Untrained Physically Informed Neural Network for Image Reconstruction of Magnetic Field Sources

Dubois, A. E.E.; Broadway, D. A.; Stark, A.; Tschudin, M. A.; Healey, A. J.; Huber, S. D.; Tetienne, J. P.; Greplova, E.; Maletinsky, P.

DOI
10.1103/PhysRevApplied.18.064076

Publication date
2022

Document Version
Final published version

Published in
Physical Review Applied

Citation (APA)
Dubois, A. E. E., Broadway, D. A., Stark, A., Tschudin, M. A., Healey, A. J., Huber, S. D., Tetienne, J. P., Greplova, E., & Maletinsky, P. (2022). Untrained Physically Informed Neural Network for Image Reconstruction of Magnetic Field Sources. *Physical Review Applied*, *18*(6), Article 064076. https://doi.org/10.1103/PhysRevApplied.18.064076

Editors' Suggestion

Untrained Physically Informed Neural Network for Image Reconstruction of Magnetic Field Sources

A.E.E. Dubois,[1,2] D.A. Broadway,[1,*] A. Stark,[2] M.A. Tschudin,[1] A.J. Healey,[3,4] S.D. Huber,[5] J.-P. Tetienne,[6] E. Greplova,[7] and P. Maletinsky[1,†]

[1] *Department of Physics, University of Basel, Klingelbergstrasse 82, Basel CH-4056, Switzerland*
[2] *QNAMI AG, Hofackerstrasse 40 B, Muttenz CH-4132, Switzerland*
[3] *School of Physics, University of Melbourne, Victoria 3010, Australia*
[4] *Centre for Quantum Computation and Communication Technology, School of Physics, University of Melbourne, Victoria 3010, Australia*
[5] *Institute for Theoretical Physics, ETH Zurich CH-8093, Switzerland*
[6] *School of Science, RMIT University, Melbourne, Victoria 3000, Australia*
[7] *Kavli Institute of Nanoscience, Delft University of Technology, GA Delft 2600, Netherlands*

(Received 10 May 2022; accepted 9 November 2022; published 26 December 2022)

The prediction of measurement outcomes from an underlying structure often follows directly from fundamental physical principles. However, a fundamental challenge is posed when trying to solve the inverse problem of inferring the underlying source configuration based on measurement data. A key difficulty arises from the fact that such reconstructions often involve ill-posed transformations and that they are prone to numerical artifacts. Here, we develop a numerically efficient method to tackle this inverse problem for the reconstruction of magnetization maps from measured magnetic stray-field images. Our method is based on neural networks with physically inferred loss functions to efficiently eliminate common numerical artifacts. We report on a significant improvement in reconstruction over traditional methods and we show that our approach is robust to different magnetization directions, both in and out of plane, and to variations of the magnetic field measurement-axis orientation. While we showcase the performance of our method using magnetometry with nitrogen-vacancy center spins in diamond, our neural-network-based approach to solving inverse problems is agnostic to the measurement technique and thus is applicable beyond the specific use case demonstrated in this work.

DOI: 10.1103/PhysRevApplied.18.064076

I. INTRODUCTION

Determining the nanoscale magnetic state of materials is crucial to developing a deeper understanding of classical and quantum magnetism [1] and the realization of next-generation spintronics technologies [2,3]. Typically, this information is difficult to achieve directly and requires, e.g., the use of large-scale imaging facilities [4]. An alternative method is to measure the fields emitted from the source, e.g., magnetic fields, and use these fields to infer information about the structure of the source. However, such a reconstruction process often involves solving an inverse problem that can be ill posed and thus prone to significant errors [Fig. 1(a)]. Despite this, by using appropriate assumptions and measurement configurations, the method of direct inversion has been used to probe different regimes of current flow [5–7] and magnetization in two-dimensional (2D) van der Waals materials [8,9]. However, extending these measurements to more exotic materials with complex magnetic structures becomes an issue, as the error from the ill-posed transformation can become larger than the signal itself.

In this work, we develop a methodology for addressing this challenge. We propose to perform the ill-posed transformation using a physically informed neural network (NN), which learns using a well-posed forward transformation [10]. We demonstrate that with our method, many of the issues that arise from traditional approaches can be overcome. While deep learning has already been employed to address inverse problems [11–15], this approach has up to now relied on training of the network with large known data sets. This can result in excellent reconstructions; however, in fields that do not generate large swaths of data, this type of technique is not applicable. While it is sometimes possible to simulate data for training, this inherently carries the risk of training the network to only solve a subset

