

Learned interferometric imaging for the SPIDER instrument

Matthijs Mars¹,^{*} Marta M. Betcke² and Jason D. McEwen^{1,3}

¹Mullard Space Science Laboratory (MSSL), University College London (UCL), Dorking RH5 6NT, UK

²Department of Computer Science, University College London (UCL), London WC1E 6BT, UK

³Alan Turing Institute, London NW1 2DB, UK

Accepted 2023 November 1. Received 2023 September 29; in original form 2023 February 20

ABSTRACT

The Segmented Planar Imaging Detector for Electro-Optical Reconnaissance (SPIDER) is an optical interferometric imaging device that aims to offer an alternative to the large space telescope designs of today with reduced size, weight, and power consumption. This is achieved through interferometric imaging. State-of-the-art methods for reconstructing images from interferometric measurements adopt proximal optimization techniques, which are computationally expensive and require handcrafted priors. In this work, we present two data-driven approaches for reconstructing images from measurements made by the SPIDER instrument. These approaches use deep learning to learn prior information from training data, increasing the reconstruction quality, and significantly reducing the computation time required to recover images by orders of magnitude. Reconstruction time is reduced to ~ 10 ms, opening up the possibility of real-time imaging with SPIDER for the first time. Furthermore, we show that these methods can also be applied in domains where training data are scarce, such as astronomical imaging, by leveraging transfer learning from domains where plenty of training data are available.

Key words: machine learning – image processing – interferometric imaging.

1. INTRODUCTION

The Segmented Planar Imaging Detector for Electro-Optical Reconnaissance (SPIDER, Kendrick et al. 2013; Duncan et al. 2015) instrument is an alternative electro-optical imaging device to current space telescopes like the *Hubble Space Telescope* or the *JWST*. While traditional electro-optical telescopes require large optics, housings, and thermal controls for the optics to attain precise measurements, SPIDER aims to decrease the volume, mass, and cost of electro-optical imagers by replacing the traditional mirrors with arrays of lenslets to gather interferometric measurements. The light captured by these lenslets is processed using photonic integrated circuit (PIC) chips into interferometric measurements. Typical interferometers still require large components to make measurements; yet by processing the light from the lenslets on an integrated chip, the size of the instrument can be reduced significantly. The SPIDER instrument is designed to be a cheaper and lighter alternative for instruments that can be used for both Earth observation and astronomical research.

The concept design for the SPIDER instrument proposed by Kendrick et al. (2013) and Duncan et al. (2015) uses 37 PICs mounted at different angles to form a planar disc. Interferometric measurements are acquired on each of the PICs, and by combining the measurements taken at each of the different angles, the telescope creates a large synthetic aperture with the diameter of the aperture equal to the largest spacing of lenslets on the PIC. A diagram of their design can be found in Fig. 1.

Beyond the initial concept, more efficient designs that make better use of the baselines (Liu, Wen & Song 2018; Liu et al.

2019) and improved lenslet arrangements to increase reconstructions quality (Lv et al. 2020; Hu et al. 2021) have been investigated. Furthermore, the PIC reconstruction capabilities were found to match the theoretical predictions of the concept design well (Chu et al. 2017). Tests of the PICs show that it is possible to measure both the amplitude and phase of the incoming light as predicted (Su et al. 2017) and that these measurements can be used for image reconstruction (Badham et al. 2017; Su et al. 2018).

Since SPIDER measures both the phase and the amplitude of the interferometric signals, the measurement process is analogous to that of radio interferometers. Interferometric imaging techniques developed for radio interferometry can thus be adapted and repurposed to recover images from the raw data acquired by the SPIDER instrument (Pratley & Mcewen 2021). Radio interferometry and aperture synthesis have played an important role in pushing the boundaries of astronomical research in the radio frequency regime, where acquiring high-resolution images is otherwise difficult because of the relatively large wavelengths considered. By combining measurements from pairs of radio telescopes, spatial frequency information can be retrieved from the observed sky. Each pair of telescopes forms a baseline and measures a so-called visibility corresponding to a Fourier coefficient of the sky brightness distribution. By sampling the Fourier plane with these measurements, an observation with a uniform aperture is approximated using aperture synthesis. Since the telescope acquires only a finite number of measurements and since the Fourier plane is not sampled uniformly, the problem is ill-posed and an accurate image cannot be reconstructed simply by inverting the Fourier transform. Instead, when inverting the measuring process directly as described a so-called dirty image is obtained: a reconstruction of the sky brightness convolved with the

* E-mail: matthijs.mars97@gmail.com

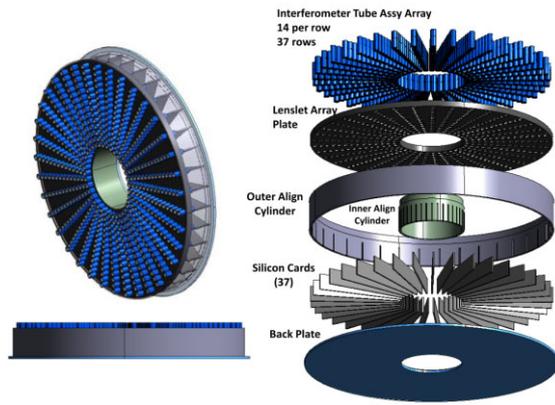


Figure 1. A diagram of a SPIDER instrument design proposed with 37 PICs with lenslets attached to them. Image credit: Kendrick et al. (2013).

point spread function (PSF) of the telescope configuration. In order to recover an accurate image of the sky brightness, techniques to regularized inverse problems are typically considered.

One method for solving this inverse problem is variational regularization. Variational regularization techniques combine a data fidelity component relative to a forward model with a regularizer in order to constrain and stabilize the problem. Regularizers for these algorithms encode prior information about the images. An example of modern model-based iterative methods applied to radio interferometry can be found in Pratley et al. (2018), which adopt convex optimization techniques to solve the variational regularization optimization problem. An application of these techniques to SPIDER imaging can be found in Pratley & McEwen (2021). While these techniques can be distributed to reduce the reconstruction time (Pratley et al. 2019), they are computationally expensive as they evaluate the measurement operator that models the instrument at each iteration of the optimization process. Besides the computational cost of these algorithms, they also adopt handcrafted priors (e.g. ℓ_1 sparsity in a wavelet representation) that, while general, are not tailored to the images of interest.

An alternative to model-based algorithms, such as variational regularization, are data-driven algorithms that learn the prior information from training data. Since the prior is implicitly specified by the training data, data-driven methods are typically able to achieve higher reconstruction quality than by using handcrafted priors, provided the training distribution matches the target distribution well. Learned methods that are completely independent of the measurement operator and attempt to learn a direct mapping from measurements to target image, however, are generally not effective. Thus, learned methods typically combine some model-based information, like the measurement operator, with prior information learned from training data. Learned data-driven approaches broadly fall into three categories, differentiated by the degree to which model-based information such as the measurement operator is encoded or leveraged. *Learned regularization methods* (e.g. Lunz, Öktem & Schönlieb 2018; Li et al. 2020) are iterative in nature and make full use of the measurement operator in each iteration; consequently, they achieve excellent reconstruction quality by exploiting full knowledge of the measurement operator, along with learned prior information, but are generally highly computationally demanding. *Learned sequential methods* (e.g. Jin et al. 2017), on the other hand, simply pre- and/or post-process data with learned models in observation and/or image space; consequently, while they nevertheless provide good quality reconstructions they are limited by the fact that they only evaluate

the measurement operator very few times (sometimes just once). This however does make these methods substantially more efficient computationally. *Learned iterative methods* (e.g. Adler & Öktem 2017), also called unrolled methods, provide a balance by designing a learned model that attempts to unroll a small number of iterations of iterative approaches, encoding the measurement operator into the model; consequently, they typically achieve superior reconstruction quality to learned sequential methods exploiting greater knowledge of the measurement operator and are more computationally efficient than learned regularization methods.

