium model with anisotropy for model-based reconstruction in magnetic particle imaging *IEEE Trans. Comput. Imaging* **10** 1588–601
Micikevicius P *et al* 2018 Mixed precision training *6th Int. Conf. on Learning Representations* (OpenReview.net) (available at: https://openreview.net/forum?id=r1gs9JgRZ)
Mohn F, Scheffler K, Ackers J, Weimer A, Wegner F, Thieben F, Ahlborg M, Vogel P, Graeser M and Knopp T 2024 Characterization of the clinically approved MRI tracer Resotran for magnetic particle imaging in a comparison study *Phys. Med. Biol.* **69** 135014
Mohtashamdolatshahi A, Kratz H, Kosch O, Hauptmann R, Stolzenburg N, Wiekhorst F, Sack I, Hamm B and Taupitzog Jörg Schnorr M 2020 *In vivo* magnetic particle imaging: angiography of inferior vena cava and aorta in rats using newly developed multicore particles *Sci. Rep.* **10** 17247
Molwitz I, Ittrich H, Knopp T, Mummert T, Salamon J, Jung C, Adam G and Kaul M G 2019 First magnetic particle imaging angiography in human-sized organs by employing a multimodal ex vivo pig kidney perfusion system *Physiol. Meas.* **40** 105002
Paysen H, Kosch O, Wells J, Loewa N and Wiekhorst F 2020 Characterization of noise and background signals in a magnetic particle imaging system *Phys. Med. Biol.* **65** 235031
Salamon J *et al* 2016 Magnetic particle/magnetic resonance imaging: in-vitro MPI-guided real time catheter tracking and 4D angioplasty using a road map and blood pool tracer approach *PLoS One* **11** 1–14
Shi G, Yin L, An Y, Li G, Zhang L, Bian Z, Chen Z, Zhang H, Hui H and Tian J 2023 Progressive pretraining network for 3D system matrix calibration in magnetic particle imaging *IEEE Trans. Med. Imaging* **42** 3639–50

Szwargulski P, Möddel M, Gdaniec N and Knopp T 2019 Efficient joint image reconstruction of multi-patch data reusing a single system matrix in magnetic particle imaging *IEEE Trans. Med. Imaging* **38** 932–44

Tay Z W *et al* 2018 Magnetic particle imaging-guided heating *in vivo* using gradient fields for arbitrary localization of magnetic hyperthermia therapy *ACS Nano* **12** 3699–713

Thieben F, Foerger F, Mohn F, Hackelberg N, Boberg M, Scheel J-P, Möddel M, Graeser M and Knopp T 2024 System characterization of a human-sized 3D real-time magnetic particle imaging scanner for cerebral applications *Commun. Eng.* **3** 47

Tsanda A, Jürß P, Hackelberg N, Grosser M, Möddel M and Knopp T 2024 Extension of the Kaczmarz algorithm with a deep plug-and-play regularizer *Int. J. Magn. Part. Imaging* **10** 748

Tsanda A, Reiss S, Scheffler K, Boberg M and Knopp T 2025a Software: Deep Learning for Restoring MPI System Matrices Using Simulated Training Data (available at: https://github.com/IBIResearch/mpi-sm-restoration) (https://doi.org/10.5281/zenodo.20038418)

Tsanda A, Reiss S, Scheffler K, Boberg M and Knopp T 2025b Supplementary data for the paper: "Deep Learning for Restoring MPI System Matrices Using Simulated Training Data (https://doi.org/10.15480/882.16265)

Tsanda A, Scheffler K, Reiss S and Knopp T 2025c Denoising the system matrix with deep neural networks for better MPI reconstructions *Int. J. Magn. Part. Imaging* **11** 810

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12

Weber A and Knopp T 2015 Reconstruction of the magnetic particle imaging system matrix using symmetries and compressed sensing *Adv. Math. Phys.* **2015** 460496

Weber A, Weizenecker J, Heinen U, Heidenreich M and Buzug T M 2015 Reconstruction enhancement by denoising the magnetic particle imaging system matrix using frequency domain filter *IEEE Trans. Magn.* **51** 1–5

Weizenecker J, Gleich B, Rahmer J, Dahnke H and Borgert J 2009 Three-dimensional real-time *in vivo* magnetic particle imaging *Phys. Med. Biol.* **54** L1

Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising *IEEE Trans. Image Process.* **26** 3142–55

Zhang L, Li J, He X, Tang Z, Li S, Zhao Y and Guo H 2025 Iterative up-and-down sampling network based on pyramid pooling and attention mechanism for 3-D system matrix recovery in MPI *IEEE Trans. Instrum. Meas.* **74** 1–12

Zhang Y, Tian Y, Kong Y, Zhong B and Fu Y 2021 Residual dense network for image restoration *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 2480–95