*davidaaron.broadway@unibas.ch
†patrick.maletinsky@unibas.ch

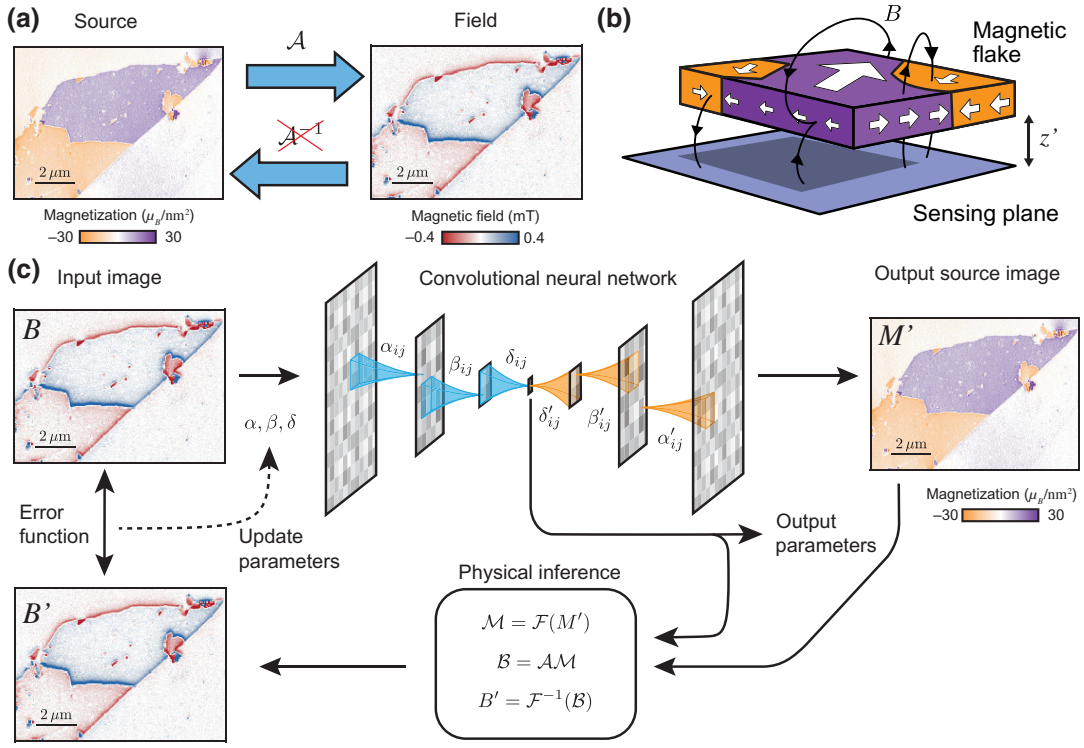

FIG. 1. The direct training method for solving inverse problems. (a) An illustration of a source-and-field reconstruction problem, where the transformation $\mathcal{A}$ is well defined from source to field but the inverse, $\mathcal{A}^{-1}$, is not. (b) An illustration of measurement of a magnetic field in a sensing plane offset from a magnetic material. (c) A diagram of an untrained physically informed neural network (UPINN), where an input field image is put through a convolutional NN, with weightings α, β, δ, etc. at each layer. The reconstructed source image is transformed back into a field image using the standard Fourier transform $\mathcal{F}$ and the well-defined forward transformation $\mathcal{A}$ and this reconstruction is used to generate the error function.

of the total number of problems and thus may converge to biased or nonphysical solutions. Likewise, training NNs on subsets of the total number of problems can result in nonphysical solutions and a lack of generalizability. Conversely, our method does not require prior training, as the forward transformation allows the network to learn the transformation for a given data set. As such, this is a broadly applicable technique for solving ill-posed reverse problems when the forward problem is well defined.

II. THE RECONSTRUCTION PROBLEM

The relationship between 2D sources of magnetization and magnetic fields can be described in the Fourier space by the following expression [16]:

$$\begin{bmatrix} \mathcal{B}_x \\ \mathcal{B}_y \\ \mathcal{B}_z \end{bmatrix} = \underbrace{\frac{-1}{\alpha}\begin{bmatrix} k_x^2/k & k_x k_y/k & ik_x \\ k_x k_y/k & k_y^2/k & ik_y \\ ik_x & ik_y & -k \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} \mathcal{M}_x \\ \mathcal{M}_y \\ \mathcal{M}_z \end{bmatrix}, \quad (1)$$

where $\mathcal{B}$ and $\mathcal{M}$ are the Fourier-transformed magnetic field and magnetization vector, respectively, $\mathcal{A}$ is the transfer matrix, k_x and k_y are the Fourier space coordinates with $k = \sqrt{k_x^2 + k_y^2}$, $\alpha = 2e^{kz'}/\mu_0$ contains an exponential that represents the propagator between the source and measurement planes, and μ_0 is the vacuum permeability. Where we use the assumptions that the magnetization vector, $\mathbf{M}(x, y)$, is confined to a 2D plane and that the stray field is measured in a parallel plane at an approximately known height z', $\mathbf{B}(x, y, z')$ [Fig. 1(b)].

The transformation from a magnetization map to a magnetic field map is a well-posed forward transformation with a unique solution. In contrast, the inverse problem of transforming from a magnetic field map to a magnetization map is ill posed, as the transformation matrix $\mathcal{A}$ is singular ($\det(\mathcal{A}) = 0$), such that there is an infinite number of solutions for $\mathbf{M}$. The problem can be simplified using prior knowledge, e.g., one often assumes that the magnetization has a uniform known direction defined by spherical angles (θ, ϕ), i.e., $\mathbf{M}(x, y) = M_{\theta,\phi}(x, y)\mathbf{\Theta}$, where $\mathbf{\Theta} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. The magnetic field measurement is also generally performed along a single direction $\mathbf{\Theta}_m = (\sin\theta_m\cos\phi_m, \sin\theta_m\sin\phi_m, \cos\theta_m)$, with measured projection B_m. Equation (1) then simplifies to

$$\mathcal{B}_m = \mathbf{\Theta_m} \cdot \mathcal{A} \cdot \mathbf{\Theta}\mathcal{M}_{\theta,\phi}, \quad (2)$$

where the transfer function is now $\mathcal{A}' = \mathbf{\Theta_m} \cdot \mathcal{A} \cdot \mathbf{\Theta}$ (which we denote simply as $\mathcal{A}$ in the following). Even so, this simplified problem is not invertible for certain values in k space, e.g., for $k = 0$, $k_{x,y} = 0$, and $k_x = -k_y \tan\phi$. Moreover, some parameters entering the transfer function may be *a priori* unknown, e.g., for an in-plane magnet ($\theta = 90^\circ$ but ϕ unknown).

Traditionally, two approaches exist to tackle this type of reconstruction. First, the analytical reconstruction tries to model the transformation $\mathcal{A}^{-1}$ using an optimization criterion and prior knowledge. For instance, regularization [17] is an analytical model that attempts to solve the ill-posed problem by minimizing an auxiliary parameter that is added to remove the undefined transformation terms at the potential cost of spatial resolution. Second, one can treat unknown values as being stochastic, where it is then possible to estimate the full posterior $p(x|y)$ with a Bayesian inference [18]. In contrast, our method, which leverages a machine-learning architecture to generate the mapping from field to source, bypasses the need to rely on the ill-posed transformation.

III. SOLVING INVERSE PROBLEMS USING NEURAL NETWORKS

Deep neural networks are characterized as universal approximators having the potential to compute any nonlinear function [20]. Thus, they are good candidates in tackling reconstruction problems [11–14]. Given a data set $\mathbf{x}_i$, the reconstruction task is modeled by

$$\mathbf{y}_i = g_\alpha(\mathbf{x}_i) + \epsilon \tag{3}$$

where $g_\alpha(\mathbf{x}_i)$ is the NN model with parameter set α and ϵ is the noise in the measurement. The goal is to develop a model g_α that will learn to operate as $\mathcal{A}^{-1}$.

While deep learning is the state-of-the-art tool for solving many tasks in natural language processing or computer vision, serious improvements are needed to completely surpass traditional methods for reconstruction tasks [11]. The main issue is that the knowledge provided to the model to learn is constrained to the training data set. In our case, the number of parameters (e.g., the magnetization direction θ, ϕ, the magnetometer stand-off z', and the magnetometer direction θ_m, ϕ_m) influencing the reconstructed magnetization is high. Thus, it requires a large data set to train the network with every possible combination of control parameters. As such, there is a risk of training the network to only solve a subset of the total number of problems, which may lead to the network converging to nonphysical solutions. Moreover, the network is likely to encounter measurements not seen in the training data as well as the presence of unseen noise. This is particularly problematic as even tiny perturbations in the input can produce artifacts in the reconstructed output [21]. Another challenge for deep learning is that the solution is based on a "black box," which leaves out any explanations of the reconstruction process. The estimations on new samples are statistical inferences based on previous learning without warranty on the new output.