Many such methods were originally proposed in the medical imaging context (e.g. computed tomography [CT], magnetic resonance imaging [MRI], photoacoustic tomography [PAT]) and are based on deep learning approaches (e.g. Adler & Öktem 2017, 2018; Arridge et al. 2019). In the field of radio astronomy, learned image reconstruction techniques were first considered by McEwen & Allam, Jr (Allam 2016), where superresolution convolutional neural networks (CNNs, Dong et al. 2016) were applied to post-process dirty radio interferometric images. These networks were considered for cases where the exact telescope PSF is known, as well as for the PSF-unaware case, highlighting the potential of learning a generalized network that works with unseen telescope configurations. Although, performance was relatively poor for this rudimentary approach. More recent applications of learned post-processing methods in radio astronomy have been considered using learned denoisers (Terris et al. 2019), convolutional autoencoders (Gheller & Vazza 2021), and superresolution networks (Connor et al. 2022). The use of plug-and-play (PnP, Venkatakrishnan, Bouman & Wohlberg 2013) denoisers within iterative reconstruction methods has also been considered for radio interferometric imaging (Terris et al. 2022), although such approaches still require many evaluations of the measurement operator during imaging.

While we draw inspiration from these prior works on learned imaging for interferometry, in this paper we develop new learned imaging techniques, specifically targeting the SPIDER instrument. Our primary goal is reducing computational cost, while of course still ensuring high-quality reconstructions. Reducing the computational cost of recovering images from raw SPIDER measurements would open up the possibility of real-time imaging with SPIDER, which would afford numerous new applications. Consequently, we develop a learned sequential method that requires only a single evaluation of the measurement operator. While reconstruction quality for this method is similar to current state-of-the-art variational regularization techniques, computational time is orders of magnitude faster. We also develop a learned iterative method, trading off a small increase in computational time compared with the learned sequential method, but that nevertheless remains orders of magnitude faster than traditional approaches, while achieving a further improvement in reconstruction quality.

The remainder of the paper is structured as follows. Section 2 introduces the measurement process of interferometric imagers and the particular imaging configuration of the SPIDER instrument, as well as discussing the inverse imaging problem. Variational and learned approaches for the inverse imaging problems are reviewed in Section 3. In Section 4, we present two approaches to model the SPIDER instrument: alongside the standard non-uniform Fourier transform approach considered previously (Pratley & McEwen 2021), we also introduce a new modelling approach for SPIDER based on the Radon transform. In Section 5, we propose our learned methods for reconstruction from interferometric measurements. Section 6 evaluates the performance of the reconstruction methods when applied to natural images, their robustness to additional noise, and

Table 1. The parameters of the SPIDER concept design proposed in Kendrick et al. (2013).

Parameter	Value
Spectral range	500–900 nm
Lenslet diameter	8.75 mm
Longest baseline	0.5 m
Number of lenslets per PIC card	24
Number of PIC cards	37
Number of spectral bins	10

generalization potential of these methods to smaller data sets of galaxy or satellite images. Finally, concluding remarks are made in Section 7.

2. SPIDER INSTRUMENT

The SPIDER instrument measures incoming light through pairs of separated lenses and combines it in waveguides on a PIC chip to form interferometric baselines. All the operations to make the interferometric measurements are performed on the chip, resulting in a small form-factor for the system. The concept design introduced in Kendrick et al. (2013) uses a linear array of lenslets attached to one PIC chip to measure several baselines using a one-dimensional (1D) interferometer. Several of these so-called spokes are then mounted radially resulting in a 2D sampling pattern. While in typical interferometers the information of all receivers can be combined into interferometric measurements, resulting in a total of $N(N-1)/2$ measurements for N lenslets, in the SPIDER design lenslets can only be combined within one PIC chip, resulting $N/2$ baselines per PIC module. To increase the number of baselines gathered from one PIC card, different wavelengths of light are measured. Because the spatial frequency measured depends on the separation of the lenslets and the wavelength of the light, the number of baselines is multiplied by the number of wavelength bins observed. The technique of incorporating spectral information from measurements at different wavelengths in a single reconstruction is called multifrequency synthesis and is common in radio interferometry (Sault & Conway 1999).

By mounting multiple PIC chips as radial spokes on a disc, a 2D interferometer is created. In doing so a radial sampling profile is created for the uv -plane. In the concept of Kendrick et al. (2013), 37 spokes of PICs with 24 lenslets each are used. The measured light frequency spectrum is from 500 to 900 nm with 10 spectral bins. This results in 120 baselines per spoke and a total of 4440 measured Fourier components. The parameters for the configuration used by Kendrick et al. (2013) can be found in Table 1 and the lenslet layout and the uv -sampling are shown in Fig. 2.

The interferometric measurements at the baseline (spatial) frequency $\xi = (u, v)$ represent samples of the 2D Fourier transform of the image $f(\chi)$, with spatial coordinate χ , as given by the van Cittert–Zernike theorem (Zernike 1938):

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\chi) e^{-i2\pi \chi \cdot \xi} d\chi, \quad (1)$$

which is a continuous unitary Fourier transform of the signal. The baselines are determined by the spatial distance between the lenslets on the PIC with respect to the observed wavelength of the light. To recover an image from the Fourier measurements one needs the inverse of equation (1),

$$f(\chi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi \chi \cdot \xi} d\xi, \quad (2)$$

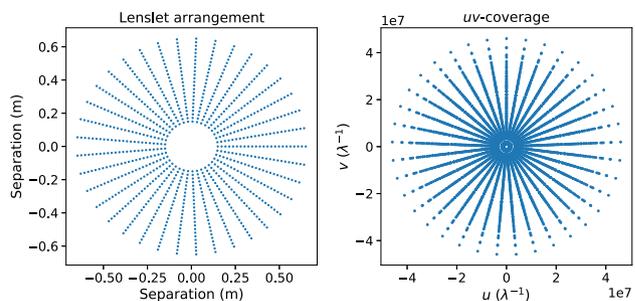


Figure 2. Left panel: The physical locations of the lenslets of the 2D interferometer as detailed in the design proposed in Kendrick et al. (2013). The lenslets on each of the radial spokes are mounted to their respective PIC. Right panel: The measured baselines of the interferometric measurements between pairs of lenslets on the PICs of the SPIDER instrument. The amount of baselines is increased by measuring at different spectral frequencies. Note that the measurements all lie in the same direction as the directions of the spokes, since measurements are only made using pairs of lenslets on 1 PIC.

which is only possible in theory when one has complete knowledge of the continuous Fourier representation $\hat{f}(\xi)$. In practice this is not the case.

The approaches for modelling the SPIDER instrument introduced in Section 4 require an (upsampled) discrete Fourier transform (DFT). The unitary DFT for an $N_1 \times N_2$ image is an operator mapping from $\mathbb{R}^{N_1 \times N_2} \rightarrow \mathbb{C}^{M_1 \times M_2}$, we henceforth use the same \hat{f} notation for the continuous and DFT as the meaning will be clear from the context. The unitary DFT is defined by

$$\hat{f}(\xi_{kl}) = \frac{1}{\sqrt{N_1 N_2}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(\chi_{ij}) e^{-i \chi_{ij} \cdot \xi_{kl}}, \quad (3)$$

and the unitary inverse DFT by

$$f(\chi_{ij}) = \frac{1}{\sqrt{M_1 M_2}} \sum_{k=1}^{M_1} \sum_{l=1}^{M_2} \hat{f}(\xi_{kl}) e^{i \chi_{ij} \cdot \xi_{kl}}. \quad (4)$$

The interferometric measurement process (equation 1) can be written in the compact, discretized form

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (5)$$

where the linear measurement operator $\Phi: X \rightarrow Y$, maps the unknown image $\mathbf{x} \in X \subset \mathbb{R}^N$ to the set of noisy measurements $\mathbf{y} \in Y \subset \mathbb{C}^M$ (here Fourier components of the weighted sky-brightness), with $\mathbf{n} \in Y \subset \mathbb{C}^M$ some type of measurement noise. The interferometric measurement operator corresponds to a non-uniformly sampled Fourier transform.