In our method, we circumvent these issues by having the neural network learn directly on each image [see Fig. 1(c)]. While this learning procedure is similar to traditional training of a neural network, i.e., the weights of the network are updated according to the loss function, it differs in that the learning itself is never used for a different image. Nor does the network learn from multiple data sets, as in traditional training. Our method removes the reliance on large data sets with ground truths for loss functions. Instead, it introduces a data-specific loss function that is defined by performing the well-posed forward transformation on the ML output, to transform it back into the initially measured experimental quantity. Thus, the error is defined as the difference between the original magnetic field and the reconstructed one. By tailoring the loss function in this manner, prior training of the network is no longer required. Furthermore, the output of the ML code is restricted to being a physically relevant result, as it must reproduce the measured field. We call this method the untrained physically informed neural network (UPINN), as opposed to the traditional pretrained model.

As this method learns directly on a single image, we do not face computational burden. Moreover, we can significantly alleviate the overtraining risk, which usually refers to the inability to generalize well on unseen data due to an overlearning of the statistical distribution of the training data set. Hence, this method can be used with convolutional NN as well as fully connected NN. We opt for the former as it converges faster and minimizes noise in the output image. In this work, we present results obtained with a convolutional NN following a "U-network architecture" [21], details of which are given in Appendix D. We also note that similar methods have been employed to reconstruct holograms using simulations fed from the output of convolutional neural networks (CNNs) [22].

IV. NEURAL-NETWORK SOLUTION CONVERGENCE

To benchmark our UPINN with real data, we first employ it on a problem that is also manageable with regular methods, namely the reconstruction of an out-of-plane magnetized source, i.e., $\theta = 0$ in Eq. (2). The data are taken using a scanning nitrogen-vacancy (N-V) magnetometer with an angle of $(\theta_m, \phi_m) = (54.7^\circ, 282^\circ)$ [9] [Figs. 2(a) and 2(b)]. Our UPINN quickly converges to a mapping that matches that of the standard method [Fig. 2(c)]. The two methods return the same average magnetization of the material, $M_z = 11.5\ \mu_B/\mathrm{nm}^2$, while the

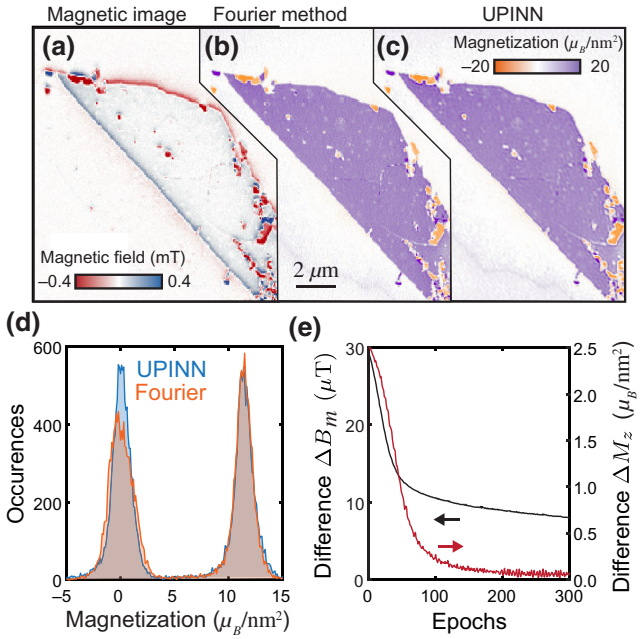

FIG. 2. Testing the neural network with out-of-plane magnetization. (a) The magnetic stray-field image $(\theta_m, \phi_m) = (54.7^\circ, 282^\circ)$ of an out-of-plane magnetized trilayer of the material CrI_3 [9]. (b) A magnetization image produced using the traditional Fourier reconstruction method [16]. (c) A magnetization image produced with the UPINN. (d) The histogram of the magnetization values for the Fourier method (orange) and the UPINN (blue). (e) The mean difference between the original magnetic field and the UPINN reconstruction as a function of the number of epochs (left axis) and the mean difference between the Fourier and UPINN reconstruction methods (right axis).

UPINN more accurately reconstructs the background magnetization, which is observable in the height and width of the zero peak in the histogram in Fig. 2(d). We attribute this difference to a circumvention of some edge-related artifacts that are generated in the Fourier method.

The Fourier method reconstructs based on the observed wave vectors in the magnetic image; as such, it has a loss of information when the magnetic material is not fully contained in the image or if the magnetic field is truncated from the field of view. Our UPINN method, on the other hand, does not have this restriction, as it is fitting the image in order to reconstruct the magnetic field, as such truncation of the magnetic field does not directly affect the reconstruction. Likewise, source materials that cross the boundary of the image do not contribute to a significant loss in the quality of the reconstruction, as there is no magnetic field information to inform the fit to change from the observed source value. In principle, the UPINN method could even be used to attempt to reconstruct sources that exist outside of the field of view by including a spatial model as done in Clement *et al.* [18] but this aspect is beyond the scope of this work.

We define a loss function, $\Delta B_m = \text{mean}(|B_m - B_{\text{NN}}|)$, that reliably estimates the quality of the reconstruction [Fig. 2(e)], reducing the average difference per pixel to < 8 nT ($< 3\%$ of the maximal magnetic field strength) within 300 epochs and mainly localized to individual pixels rather than large areas. Increasing the number of epochs eventually leads to the condition where the UPINN begins to overfit the image and reconstructs the noise directly. While the error function does not reflect this overtraining, it is visually evident and is observed as an increase in the signal-to-noise ratio (SNR) of the output image (for details, see Appendix E) and thus can be selected out *a posteriori*. We compare the UPINN reconstruction to that from the traditional method, $\Delta M_z = \text{mean}(|M_F - M_{\text{UPINN}}|)$ (Fig. 2(e), right axis), where M_F (M_{UPINN}) is the reconstructed image via the Fourier (UPINN) method. This comparison demonstrates that our model converges to a similar solution within 200 epochs, returning an average difference of $\Delta M_z \approx 0.07\ \mu_B\text{nm}^{-2}$ ($< 1\%$ of the maximal magnetization strength) before becoming overfitted, which is observed by the increase in the noise of the difference between the two methods.

V. DETERMINING THE MAGNETIZATION DIRECTION

We now move to solving the more difficult problem of reconstructing the magnetization in a material with an arbitrary but piecewise constant-magnetization direction. These types of more general spin textures generate two issues in traditional reconstruction. First, the in-plane component of the magnetization has more undefined terms than the case of purely out-of-plane magnetization [16]. Second, the introduction of the additional magnetization angles greatly expands the parameter space. Here, we take simulated data in order to reliably define the magnetization angle [Fig. 3(a)] and generate magnetic field images, from which we try to reconstruct the underlying spin texture [Fig. 3(b)]. To begin with, we confine the magnetization to be purely in plane ($\theta = 90^\circ$), which is a valid restriction for materials with known easy-plane magnetic anisotropy. The UPINN performs well in reconstructing the simulated data with a high degree of accuracy and is capable of determining the in-plane angle. Performing this experiment 1000 times with random magnetization angles [Fig. 3(c)] demonstrates that the UPINN does indeed accurately predict the angle, with an uncertainty of $< 2\%$ for an SNR of 20.

In some cases, the experimental conditions or the material quality may lead to a canting of the magnetization direction [Fig. 3(d)], resulting in the magnetization having a component that is out of the plane [Fig. 3(e)]. The UPINN is able to determine the degree of this canting [Fig. 3(f)] with an uncertainty of $< 1\%$ for an SNR of 20.

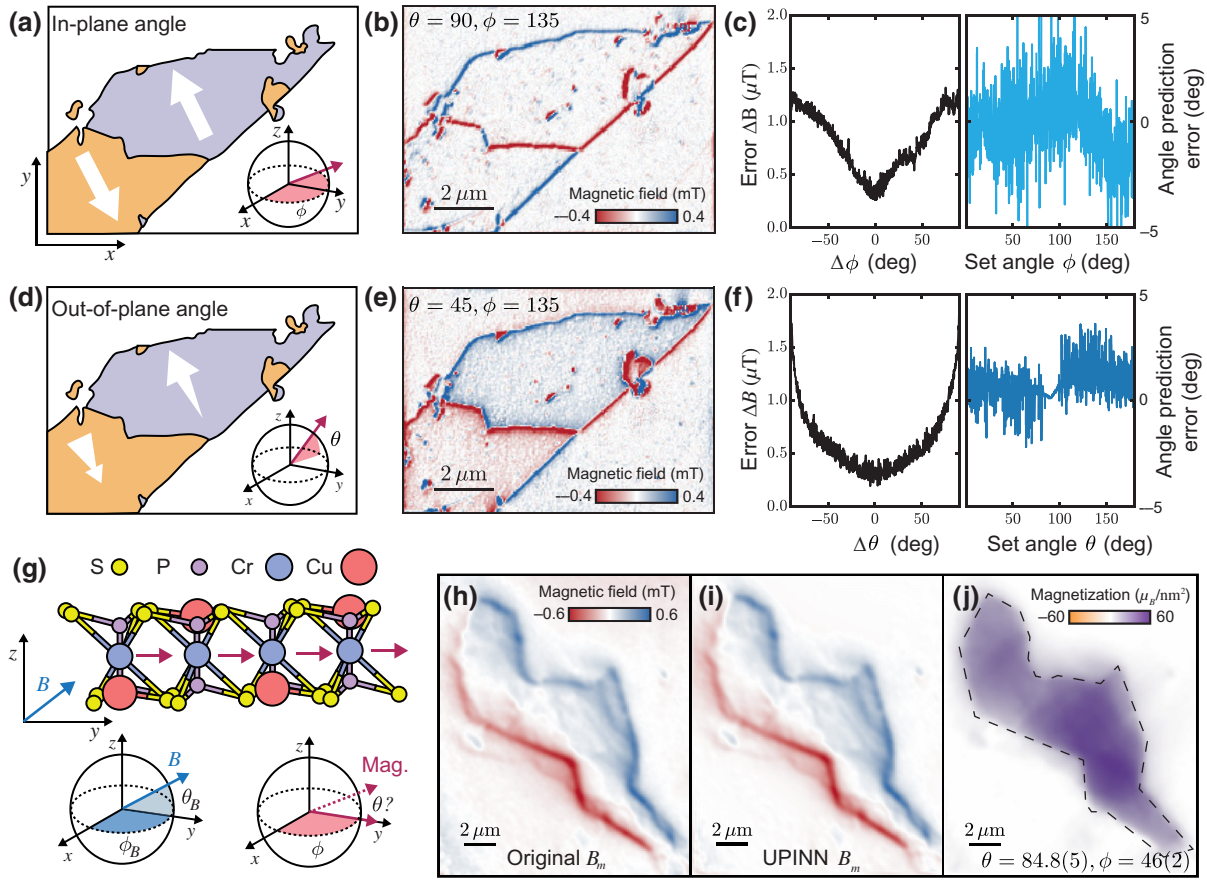

FIG. 3. Determination of the arbitrary magnetization direction of easy-axis magnets with a neural network. (a) An illustration of a magnetization image that is used to simulate magnetization with an arbitrary direction. (b) An example simulation of the magnetic field (B_z) image from a magnetization direction of $(\theta, \phi) = (90°, 135°)$. (c) A comparison of the estimated in-plane angle (ϕ) errors from the UPINN. Left panel: the error-function evolution for different assumptions of the angle, showing a minimum for the correct assumption. Right panel: the error in the angle prediction for different actual angles. (d) An illustration of magnetization canting out of plane to an arbitrary direction. (e) An example simulation of the magnetic field (B_z) image from a magnetization direction of $(\theta, \phi) = (45°, 135°)$. (f) The same as (c) but for out-of-plane (θ) angles. (g) An illustration of a monolayer of $CuCrP_2S_6$, which has a spontaneous magnetization in plane, with an applied magnetic field at $(\theta_B, \phi_B) = (54.7°, 90°)$ canting the spins out of the plane with an unknown angle. (h) The magnetic image of a 300-nm-thick $CuCrP_2S_6$ flake taken with a magnetometer and magnetic field direction of $(\theta_{m,B}, \phi_{m,B}) = (54.7°, 45°)$. (i),(j) The reconstructed magnetic field (i) and magnetization (j) of the $CuCrP_2S_6$ flake with the predicted magnetization angle of $(\theta, \phi) = (84.8(5)°, 46(2)°)$. The dashed line in (j) is a guide for the material edges, where the magnetization should be nonzero inside this region only.

To demonstrate this capability, we perform magnetization reconstructions on experimental data for an in-plane magnetized material [$CuCrP_2S_6$, Fig. 3(g)] that is measured with a nonplanar magnetic field [23], resulting in an unknown canting of the magnetization, and a magnetic field image shown in Fig. 3(h). The UPINN reconstruction [Figs. 3(i) and 3(j)] converges to a magnetization image that produces an average difference in the magnetic fields of $\Delta B \sim 2\ \mu$T, which corresponds to 3% of the rms magnetic signal strength. Here, the UPINN predicts an in-plane angle of $\phi = 46(2)°$ and an out-of-plane magnetization angle of $\theta = 84.8(5)°$. The former nearly matches the in-plane direction of the applied magnetic field $\phi_B = 45°$, which sets the magnetization direction within the easy plane of the sample, while the value determined for the out-of-plane magnetization angle θ is indicative of the out-of-plane component of the applied magnetic field ($\theta_B = 54.7°$) canting the spins out of being purely in-plane. The spin canting scales with the strength of the applied magnetic field (Appendix F), indicating that this is indeed a true spin canting rather than an artifact from the reconstruction process.