The limited number of lenslets of the telescope results in a limited sampling of the uv -plane. Since the Fourier domain is sampled incompletely, the measurement operator is ill-posed, and it cannot simply be inverted to find a solution to the inverse problem. Because of the ill-posedness of the operator, prior information on the solution is needed to regularize the inversion. Therefore, approaches such as sparse regularization are needed to stably recover a solution for the reconstruction. Underlying the sparse regularization is the idea that natural signals (e.g. astronomical images) are sparse or compressible in a suitable basis or frame (e.g. wavelet bases).

3. INVERSE IMAGING APPROACHES

We briefly recall the state-of-the-art variational techniques that are applied to solve the interferometric imaging problem and how these

methods can be enhanced or replaced by using learned approaches that make use of deep learning.

3.1 Variational regularization

To recover a solution to the inverse problem stated in equation (5), we find a solution for \mathbf{x} that trades off matching the data and the prior information encoded in the regularizer. Variational regularization approaches do this by posing an appropriate optimization problem, which is then often solved by proximal optimization algorithms.

3.1.1 Optimization problems

Trading off data fidelity and prior information can be achieved by obtaining a solution to the following minimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in X} \mathcal{L}(\Phi \mathbf{x}, \mathbf{y}) + \lambda \mathcal{S}(\mathbf{x}), \quad (6)$$

where λ is the regularization parameter, and $\mathcal{L}(\Phi \mathbf{x}, \mathbf{y})$ and $\mathcal{S}(\mathbf{x})$ are a data fidelity and regularization functional, respectively.

A common optimization strategy is to use a regularizer that promotes sparsity in a particular frame. The signal $\mathbf{x} \in \mathbb{R}^N$ can be represented in a dictionary $\Psi \in \mathbb{R}^{N \times D}$ by $\mathbf{x} = \Psi \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^D$. Natural signals typically exhibit stronger sparsity if the dictionary is redundant, i.e. $D > N$ (Gribonval & Nielsen 2003; Bobin et al. 2007; Starck, Murtagh & Fadili 2010). It is therefore often beneficial to use dictionaries which are a concatenation of different basis of frames, e.g. Dirac and Daubechies wavelets (Carrillo, McEwen & Wiaux 2012; Carrillo et al. 2013).

The minimization problem can be expressed in the synthesis formulation, where the image \mathbf{x}^* is synthesized from dictionary elements by $\mathbf{x}^* = \Psi \boldsymbol{\alpha}^*$. Alternatively, the problem is expressed in the so-called analysis setting where the image \mathbf{x}^* is recovered directly while promoting the sparsity of $\Psi^* \mathbf{x}^*$. If Ψ is a tight frame, the adjoint Ψ^* is equal to the self-inverse Ψ^\dagger and the two formulations are equivalent. However, in a general overcomplete dictionary (such as a concatenation of two bases) the analysis formulation is often found to yield superior reconstruction quality (e.g. Carrillo et al. 2012, 2013).

The data fidelity term is typically the ℓ_2 -norm of the residuals (which can also be interpreted as the log-likelihood for the case of Gaussian noise). The optimization functional in the unconstrained, analysis setting then reads

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in X} \|\Phi \mathbf{x} - \mathbf{y}\|_{\ell_2}^2 + \lambda \|\Psi^\dagger \mathbf{x}\|_{\ell_1}, \quad (7)$$

assuming identity covariance (typically measurements are weighted to have unit variance; as discussed in Section 4.1).

In the unconstrained setting, the optimization depends on a good choice of the hyperparameter λ to find a good balance between the data fidelity and the sparsity of the signal. The problem can also be formulated as a constrained optimization problem

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in X} \|\Psi^\dagger \mathbf{x}\|_{\ell_1}, \quad \text{s.t.} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_{\ell_2} < \epsilon, \quad (8)$$

where we seek the most sparse $\Psi^\dagger \mathbf{x}$ that satisfies the constraint with some $\epsilon > 0$ on the data misfit. The hyperparameter ϵ can be estimated from an estimate of the noise level (Pratley et al. 2018).

3.1.2 Proximal optimization algorithms

The optimization problems described in the previous subsection can often be solved by proximal optimization algorithms that leverage

a proximity (or proximal) operator. The proximity operator of a (proper, semi-continuous) convex function λh (with $\lambda > 0$) maps $\mathbf{v} \in \mathbb{R}^N$ to a unique solution to the (strongly convex) minimization problem

$$\text{prox}_{\lambda h}(\mathbf{v}) = \arg \min_{\mathbf{x} \in X} \lambda h(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_{\ell_2}^2. \quad (9)$$

The parameter λ sets the balance between the squared ℓ_2 -distance to \mathbf{v} and the value of h . Many common proximal operators admit an analytical solution or at least a linear time iterative solution. The fixed point of a proximal operator is the global minimum of h (Boyd & Vandenberghe 2004; Combettes & Pesquet 2011).

Proximal splitting methods use proximal operators to estimate the solution to the inverse problem by splitting the objective function in separate steps for the different optimization functionals. A review of different proximal splitting methods can be found in Combettes & Pesquet (2011). The simplest proximal splitting algorithm is the proximal gradient method, which consists of a gradient update step, followed a proximal update step:

$$\mathbf{x}_{i+1} = \text{prox}_{\lambda \mathcal{S}}(\mathbf{x}_i - \nabla \mathcal{L}(\Phi \mathbf{x}_i)). \quad (10)$$

In this manner, sparsity-promoting priors $\mathcal{S}(\mathbf{x})$ that are not differentiable can be supported (e.g. ℓ_1 sparsity in a wavelet basis).

Proximal splitting methods have found numerous applications to radio interferometric imaging problems in astronomy: Douglas–Rachford splitting (e.g. Carrillo et al. 2012, 2013), simultaneous-direction method of multipliers (e.g. Carrillo, McEwen & Wiaux 2014), alternating direction method of multipliers (e.g. Pratley et al. 2018; Pratley & McEwen 2021), and the proximal gradient method (Cai, Pereyra & McEwen 2018a, b; Cai, Pratley & McEwen 2019). Besides reconstruction, proximal methods have also been developed to perform uncertainty quantification for radio interferometric imaging (Cai et al. 2018a, b; Cai, McEwen & Pereyra 2022).

In this paper, we compare our new learned methods with a primal dual hybrid gradient (Chambolle & Pock 2011) method for finding a solution to the constrained analysis problem (equation 8) as described in Onose et al. (2016). For the dictionary representation of our signal we chose a representation similar to the one used in the sparsity averaging reweighted analysis (SARA; Carrillo et al. 2012), i.e. a combination of the Dirac basis as well as the first eight Daubechies wavelets, Db1–Db8 (Daubechies 1992), which was shown to work well for astronomical reconstruction (Carrillo et al. 2012, 2013).

3.2 Learned methods

Traditional approaches handle the ill-posedness of the inverse problems by using prior information through the use of handcrafted regularizers that, while general, fail to capture detailed prior information of real data. Learned methods attempt to overcome this by instead enforcing the prior information implicitly specified by training data, ensuring that the prior information promotes images which are in some sense similar to the training data.

Since learned methods rely on learning the prior information for the imaging problem from the data they are provided, reconstructive power depends on the quality and quantity of the training data provided. Furthermore, learned methods have to be trained before evaluation can take place, a process that may take some time, yet only has to be performed once. Once training is done, imaging can then be performed rapidly. If training data in the form of input-output pairs are available, the network can be trained in a supervised approach, with the network receiving the input measurements and their respective targets. When input-output pairs are not available,

the networks can be learned adversarially, where the network is trained using a distribution of measurements to fit an independent distribution of target outputs, which is sometimes interpreted as semi-supervised learning. If no training outputs are available the methods need to be learned in a self-supervised way.