VI. COMPARISON OF MODELS

We perform comparisons of the magnetization reconstruction for several different networks: the CNN and the fully connected neural network (FCNN), both of which

are untrained, and a CNN that is trained with simulations. We do not train a FCNN because of the computational burden and its poor generalization abilities. In order to train the CNN, we create a data set with randomized magnetization shapes and domains, which is used with the forward solution to obtain magnetic field images with the associated ground truth. The value and direction of the reconstructed magnetization is influenced by the value of the magnetic field as well as five parameters: the magnetization angles θ and ϕ, as well as the angles of the sensor (here, the N-V orientation) θ_{N-V} and ϕ_{N-V}, and the sample-to-sensor stand-off d_{N-V}. All these parameters are randomly generated in order to create different combinations for the training. Additionally, we include random noise in the magnetic field image. The comparison of the reconstructions of the different networks is shown in Fig. 4.

The magnetization obtained with an untrained CNN is more resistant to noise compared to the untrained FCNN due to the filtering effects of convolutional layers. Despite being sharper, the magnetization obtained with a trained CNN shows artifacts due to the presence of unseen noise, namely the gradient in the input image, and produces a drastically different overall magnitude, which can be attributed to the reconstruction of inverse-direction magnetization outside of the material. However, the sharpness of the trained model does indicate that with a large and diverse training data set, a combination of training and the physically inferred loss function may produce a more accurate result.

In contrast, the traditional Fourier reconstruction method generally amplifies noise when reconstructing in-plane magnetization which requires spatial filtering to minimize [16] and, as such, is prone to artifacts in the background that can be larger than the reconstructed magnetization. In principle, this background can be removed with knowledge of the system [8] but this still requires high-quality data and is unusable for magnetization images that are either noisy or do not capture a significant known zero-magnetization region to normalize the image. UPINN offers a robust method to minimize the effect of these artifacts while still reconstructing a physically relevant representation of the magnetization. Additionally, UPINN does not add a significant computational burden, with image processing taking a few minutes for larger (512×512 pixels) images, which, for a postprocessing technique, is an acceptable time frame.

VII. CONCLUSIONS

We develop a method for solving ill-posed inverse problems with neural networks that use the well-posed forward problem. This method is comparable to traditional Fourier reconstruction methods when the ill-posed component is minimal but greatly outperforms them when more significant ill-posed terms are introduced into the transformation or when some parameters are *a priori* unknown. We demonstrate that our technique is capable of reconstructing the magnetization of in-plane magnetic material, which hitherto has been difficult to reconstruct. Additionally, we show that the UPINN method is capable of predicting the magnetization angle for an arbitrary direction, opening up the possibility of detailed studies of changes in magnetization direction. Furthermore, our method can also be extended to other transformation problems, such as the reconstruction of current density. Finally, we emphasize that while we demonstrate this technique using data obtained with nitrogen-vacancy centers in diamond, the technique is agnostic to sensor and only requires the magnetic field map (of any projection). As such, our method readily applies to other nanoscale-magnetometry techniques such as, e.g., scanning nano-SQUID (superconducting quantum interference device) magnetometry [24].

The PYTHON-based code for the machine-learning reconstruction can be found in Ref. [19].

ACKNOWLEDGMENTS

We acknowledge financial support from the National Centre of Competence in Research (NCCR) Quantum Science and Technology (QSIT), a competence center funded by the Swiss National Science Foundation (SNF), through SNF Project No. 188521, and by the European Research Council (ERC) consolidator grant project QS2DM. This work was supported by the Australian Research Council (ARC) through Grants No. CE170100012 and No. FT200100073.We would like to thank Sharidya Rahman, Boqing Liu, and Yuerui Lu for fabrication of the $CuCrP_2S_6$ sample. The machine-learning architecture was designed by A.E.E.D. and D.A.B. with assistance from P.M., A.S., E.G., and S.D.H. Author M.A.T. took the measurements of the CrI_3 sample for in-plane simulations and testing of the machine-learning code. A.J.H. and J.-P.T. took measurements of the $CuCrP_2S_6$ sample for testing the angle determination. A.E.E.D., D.A.B., and P.M. wrote the manuscript. All authors discussed the data and commented on the manuscript.

APPENDIX A: REGRESSION TASK USING A FEED-FORWARD FULLY CONNECTED NETWORK

A deep-learning architecture is made of neurons that are stacked in layers, where a neuron z is a linear combination of a weight w and a bias b, such that $z = x_i w_1 + b_1$ for a given input x_i.

A network that consists of a single hidden layer l with q neurons holds the following relationship:

$$y_i \approx g_\alpha(\mathbf{x}_i) = f\,(x_i W_1 + b_1) W_2 + b_2 \qquad \text{(A1)}$$

where W_1 and W_2 are weight matrices of size $(d \times q)$ and $(q \times k)$ and the bias vectors b_1 and b_2 are of size q and k, respectively, in which d is the size of the input data set, q is the number of neurons in the layer, and k is the number of neurons in the output layer, denoted by the subscript 2. The weights and bias vectors are the set of parameters α that the model has to update while training. The activation function f (.) is a nonlinear transformation carried out on each neuron at each layer. A detailed description of this type of neural network can be found in Ref. [25].

In general, the approximation of complex nonlinear functions can be dramatically improved by increasing the number of neurons z in each layer or by adding more layers l [20]. In multiple-layer architectures, the output of each activation function becomes the next-layer input, which is then transformed by a new weight matrix and bias vector. At the end of such a multilayer network, a specific activation function f (.)—called the output function—is used to define the output format in the desired form. In the method presented in this work, the output is an image that is the same size as the input image but this can also be used to restrict the output to possibilities for classification problems.

APPENDIX B: TRAINING A NEURAL NETWORK

In order to train a network, one needs a data set made of pairs of inputs and outputs (X, Y). At the end of a forward pass, the network estimates an output Y' from the given input, X. The difference between the ground truth Y^T and the estimation Y' is computed to obtain the loss function. A common loss function is the mean absolute error, given by

$$C(Y^T, Y') = \frac{1}{n}\sum_{i=1}^{n} \left|g(x)_i - y_i^t\right|, \tag{B1}$$

where n is the number of output parameters, $g(x)_i$ is the network estimation for the ith parameter, and y_i^t is the corresponding ground truth. In order to update the network parameters, a back-propagation algorithm evaluates the gradients of the loss function $\nabla_\alpha C$ with respect to the weights and biases. This procedure begins by calculating the gradients from the last layer and continues layer by layer using the chain rule, which qualifies how much the value of each parameter affects the final loss of the network.

Once the gradients have been evaluated, the optimizer defines the parameter update. Commonly, the simplest version of this optimizer can be used, the standard stochastic gradient descent. At each iteration i, it updates individual values of parameters α based on their respective gradients and a learning rate η:

$$\alpha^{i+1} = \alpha^i - \eta^i \nabla_\alpha C^i. \tag{B2}$$

TABLE I. The architecture of the convolutional neural network. conv, convolutional.

	Layer type	Filters	Kernel size	Stride	Neurons
	conv	8	5	1	
	conv (ROI)	8	5	1	
	conv	8	5	2	
	conv	16	5	2	
	conv	32	5	2	
	conv	64	5	2	
Image	conv	128	5	2	
	conv	64	5	2	
	conv	32	5	2	
	conv	16	5	2	
	conv	8	5	2	
	conv	1	5	2	
	conv	1	3	1	
	Dense				128
Parameters	Dense				64
	Dense				1

The whole process is set as an optimization problem $\arg\min_\alpha C$, where forward feeding is repeated until some stopping criteria are met.

APPENDIX C: TRAINING AND LEARNING OF A PURELY PHYSICALLY INFERRED NEURAL NETWORK

Usually, regularization acts on the weights α in order to overcome overfitting, where a physics regularization deals with the implementation of knowledge into the model. This added knowledge constrains the learning to a desired physical solution by adding a penalty term $\Phi(Y^o)$ to the loss function C. In our case, this term is the forward solution for reconstructing the magnetic field from the magnetization. Thus, at the end of each iteration, the magnetic field is computed from the current estimated magnetization. By doing so, it can be compared with the original magnetic

TABLE II. The architecture of the fully connected neural network.

	Layer type	Filters	Kernel size	Stride	Neurons
	Dense				256
	Dense				128
	Dense				64
	Dense				32
Image	Dense				16
	Dense				32
	Dense				64
	Dense				128
	Dense				256
	Dense				128
Parameters	Dense				64
	Dense				1

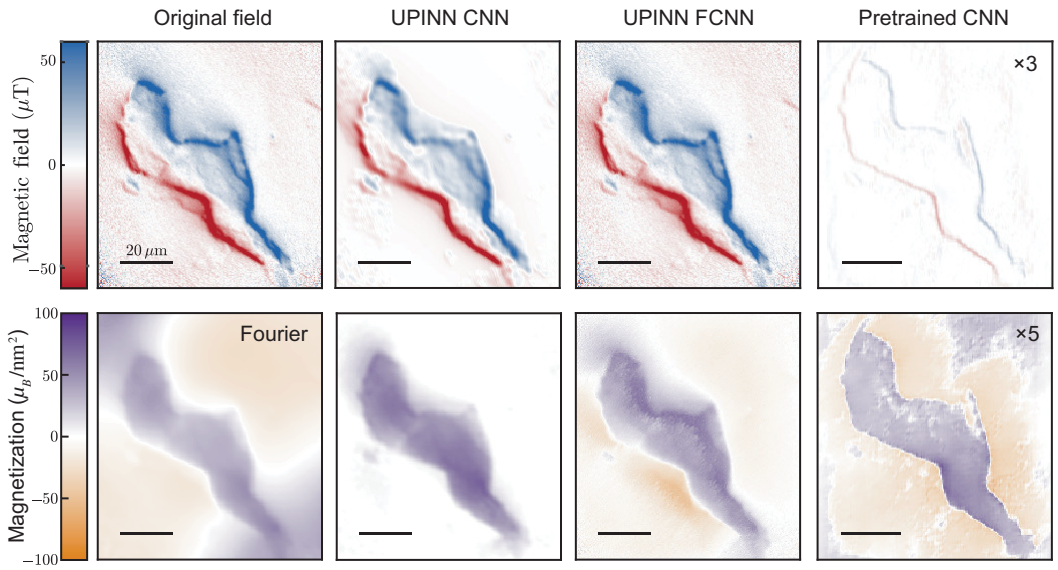

FIG. 4. A comparison of the different reconstruction methods on the same magnetic field image. The top row compares the original magnetic field to the reconstructed magnetic field from the neural-network approaches: UPINN CNN, UPINN FCNN, and trained CNN. The bottom row compares the magnetization reconstruction from the traditional Fourier method and the same neural networks as the top row. All of the reconstructions use a set magnetization angle of $(\theta, \phi) = (90°, 45°)$ to compare just the spatial-reconstruction capabilities. The Fourier method includes the use of a spatial filter to minimize the noise [16], which may mask magnetization information in more complex images.

field X^0 at each iteration and be included in the loss function. The complete loss function with physics constraint C_r takes the form

$$C_r(Y^t, Y^o) = C(Y^t, Y') + \lambda[C_p(X^0, \Phi(Y'))], \qquad \text{(C1)}$$

where λ is a hyperparameter controlling the influence of the physics term and C_p refers to the physical-inference cost function. By including this addition term, one can ensure that even on a subset of all possible data, the model will learn the proper function.

In the situation where no ground truths are available, it is possible to get rid of the first part of the loss function all together, keeping only the physics term. The loss function then becomes

$$C_r(Y^t, Y') = C_p(X^0, \Phi(g(X^0))). \qquad \text{(C2)}$$

The ground truth Y^t disappears from the loss function and this one only depends on the original input X^0 and the learned function $g(.)$ that gives the network estimation. This modification allows a model to train without the required ground-truth value. Moreover, it can directly learn on a single new input without previous training. So instead of traditionally training on a data set and using the learned weights to make new predictions, a network updates its parameters α directly on the measured quantity X^0.

The output of the network is no longer a statistical inference but a solution depending on the well-posed forward transformation. Thus, the knowledge of the model is not limited to a previous data set but is directly taken from the input. Seeing that the loss function depends on the well-posed forward transformation, it can be ensured that the proposed solution is physically relevant.

APPENDIX D: MODEL AND NETWORK ARCHITECTURE

The possibility of training thenetwork directly on a single image makes it possible to use different architecture without facing a computational problem. In this main text, we use a CNN following a U-net architecture, where the cost function is the mean average loss between the reconstructed magnetic field and the original one. Seeing that the convolutional layer requires a fixed-size input, a border is added to the input image to match the correct format. In order to make sure that this border is not taken into account in the reconstruction, a region-of-interest (ROI) layer is inserted to focus the learning of the network on the desired region [26]. Out of the six parameters influencing the reconstruction of the magnetization, four can be deduced by the model directly. It consists of the 2 N-V angles and the 2 magnetization angles. For this purpose, a subnetwork—fully connected—is forked from the main network for each one of those parameters. The architecture for the convolutional neural network is displayed in Table I. We also implement a FCNN) to compare with, the architecture of which is displayed in Table II.