Learned methods that are completely independent of the measurement operator and learn a direct mapping from measurements to target are generally not effective since they are difficult to train and require huge volumes of training data (Adler & Öktem 2017). Therefore, most learned approaches incorporate the measurement operator in some capacity to encode or leverage the mapping from the measurement to the reconstruction space. The learned network is then constructed around that. As briefly overviewed in Section 1, learned data-driven imaging approaches broadly fall into three categories, differentiated by the degree to which model-based information such as the measurement operator is encoded or leveraged. By making greater use of the measurement operator superior performance can often be achieved but at the cost of increased computational time. We subsequently review each of these three classes of approach.

3.2.1 Learned regularization

One method for learning the prior information from training data is by using a learned regularizer in conjunction with traditional optimization schemes. These algorithms aim to replace the regularizer \mathcal{S} in equation (6) with a learned regularizer, encoding the prior information in the training data.

A common approach is to replace the application of the proximal operators with a learned map to compute the update. These methods typically adopt PnP denoisers that are used instead of a proximal update step (Venkatakrisnan et al. 2013; Ryu et al. 2019). A variant of a PnP denoiser was applied to radio interferometry by Terris et al. (2022).

Various other pioneering approaches to learned regularization have also been considered. The regularizer can be formed as the norm of a learned dictionary that translates to a sparse representation of the data (Xu et al. 2012) or by using a constraint based on a learned scattering transform (Dokmanić et al. 2016) in place of a traditional regularizer. Alternatively, methods are used that implement deep neural networks to act as regularizers (e.g. Kobler et al. 2020; Li et al. 2020) as well as methods that train adversarially learned neural networks (e.g. Lunz et al. 2018; Mukherjee et al. 2020).

While learned regularization approaches often achieve excellent reconstruction quality since they make full use of the measurement operator, along with learned prior information, they are computationally demanding since the full measurement operated must be evaluated for each iteration.

3.2.2 Learned sequential methods

An alternative approach is to consider a model that is split up into a sequence of operations. Sequential models are a composition of a learned operator acting in the data space $\mathbf{C}_\theta : Y \rightarrow Y$, an adjoint or pseudo-inverse mapping from the data to the reconstruction space, $\mathbf{A} : Y \rightarrow X$, and a learned operator acting in the image space $\mathbf{B}_\theta : X \rightarrow X$, where learned operators depend on the parameters θ (e.g. Zheng et al. 2020):

$$\Phi_\theta^\dagger = \mathbf{B}_\theta \circ \mathbf{A} \circ \mathbf{C}_\theta. \quad (11)$$

When either the operator in the reconstruction domain, \mathbf{B}_θ , or in the data domain, \mathbf{C}_θ , is set to be the identity operator, we

recover learned pre-processing or learned post-processing methods, respectively. Most sequential methods are of the post-processing type since they are relatively easy to train as learning of the network is decoupled from the mapping from measurement to reconstruction space; the post-processing network can thus be trained independently, reducing training time substantially (e.g. Chen et al. 2017; Jin et al. 2017; Yi & Babyn 2018). Post-processing methods have been used in astronomy using superresolution networks (Allam 2016; Connor et al. 2022), denoisers (Terris et al. 2019), and convolutional autoencoders (Gheller & Vazza 2021).

Post-processing methods only apply the adjoint or pseudo-inverse of the measurement operator once and so are highly computationally efficient. However, since the measurement operator is not continually leveraged, reconstruction quality suffers as a consequence. To compensate sequential models typically employ more elaborate network architectures (such as U-Nets; Ronneberger, Fischer & Brox 2015) than those adopted in learned iterative methods. Nevertheless, learned post-processing approaches can sometimes struggle to achieve the reconstruction quality of learned iterative approaches.

3.2.3 Learned iterative methods

Learned iterative methods unroll a small, fixed number of iterations of an iterative solver, e.g. proximal gradient method, and replace the proximal operators with learned CNNs (Gregor & LeCun 2010). The CNNs used are typically small feed-forward networks with just a few convolutional layers (e.g. Adler & Öktem 2017; Putzky & Welling 2017; Yang et al. 2017), in contrast to the architectures used in sequential models. To turn the proximal gradient method of equation (10) into a learned method, we replace the proximal operator with a learned network $\Lambda_{\theta,i}$:

$$\mathbf{x}_{i+1} = \Lambda_{\theta,i}(\mathbf{x}_i - \nabla \mathcal{L}(\Phi \mathbf{x}_i)), \quad (12)$$

where $\Lambda_{\theta,i}$ can be a different network for each iteration k or one network, $\Lambda_{\theta,i} = \Lambda_\theta$. The latter scenario results in learning an approximation to the proximal operator.

Hauptmann et al. (2020) propose evaluating the operator at different, increasingly finer resolutions/scales such that the full resolution operator only needs to be evaluated once. Combining this with a multiscale neural net like the U-Net (Ronneberger et al. 2015) cuts down evaluation time (and thus training time) significantly, while also reducing memory requirements. Other advancements are considered by learning adversarially trained versions of existing models (Mukherjee, Öktem & Schönlieb 2021b). Also, combining approaches using learned regularizers with learned iterative methods, provides algorithms that are closer to the traditional optimization methods and allow proving rigorous convergence results (Mukherjee et al. 2021a). Finally, the inclusion of the explicit measurement operator in the learned iterative models improves the robustness and generalizability (Boink, Manohar & Brune 2020), while at the same time reducing the number of trainable parameters and hence reducing the requirement of large training data volumes.

Since learned iterative methods only unroll a small number of iterations, the measurement operator is only evaluated sparingly, making the learned iterative approaches typically faster than the iterative solvers on which they are based. While the integration of the measurement operator in the network does require its evaluation during training, resulting in considerably longer training times than for the sequential models, encoding the measurement operator in the model typically results in superior performance to learned sequential methods, for only a modest increase in computational time.

4. MODELLING THE SPIDER INSTRUMENT

In order to reconstruct an image from interferometric measurements, we need a model for our measurement process in the form of a measurement operator. The accuracy of the measurement operator is a critical factor for the accuracy of the reconstruction and its computational complexity is reflected in both training and reconstruction time.

For interferometric imaging, the measurement process is given by equation (1). In a discretized version of equation (1), following the notation of equation (5), the measurement operator $\Phi : \mathbb{R}^N \rightarrow \mathbb{C}^M$ is a non-uniform discrete Fourier transform (NUDFT) mapping an image in \mathbb{R}^N to M non-uniformly distributed Fourier measurements.

For an image $\mathbf{x} = f(\chi)$, with spatial pixel coordinates χ_{ij} at the i th and j th pixels of the image, and Fourier measurements $\mathbf{y} = \hat{f}(\xi)$ at non-uniformly distributed Fourier coordinates $\xi_k = (u_k, v_k)$ (in contrast to equations 3 and 4), the unitary NUDFT can be expressed as

$$\hat{f}(\xi_k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(\chi_{ij}) e^{-i\chi_{ij} \cdot \xi_k}, \quad (13)$$

which is a finite sum over all pixels (i, j) of an $N_1 \times N_2$ image with a total number of $N = N_1 \times N_2$ pixels.

To construct the dirty image $f_D(\chi)$, the adjoint of the measurement operator is applied, which maps from the non-uniformly distributed Fourier measurements to a uniformly sampled image

$$f_D(\chi_{ij}) = \frac{1}{\sqrt{N}} \sum_{k=1}^M \hat{f}(\xi_k) e^{i\chi_{ij} \cdot \xi_k}. \quad (14)$$

We can approximate the inverse Fourier transform (equation 2) by evaluating a weighted, finite Fourier series at the non-uniformly distributed Fourier coordinates. The accuracy of the resulting pseudo-inverse DFT is determined by the choice of sampling ξ_k and the corresponding measurement weights $w(\xi_k)$ and is given by

$$f(\chi_{ij}) \approx \sum_{k=1}^M w(\xi_k) \hat{f}(\xi_k) e^{i\chi_{ij} \cdot \xi_k}. \quad (15)$$

The complexity of a naive realization of equations (13), (14), and (15) would be $\mathcal{O}(NM)$; however, efficient numerical NUDFT schemes make use of interpolation to map the non-uniformly distributed measurements to a uniformly sampled grid and take advantage of the fast Fourier transform (FFT) to reduce complexity.

In this section, we introduce two efficient methods for modelling the measurement operator of the SPIDER instrument. First, we consider the non-uniform fast Fourier transform (NUFFT, Duijndam & Schonewille 1997), which can be used for arbitrary sampling distributions. Second, we present a new approach that exploits the specific sampling distribution of the SPIDER instrument, making use of similarities of the problem to that of parallel Radon transform of X-ray tomography.

4.1 Non-uniform fast Fourier transform

Using FFTs with a (de)gridding interpolation operator to approximate the non-uniformly distributed Fourier measurements lowers the computational cost significantly. The standard operator used in radio interferometry can be expressed as a composition of operators (e.g. Pratley et al. 2018):

$$\Phi = GFZD, \quad (16)$$

where $D : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is an operator that corrects for the gridding operation (discussed further below), $Z : \mathbb{R}^N \rightarrow \mathbb{R}^{\alpha^2 N}$ is a zero-

padding operator that zero pads the image in the spatial dimension to provide oversampling of a factor α in each direction in the Fourier domain, $F : \mathbb{C}^{\alpha^2 N} \rightarrow \mathbb{C}^{\alpha^2 N}$ is a 2D unitary FFT operator, and $G : \mathbb{C}^{\alpha^2 N} \rightarrow \mathbb{C}^M$ is the degridding operator that interpolates the measurements off of the uniform Fourier grid via convolution with a spreading kernel centred at the irregularly spaced measurements points in the Fourier domain. The operator D corrects for the convolution with this spreading kernel by a pointwise division of the image with the Fourier transform of the spreading kernel.

Equation (16) describes the measurement process, going from the observable to the measurement space. The adjoint of this operation, $\Phi^* : \mathbb{C}^M \rightarrow \mathbb{R}^N$ is readily obtained as

$$\Phi^* = DZ^*F^\dagger G^*, \quad (17)$$

where we made use of the self-adjointness of the convolution correction operator ($D^* = D$). The gridding operation $G^* : \mathbb{C}^M \rightarrow \mathbb{C}^{\alpha^2 N}$ convolves the non-uniformly distributed measurements with the spreading kernel to grid them on a uniformly sampled mesh, on which the inverse FFT, $F^\dagger : \mathbb{C}^{\alpha^2 N} \rightarrow \mathbb{C}^{\alpha^2 N}$, is applied to obtain an image, which is restricted to \mathbb{R}^N via symmetric cropping through the operator $Z^* : \mathbb{R}^{\alpha^2 N} \rightarrow \mathbb{R}^N$, and deconvolved via pointwise division with the Fourier inverse of the spreading kernel, using the operator D , to correct for the gridding operation.

The efficiency of the NUFFT schemes hinges upon a suitable choice of the spreading kernel, in particular its localization in the spatial and frequency domains and the ease of computing the Fourier transform of the kernel for deconvolution. The kernel needs to have a small support in frequency space to be computationally inexpensive, while also having a small support in the image domain as to minimize the effects of aliasing. Here, we adopt the Kaiser–Bessel (KB) kernel (Jackson et al. 1991; Fessler & Sutton 2003). The kernel is truncated in the Fourier domain to limit its support and therefore its computational cost. Since these kernels are linearly separable, i.e. the kernel can be written $\hat{c}(u, v) = \hat{c}(u)\hat{c}(v)$, we consider the kernel in only one dimension:

$$\hat{c}(u) = \begin{cases} \frac{1}{J} \frac{I_0\left(\beta\sqrt{1-\left(\frac{2u}{J}\right)^2}\right)}{I_0(\beta)} & \text{for } |u| \leq J/2, \\ 0 & \text{for } |u| > J/2, \end{cases} \quad (18)$$

where I_0 is the zeroth-order modified KB function, β determines the shape of the kernel, and J the support of the kernel. The correction for applying this convolutional kernel can be calculated analytically by taking the inverse Fourier transform of the kernel to get (Jackson et al. 1991; Fessler & Sutton 2003)

$$c(x) = \left[\frac{\sin\left(\sqrt{\pi^2 x^2 J^2 - \beta^2}\right)}{\sqrt{\pi^2 x^2 J^2 - \beta^2}} \right]^{-1}. \quad (19)$$

The correction operator can then be found by calculating

$$D(\chi_{ij}) = c\left(\frac{i}{N_1} - \frac{1}{2}\right) c\left(\frac{j}{N_2} - \frac{1}{2}\right). \quad (20)$$

Using $\beta = 2.34J$ for the spread of this KB kernel gives similar performance to the optimal min–max interpolation kernel proposed by Fessler & Sutton (2003).

The pseudo-inverse of the NUFFT is obtained by including the measurement weights in the Fourier sum:

$$\Phi^\dagger = \Phi^* W = DZ^* F G^* W, \quad (21)$$

where W is a diagonal matrix with elements $W_{kk} = w(\xi_k)$ given by the weight for each of the measurements.

In radio astronomy, interferometric measurements are typically weighted according to one of three weighting schemes: natural, uniform, or robust weighting (Taylor et al. 1999). Natural weighting maximizes the sensitivity of the observation and weights each visibility by the uncertainty on the measurement $\mathbf{W}_{k,k}^{\text{Natural}} = \sigma_k^{-2}$, with σ_k the uncertainty on the measured visibility \mathbf{y}_k . Natural weighting is akin to statistical whitening of data and ensures unit variance of the uncertainty on the measurements and gives the highest signal-to-noise ratio for detecting weak sources.

Since there is typically a higher density of samples near the (u , v) origin in radio interferometry, natural weighting emphasizes low-frequency measurements, which can be undesirable when imaging sources with both large- and small-scale structure. Uniform weighting is an alternative weighting scheme that accounts for this by weighting the density of the sampling distribution by $\mathbf{W}_{k,k}^{\text{Uniform}} = \sigma_k^{-2} N_s^{-1}(k)$, where $N_s(k)$ is typically the number of measurements in a symmetrical region (either square or circular) with width s around the measurement \mathbf{y}_k . If the measurement uncertainties are varying this can be replaced by the sum of the reliability weights in this region $N_s = \sum_{k'} \sigma_{k'}^{-2}$, where k' are the indices of the measurements in the local region around measurement k . The width s of the region is typically chosen to be the size of the size of the elements of the Fourier grid.

Lastly, robust weighting considers a trade-off between natural and uniform weighting, controlled by a robustness parameter. In this paper, we use uniform weighting and weight the measurements according to the density of the sampling distribution as well as their uncertainties.

4.2 Sub-scale operators

An approach to reduce the computational cost of the measurement operator is to evaluate it at a lower resolution and restricted Fourier space. The forward and adjoint of the measurement operator can be evaluated at a series of scales by applying a filter bank of low pass and high pass partition of unity filters; see Pan & Betcke (2022) for the details of such filter banks. Instead of using smooth partition of unity filters as used in Candès et al. (2006) and Pan & Betcke (2022), we use binary 0–1 partition of unity filters. Since the NUFFT evaluates the FFT on an upsampled grid by zero-padding the image, the periodicity in the image domain and the corresponding decay of the Fourier coefficients are reinstated. We refer to these sub-scale measurement operators as $\Phi_i : \mathbb{R}^{N_i} \rightarrow \mathbb{C}^{M_i}$, its adjoint $\Phi_i^* : \mathbb{C}^{M_i} \rightarrow \mathbb{R}^{N_i}$, and its pseudo-inverse $\Phi_i^\dagger : \mathbb{C}^{M_i} \rightarrow \mathbb{R}^{N_i}$, operating on the reduced image scale N_i and the number of measurements restricted to the reduced Fourier space, M_i . Since these sub-scale operators are evaluated on a smaller image scale as well as a restricted Fourier space, they are considerably more computationally efficient than the full-scale measurement operators.