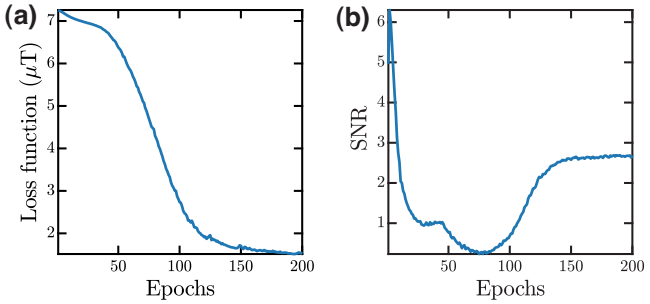

FIG. 5. An example of the evolution of different error sources as a function of the number of training epochs. (a) The evolution of the loss function, $\Delta B_m = \text{mean}(|B_m - B_{\text{NN}}|)$. (b) The evolution of the SNR $= ((\max(M) - \min(M))/\text{std}(M))$ of the reconstructed magnetization image.

APPENDIX E: NOISE REGULATION

One issue with implementing any reconstruction method is that noise in the original data can be amplified during the reconstruction process. Thus data with excessive noise are often ruled as ineligible for reconstruction, requiring either additional measurement time or a new measurement to be taken. This reconstruction is no different; in fact, the ill-posed problem is known to amplify noise [16]. With a trained neural network, one can generalize a model to get rid of the noise in the reconstruction. However, one still faces the risk of unseen perturbations creating artifacts in the reconstruction (see Fig. 4). The untrained neural network reconstructs the image without amplification of the noise; instead, the noise is reflected in the reconstructed image directly. There is a trade-off here between the generalization power offered by a trained network and the accuracy and robustness offered by our direct network.

As the untrained neural network learns on the input directly, the quality of the reconstruction is directly linked to the quality of the input. Thus, we can easily improve the reconstruction quality by minimizing the noise beforehand. However, removing noise always carries the risk of removing valuable information, which is particularly damaging when dealing with precise measurements. Isolating the denoising from the reconstruction is valuable, as it allows for the possibility of using noisy images without the need for additional training and it facilitates double checking of the denoising process itself for loss of information. In contrast, a trained neural-network model would have to learn the denoising process and can thus generalize incorrectly for a given data set. In this work, we do not include any direct denoising of the data but, in principle, it may help to improve results if implemented correctly.

In the comparison of models, we see that the CNN has the power to be more noise resistant compared to the FCNN due to the convolutional layers. However, the longer the untrained model learns, the more noise is taken into account into the reconstruction. In order to avoid this situation, we include an early stopping criteria, a SNR metric, defined as $[\max(M) - \min(M)]/\text{std}(M)$. Once the SNR has reached a minimum, the learning stops even though the loss function can still be reduced. For a typical behavior of the loss function and the SNR for an increasing number of epochs, see Fig. 5.

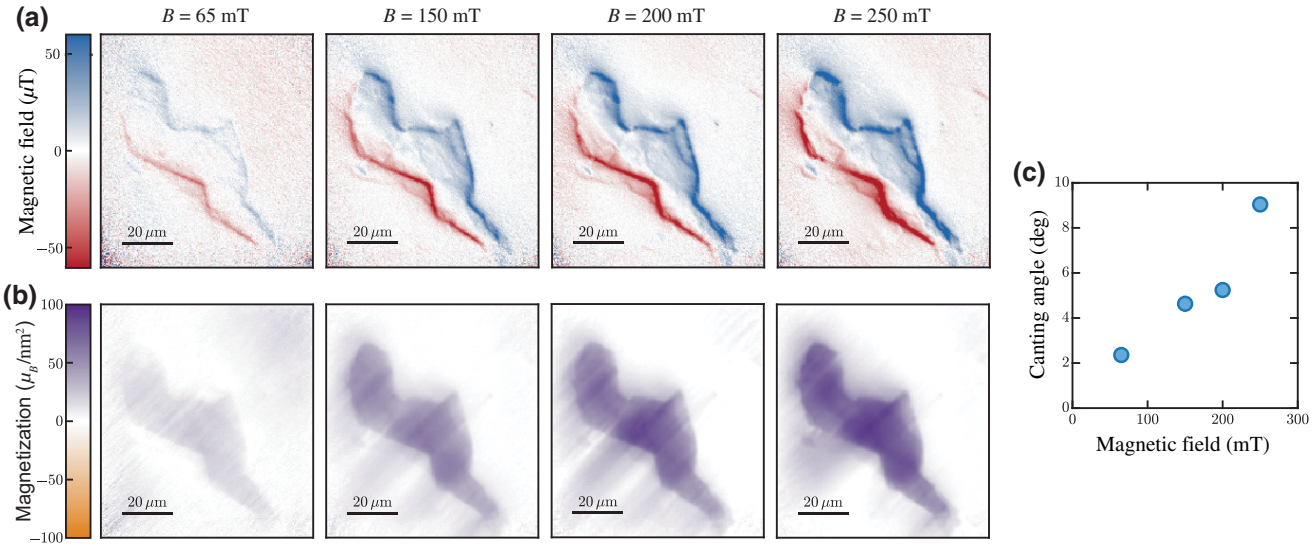

FIG. 6. The determination of out-of-plane spin canting with neural-network training. (a) Magnetic images of the same flake in the main text taken at different applied magnetic fields $B = (65, 150, 200, 250)$ mT, where the field is applied along an in-plane angle of $B_\phi = 45^\circ$ and an angle from the z axis of $B_\theta = 54.7^\circ$ (b) Corresponding magnetization reconstructions from the above magnetic fields. (c) The predicted spin-canting angle $(90^\circ - \theta)$ from purely in plane versus the applied magnetic field.

APPENDIX F: DETERMINATION OF OUT-OF-PLANE SPIN CANTING

While the determination of the spin angle in simulations is relatively robust to noise, it is important to qualify how reliable the technique is on real data. Unfortunately, with data sets like these, we do not have the ground truth. However, we can measure trends in the spin orientation to determine if these make physical sense. Here, we examine the reconstructed out-of-plane angle of $CuCrP_2S_6$ as a function of the applied magnetic field.

$CuCrP_2S_6$ is an in-plane A-type antiferromagnet, which in bulk has a small difference in susceptibility between the in-plane and out-of-plane directions [27], while density-functional theory (DFT) calculations show that a weak anisotropy exists in the monolayer limit [28]. A naive single-domain shape-anisotropy calculation (agreeing with monolayer DFT calculations) predicts negligible canting over the range of fields probed ($< 2^\circ$ at 250 mT); however, it is unlikely to capture the complex interplay between the various anisotropies and exchange terms in thicker flakes. Our NN analysis shows a monotonic increase in the canting angle (see Fig. 6) as expected but the increase is much more rapid than the naive prediction. As this is an A-type antiferromagnet and is relatively thick (approximately 300 layers), it is possible that the monolayer DFT calculations are no longer valid in this regime.