4.3 NU-Radon method

For the NUFFT, the measurement weights are determined by considering the 2D sampling density of the instrument. In this section, we propose an alternate, principled approach to derive these weights based on the similarities of the SPIDER sampling to the sampling induced by the 2D Radon transform.

The Radon transform,

$$\mathbf{R}f(r, \phi) = \mathbf{R}_\phi f(r) = \int_{x=\phi=r} f(x) dx, \quad (22)$$

models the linear attenuation of photons passing through an object. It maps the attenuation to a family of integrals along parallel lines,

parametrized in polar coordinates ϕ , the angle of the projection, and r the radial distance from the origin along that line. The Fourier slice theorem

$$\mathbf{R}_\phi^\dagger f(\rho) = \hat{f}(\rho, \phi), \quad (23)$$

expresses the equivalence of the 2D Fourier transform in polar coordinates and the 1D Fourier transform along the detector coordinate of the Radon transform, where ρ is the distance in the Fourier domain along the radial spoke at angle ϕ .

The adjoint operation of the Radon transform is the back-projection which can be used to recover the dirty reconstruction through

$$f_D(r, \phi) = \int_0^\pi \int_{-\infty}^\infty \mathbf{R}_\phi^\dagger f(\rho) e^{i2\pi r \rho} d\rho d\phi. \quad (24)$$

The inverse of the radon transform is the *filtered* back-projection and includes a ramp filter $|\rho|$, which corresponds to the Jacobian of the change from Cartesian to polar coordinates and which boosts the power of high-frequency measurements:

$$f(r, \phi) = \int_0^\pi \int_{-\infty}^\infty \mathbf{R}_\phi^\dagger f(\rho) e^{i2\pi r \rho} |\rho| d\rho d\phi. \quad (25)$$

Using the Fourier slice theorem, we can express the NUDFT (equation 13) in terms of the Radon transform, and in polar coordinates, by

$$\hat{f}(\xi_k) = \hat{f}(\rho_k, \phi_k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{\sqrt{N}} \mathbf{R}_{\phi_k} f(r_i) e^{-i r_i \rho_k}, \quad (26)$$

where the measurements are gathered by taking a 1D NUDFT of the Radon transform of the signal along a spoke at angle ϕ_k and \sqrt{N} corresponds to the diameter of the circle inscribed into the square image. Similarly, we discretize the back-projection (equation 24) to obtain the dirty reconstruction

$$f_D(\mathbf{x}) = \frac{1}{\sqrt{N}} \sum_{p=1}^P \sum_{q=1}^Q \mathbf{R}_{\phi_p}^\dagger f(\rho_q) e^{i \rho_q \mathbf{x} \cdot \phi_p}, \quad (27)$$

where we Fourier transform the Q non-uniformly distributed Fourier measurements at radial distances ρ_q for each of the spokes, followed by back-projecting the signals for each of the P projection angles ϕ_p .

Both the forward and adjoint operations of the measurement operator use a 1D NUDFT which can be accelerated by using (de)gridding in the form of a 1D NUFFT. The combined method of using the Radon transform and the NUFFT can be described as

$$\Phi = \mathbf{N} \mathbf{R}, \quad (28)$$

where $\mathbf{R} : \mathbb{R}^N \rightarrow \mathbb{R}^{P\sqrt{N}}$ is the Radon transform, and $\mathbf{N} : \mathbb{C}^M \rightarrow \mathbb{R}^{P\sqrt{N}}$ is the 1D NUFFT. Similarly, the adjoint operation is given by

$$\Phi^* = \mathbf{R}^* \mathbf{N}^*, \quad (29)$$

where $\mathbf{R}^* : \mathbb{R}^{P\sqrt{N}} \rightarrow \mathbb{R}^N$ is the back-projection operation, and $\mathbf{N}^* : \mathbb{C}^M \rightarrow \mathbb{C}^N$ is the 1D adjoint NUFFT.

To obtain the pseudo-inverse of the measurement operator, part of the measurement weights are now informed by the ramp filter in the inverse of the Radon transform (equation 25). The ramp filter for our non-uniformly distributed measurements is calculated as $w_{\text{Radon}}(\xi) = \|\xi\|_{\ell_2}$. The pseudo-inverse of the measurement operator is then defined as

$$\Phi^\dagger = \mathbf{R}^\dagger \mathbf{N}^\dagger, \quad (30)$$

where $\mathbf{R}^\dagger : \mathbb{R}^{P\sqrt{N}} \rightarrow \mathbb{R}^N$ is the filtered back-projection (with weights $w_{\text{Radon}}(\xi)$), and $\mathbf{N}^\dagger : \mathbb{C}^M \rightarrow \mathbb{C}^N$ is the 1D inverse NUFFT with

Table 2. The computational complexity of the different methods for modelling the measurement operator of the SPIDER instrument in terms of the number of pixels N , the number of measurements M , the number of projection angles P , and the support of the gridding kernels in pixels J .

	Fourier transform	Gridding	Radon transform	Total
NUDFT	$\mathcal{O}(NM)$	–	–	$\mathcal{O}(NM)$
NUFFT	$\mathcal{O}(N \log N)$	$\mathcal{O}(MJ^2)$	–	$\mathcal{O}(N \log N + MJ^2)$
NU-Radon	$\mathcal{O}(\sqrt{N} \log \sqrt{N})$	$\mathcal{O}(MJ)$	$\mathcal{O}(NP)$	$\mathcal{O}(\sqrt{N} \log \sqrt{N} + MJ + NP)$

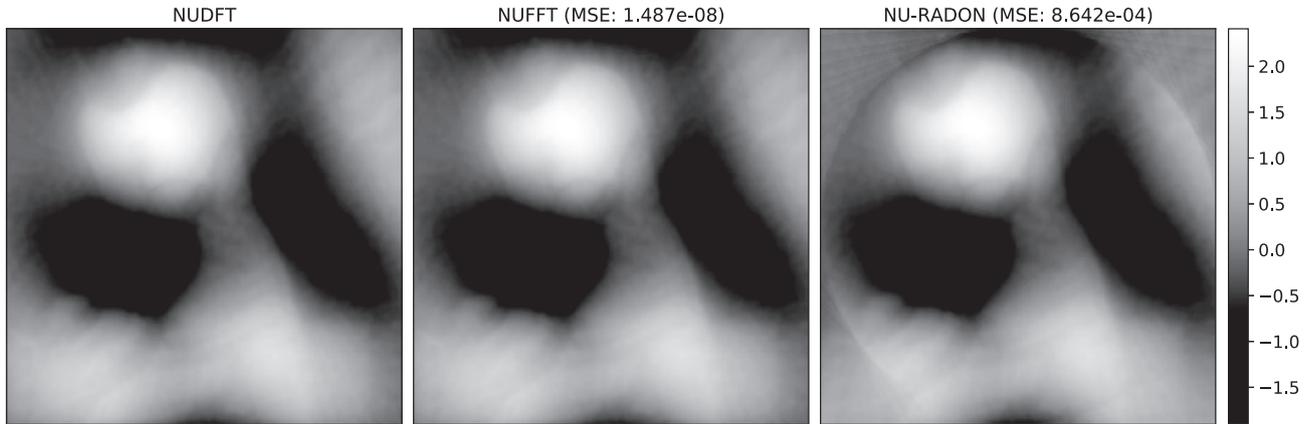


Figure 3. Generating a dirty reconstruction by applying the adjoint of the measurement operator, using the NUDFT, to interferometric measurements of the SPIDER instrument (left panel), compared with accelerated methods for approximation of the adjoint operation using the NUFFT (centre panel) and the NU-Radon (right panel) approaches as detailed in Sections 4.1 and 4.3, respectively. The MSE of the two accelerated approaches compared with the NUDFT is shown, where the error is only calculated inside the circular aperture that limits the NU-Radon method.

measurements weights based on the sampling density of the lenslets along one spoke.

4.4 Comparison

We implement both the NUFFT and NU-Radon methods in our LEIA¹ code. The NUFFT approach is implemented on the CPU using NUMPY² as a backend, as well as with GPU acceleration implemented in TENSORFLOW.³ The NU-Radon method is implemented by using a modified version of the SCIKIT-IMAGE⁴ Radon transform.

An overview of the computational complexity of the algorithms can be found in Table 2, where the complexity is split into three operations: calculating the Fourier transform, (de)gridding non-uniformly distributed measurements, and computing the Radon transform (when relevant).⁵ The NUDFT is computationally the most expensive approach, since the evaluation of the DFT takes $\mathcal{O}(NM)$ operations. Both the NUFFT and NU-Radon reduce the computational complexity by utilizing the FFT for calculating the Fourier transform, though these gains are offset by the computational cost for the gridding of the measurements and/or the calculation of the (inverse) Radon transform.

For the particular SPIDER configuration considered ($N = 256 \times 256$, $M = 4440$, and $P = 37$), the NUFFT is the best approach in terms of computational complexity. It might be beneficial to use the NU-Radon method when the number of measurements is large (and

therefore the computation time is dominated by the (de)gridding operations) or when the number of image pixels is large and the number of spokes is modest ($P \leq \log N$).

Fig. 3 shows the effect of reconstructing a dirty image using the adjoint of the measurement operator applied to a set of simulated Fourier measurements taken with the SPIDER instrument. For the configuration of the SPIDER instrument, it appears that the NUFFT is a more accurate approximation to the NUDFT.

Fig. 3 indicates that the dirty reconstructions are mostly dominated by low-frequency structures. The measurement weights used in the pseudo-inverse operations boost the power of high-frequency measurements by weighting the measurements based on the 2D sampling density for the NUFFT and the 1D sampling density along each spoke for the NU-Radon approach (which also includes the weights of the filtered back-projection).

Fig. 4 compares the different pseudo-inverse implementations. The figure indicates that the pseudo-inverse NUFFT provides a good trade-off between introducing high-frequency information and not having strong high-frequency artefacts. The pseudo-inverse NU-Radon implementation provides sharper features, yet also includes more high-frequency artefacts. In the remainder of the paper, we use the NUFFT approach for our experiments due to the lower computational cost of its implementation and its generalizability to any sampling distribution for future applications.

5. LEARNED INTERFEROMETRIC IMAGING FOR SPIDER

In this section, we present our two learned approaches for the SPIDER imaging problem. Both approaches reconstruct images from interferometric measurements and use the adjoint or pseudo-inverse of the measurement operator to create an initial reconstruction. Our approaches fit within the existing learned post-processing framework

¹<https://github.com/astro-informatics/LeIA>

²<https://numpy.org/>

³<https://www.tensorflow.org/>

⁴<https://scikit-image.org/>

⁵Recall N denotes the number of image pixels, M the number of measurements, J the support of the spreading kernel, and P the number of spokes of the SPIDER instrument.

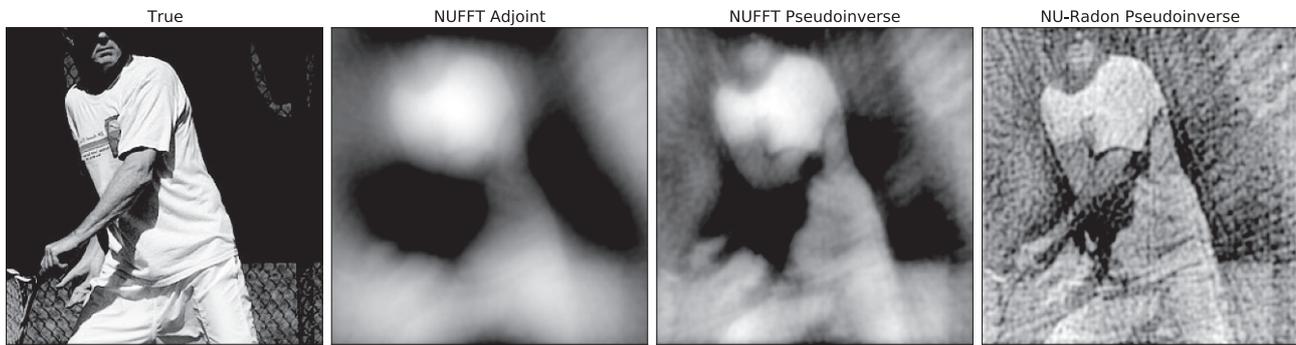


Figure 4. Measurements are simulated from the original image from the COCO data set of natural images shown in the left-most image, using a NUFFT with the SPIDER sampling pattern. Generating images by applying adjoint NUFFT operation, the pseudo-inverse NUFFT operation, and the pseudo-inverse NU-Radon operation are shown, respectively, from left to right.

(e.g. Jin et al. 2017), with the second method introducing a novel post-processing architecture.

The first method is a post-processing learned sequential method to improve the initial reconstruction estimated by the pseudo-inverse and remove artefacts. This approach is highly computationally efficient, requiring only one application of the pseudo-inverse of the measurement operator to compute the initial reconstruction. We expect final reconstruction quality to be good but limited since no further knowledge of the measurement operator is exploited beyond the initial reconstruction.

The second approach is a learned iterative method and differs to the former by including updates to the reconstruction that utilize knowledge of the measurement operator. Since additional application of the measurement operator is required this results in a modest increase in computational time compared with our post-processing method. However, we expect this to be offset by an increase in reconstruction quality due to further exploitation of the measurement operator.

We do not consider learned regularization approaches further since the iterative nature of such approaches would result in a significant increase in computational time, and our primary objective is a very low computational cost during imaging in order to facilitate real-time imaging with SPIDER.

5.1 Learned post-processing (U-Net)

Learned post-processing methods are a special case of the sequential methods discussed in Section 3.2.2, where the learned operator applied in the data domain, C_θ , is replaced with an identity operator. This makes the networks easier to train as the training is performed in the image domain without the need to invoke the measurement operator, speeding up the training process. Our method only post-processes an initial reconstruction and is in this sense similar to a denoiser. The solution can be written as

$$\mathbf{x}^* = \Phi_\theta^\dagger \mathbf{y} = \Lambda_\theta \Phi^\dagger \mathbf{y}, \quad (31)$$

with Λ_θ the learned network, and Φ^\dagger the pseudo-inverse of the measurement operator. Fig. 5 shows a schematic representation of this approach.

The architecture of our network is based on the U-Net (Ronneberger et al. 2015) network architecture, which was originally designed for biomedical image segmentation. U-Nets can also be used for denoising problems by replacing the two segmentation specific output layers with a single output layer returning the denoised image. In this form, U-Net is a convolutional autoencoder with skip

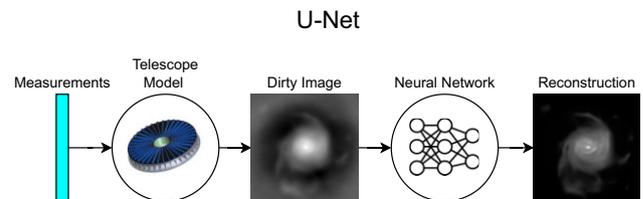


Figure 5. The learned post-processing approach takes the interferometric measurements and uses the telescope model to create an initial reconstruction, in our case the dirty image. This is then passed through the learned post-processing network to create the final reconstruction.

connections at each of the scales. Our network includes blocks of 2D convolutions with batch normalization and rectified linear unit (ReLU) activation functions at each scale, followed by MaxPool layers to sub-sample on a downward branch. In the decoder branch, the upsampling is achieved by transposed convolution layers. A schematic representation of the architecture can be found in Fig. 7.