Precise measurement of the magnetization angle is difficult to obtain by other means; thus this demonstrates the power of this NN approach to determine small perturbations of spin alignments.

[1] F. Casola, T. van der Sar, and A. Yacoby, Probing condensed matter physics with magnetometry based on nitrogen-vacancy centres in diamond, Nat. Rev. Mater. **3**, 17088 (2018).

[2] R. Wiesendanger, Nanoscale magnetic skyrmions in metallic films and multilayers: A new twist for spintronics, Nat. Rev. Mater. **1**, 16044 (2016).

[3] K. S. Burch, D. Mandrus, and J.-G. Park, Magnetism in two-dimensional van der Waals materials, Nature **563**, 47 (2018).

[4] C. Donnelly, M. Guizar-Sicairos, V. Scagnoli, S. Gliga, M. Holler, J. Raabe, and L. J. Heyderman, Three-dimensional magnetization structures revealed with x-ray vector nanotomography, Nature **547**, 328 (2017).

[5] J.-P. Tetienne, N. Dontschuk, D. A. Broadway, A. Stacey, D. A. Simpson, and L. C. L. Hollenberg, Quantum imaging of current flow in graphene, Sci. Adv. **3**, e1602429 (2017).

[6] M. J. H. Ku, T. X. Zhou, Q. Li, Y. J. Shin, J. K. Shi, C. Burch, L. E. Anderson, A. T. Pierce, Y. Xie, A. Hamo, U. Vool, H. Zhang, F. Casola, T. Taniguchi, K. Watanabe, M. M. Fogler, P. Kim, A. Yacoby, and R. L. Walsworth, Imaging viscous flow of the Dirac fluid in graphene, Nature **583**, 537 (2020).

[7] U. Vool, A. Hamo, G. Varnavides, Y. Wang, T. X. Zhou, N. Kumar, Y. Dovzhenko, Z. Qiu, C. A. C. Garcia, A. T. Pierce, J. Gooth, P. Anikeeva, C. Felser, P. Narang, and A. Yacoby, Imaging phonon-mediated hydrodynamic flow in WTe_2, Nat. Phys. **17**, 1216 (2021).

[8] D. A. Broadway, S. C. Scholten, C. Tan, N. Dontschuk, S. E. Lillie, B. C. Johnson, G. Zheng, Z. Wang, A. R. Oganov, S. Tian, C. Li, H. Lei, L. Wang, L. C. L. Hollenberg, and J.-P. Tetienne, Imaging domain reversal in an ultrathin van der Waals ferromagnet, Adv. Mater. **2003314**, 2003314 (2020).

[9] L. Thiel, Z. Wang, M. A. Tschudin, D. Rohner, I. Gutiérrez-Lezama, N. Ubrig, M. Gibertini, E. Giannini, A. F. Morpurgo, and P. Maletinsky, Probing magnetism in 2D materials at the nanoscale with single-spin microscopy, Science **364**, 973 (2019).

[10] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2019).

[11] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, Using deep neural networks for inverse problems in imaging: Beyond analytical methods, IEEE Signal Process. Mag. **35**, 20 (2018).

[12] M. Prato and L. Zanni, Inverse problems in machine learning: An application to brain activity interpretation, J. Phys.: Conf. Ser.Open Access **135**, 12085 (2008)., mL paper

[13] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, Deep learning techniques for inverse problems in imaging (2020), mL paper.

[14] Y. Khoo and L. Ying, SwitchNet: A neural network model for forward and inverse scattering problems, SIAM J. Sci. Comput. **41**, A3182 (2019).

[15] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, Solving inverse problems using data-driven models, Acta Num. **28**, 1 (2019).

[16] D. Broadway, S. Lillie, S. Scholten, D. Rohner, N. Dontschuk, P. Maletinsky, J.-P. Tetienne, and L. Hollenberg, Improved Current Density and Magnetization Reconstruction through Vector Magnetic Field Measurements, Phys. Rev. Appl. **14**, 024076 (2020).

[17] A. Y. Meltzer, E. Levin, and E. Zeldov, Direct Reconstruction of Two-Dimensional Currents in Thin Films from Magnetic-Field Measurements, Phys. Rev. Appl. **8**, 064030 (2017).

[18] C. B. Clement, J. P. Sethna, and K. C. Nowack, Reconstruction of current densities from magnetic images by Bayesian inference (2019), ArXiv:1910.12929.

[19] A. Dubois, D. Broadway, A. Stark, M. A. Tschudin, A. J. Healey, S. D. Huber, J.-P. Tetienne, E. Greplova, and P. Maletinsky, Untrained physically informed neural network for image reconstruction of magnetic field sources (2022), https://doi.org/10.5281/zenodo.7197632.

[20] A. J. Healey, S. Rahman, S. C. Scholten, I. O. Robertson, G. J. Abrahams, N. Dontschuk, B. Liu, L. C. L. Hollenberg, Y. Lu, and J.-P. Tetienne, Varied magnetic phases in a van der Waals easy-plane antiferromagnet revealed by nitrogen-vacancy center microscopy, ACS Nano **16**, 12580 (2022).

[21] O. Ronneberger, P. Fischer, and T. Brox, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Springer International Publishing, Cham, 2015), p. 234.

[22] M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, DeepCGH: 3D computer-generated holography using deep learning, Opt. Express **28**, 26636 (2020).

[23] A. J. Healey, S. Rahman, S. C. Scholten, I. O. Robertson, G. J. Abrahams, N. Dontschuk, B. Liu, L. C. L. Hollenberg, Y. Lu, and J. P. Tetienne, Varied magnetic phases in a van der Waals easy-plane antiferromagnet revealed by nitrogen-vacancy center microscopy (2022), ArXiv ArXiv:2204.09277.

[24] D. Vasyukov, Y. Anahory, L. Embon, D. Halbertal, J. Cuppens, L. Neeman, A. Finkler, Y. Segev, Y. Myasoedov, M. L. Rappaport, M. E. Huber, and E. Zeldov, A scanning superconducting quantum interference device with single electron spin sensitivity, Nat. Nanotechnol. **8**, 639 (2013).

[25] M. A. Nabian and H. Meidani, Physics-driven regularization of deep neural networks for enhanced engineering design and analysis, J. Comput. Inf. Sci. Eng. **20**, 011006 (2020).

[26] S. Eppel, Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels (2017), ArXiv ArXiv:1708.08711.

[27] P. Colombet, A. Leblanc, M. Danot, and J. Rouxel, Structural aspects and magnetic properties of the lamellar compound $Cu_{0.50}Cr_{0.50}PS_3$, J. Solid. State. Chem. **41**, 174 (1982).

[28] J. Qi, H. Wang, X. Chen, and X. Qian, Two-dimensional multiferroic semiconductors with coexisting ferroelectricity and ferromagnetism, Appl. Phys. Lett. **113**, 043102 (2018).