Our learned post-processing is computationally efficient as it only evaluates the measurement operator once per image for both evaluation and training. However, this also limits the extent to which the network can utilize the measurement model.

5.2 Learned iterative method (gradient U-Net)

An alternative approach is to apply learned iterative methods that evaluate the measurement operator several times to improve the reconstruction quality by including measurement information at several stages of the reconstruction process. While some methods in this category closely mimic optimization methods, the method we propose takes inspiration from multiscale methods. Our approach, the gradient U-Net (GU-Net), expands on the U-Net architecture by calculating the gradient of the data fidelity after every downsampling and after every upsampling operation, as well as after the input layer, to incorporate measurement information at every scale of the network. Fig. 6 shows a diagram summarizing the learned iterative reconstruction process. How we implement the network is described below and a schematic representation of the network can be found in Fig. 7.

In order to add this measurement information at different scales of the network, we use sub-scale operators, $\Phi_i : \mathbb{R}^{N_i} \rightarrow \mathbb{C}^{M_i}$ (see Section 4.2), where the size of the image space N_i decreases by powers of four (the images are halved in each direction; $N_i = N/4^i$), and the size of the measurement space M_i is determined by applying

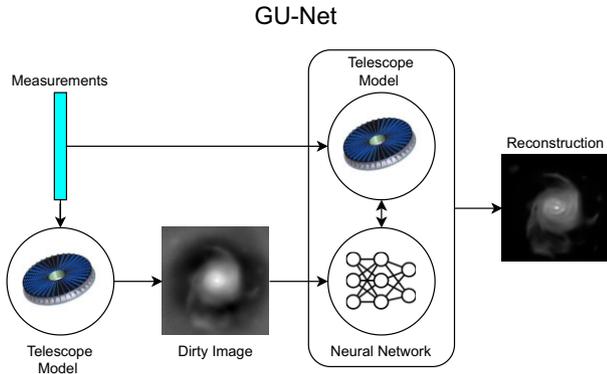


Figure 6. For the learned iterative approach, the interferometric measurements are used to create an initial reconstruction using the telescope model, in our case this results in the dirty image. This is then used as input for the learned iterative network. At several stages of the network, the telescope model is used together with the original measurements to provide model-based updates to the reconstruction.

a low-pass filter that restricts the measurements to $\{(u, v): u \leq u_{\max}/2^i \wedge v \leq v_{\max}/2^i\}$, where u_{\max} and v_{\max} are the longest baselines observed in each direction.

The first method to incorporate the measurement operator into the reconstruction process is to calculate the gradient of the data fidelity at each scale of the network: $\mathcal{L}(\Phi_i \mathbf{x}_i, \mathbf{y}_i) = \frac{1}{2} \|\Phi_i \mathbf{x}_i - \mathbf{y}_i\|_{\ell_2}^2$. This is given by

$$\nabla_{\mathbf{x}_i} \mathcal{L}(\Phi_i \mathbf{x}_i, \mathbf{y}_i) \propto \Phi_i^* (\Phi_i \mathbf{x}_i - \mathbf{y}_i), \quad (32)$$

where $\mathbf{x}_i \in \mathbb{R}^{N_i}$ is the first channel of the current iterate at scale i of the network and $\mathbf{y}_i \in \mathbb{C}^{M_i}$ is the low-pass filtered measurement vector of visibilities.

By analogy to the Radon transform, the gradient does not contain the high-frequency features which are reinstated via the ramp filter in the filtered back-projection. Therefore, we also include the filtered gradient where the high-frequency features are boosted using a sampling density based filter. This is analogous to taking the gradient of a weighted least-squares norm with the measurement weights the same as the uniform weights described in Section 4.1:

$$\nabla_{\mathbf{x}_i}^f \mathcal{L}(\Phi_i \mathbf{x}_i, \mathbf{y}_i) \propto \Phi_i^* (\mathbf{W}_i^{\text{Uniform}} (\Phi_i \mathbf{x}_i - \mathbf{y}_i)), \quad (33)$$

where $\mathbf{W}_i^{\text{Uniform}} \in \mathbb{R}^{M_i \times M_i}$ is a diagonal matrix with the measurement weights for the sub-selected measurements as the diagonal elements.

In addition to the gradient information, we also add the scale-restricted dirty reconstruction (adjoint of the measurement operator applied to the noisy measurements):

$$\mathbf{x}_{i,\text{dirty}} = \Phi_i^* \mathbf{y}_i. \quad (34)$$

In doing so, the network can learn an update between the dirty reconstruction, the current iterate passed down by the network, and the gradient and filtered gradient information based on the current iterate.

The measurement information encoded in the (sub-scale) gradient and filtered gradient, and the dirty reconstruction is added at each scale of the U-Net architecture after downsampling and after upsampling. The added measurement information at each scale of the network is given by

$$\tilde{\mathbf{x}}_i = \Lambda_{i,\theta} (\mathbf{x}_i, \nabla_{\mathbf{x}_i} \mathcal{L}(\Phi_i \mathbf{x}_i, \mathbf{y}_i), \nabla_{\mathbf{x}_i}^f \mathcal{L}(\Phi_i \mathbf{x}_i, \mathbf{y}_i), \Phi_i^* \mathbf{y}_i), \quad (35)$$

where $\Lambda_{i,\theta}$ is the learned convolutional operator to combine the measurement operator information. At each scale of the network, the

gradient information is calculated for the first channel of the current iterate. The learned convolutional operator takes the three channels (gradient, filtered gradient, and dirty reconstruction) representing the measurement information and convolves it with a number of filters matching the number of channels of the current iterate in the network, so as not to dilute the measurement operator information when combining it with the original channels.

Evaluating the added measurement information in equation (35) results in two evaluations of the measurement operator (one each for the gradient and filtered gradient) and three evaluations of the adjoint operator (one each for the gradient and filtered gradient and one for the scale-restricted dirty reconstruction). Since the measurement information is added after down and upsampling, this results in four full-scale evaluations of the forward and six full-scale evaluations of the adjoint of the measurement operator for the added measurement information. An additional evaluation of a single full-scale pseudo-inverse of the measurement operator is also required for the initial reconstruction.

This method is computationally more expensive than the U-Net post-processing method because the measurement operator has to be evaluated several times. However, as most of the evaluations of the measurement operator are at reduced scales, the impact on reconstruction time is kept to a minimum. Furthermore, inclusion of multiple iterations with the (albeit downsampled) measurement operator enhances the impact of the model on the reconstruction, improving robustness and generalizability of the model. Similar approaches that leverage multiscale evaluation of the forward/adjoint operators to reduce computational cost with application in X-ray CT imaging were proposed in Hauptmann et al. (2020) and Trent (2020).

6. EXPERIMENTS AND RESULTS

In order to evaluate the reconstruction performance and computational cost of the learned imaging methods proposed we perform reconstructions on simulated measurements and compare the reconstructions the ground truths. To assess reconstruction quality, we compute the peak signal-to-noise ratio (PSNR) to measure how well the reconstruction matches the measurements and the structural similarity index measure (SSIM, Wang et al. 2004) to assess the structural similarity of the reconstructions. We compare the two learned approaches described in Section 5 to using a reconstruction obtained with the pseudo-inverse (equation 21) and a primal-dual method as described in Section 3.1.2 (the implementation uses OptimusPrimal⁶ and runs for 300 iterations), which produces results on par with the current state-of-the-art in astronomical interferometric reconstruction. Specifically, we discuss the different data sets used, the measurement simulation process, network training, and reconstruction results. Besides assessing the computational cost and the reconstruction quality of the different approaches, we also evaluate the robustness of the methods proposed to additional noise and generalizability to other smaller data sets through transfer learning.

Implementations for the measurement operators, the learned imaging techniques and the routines used to simulate interferometric measurements and train the neural networks can be found in our LEIA⁷ codebase.

⁶<https://github.com/astro-informatics/Optimus-Primal>

⁷<https://github.com/astro-informatics/LeIA>

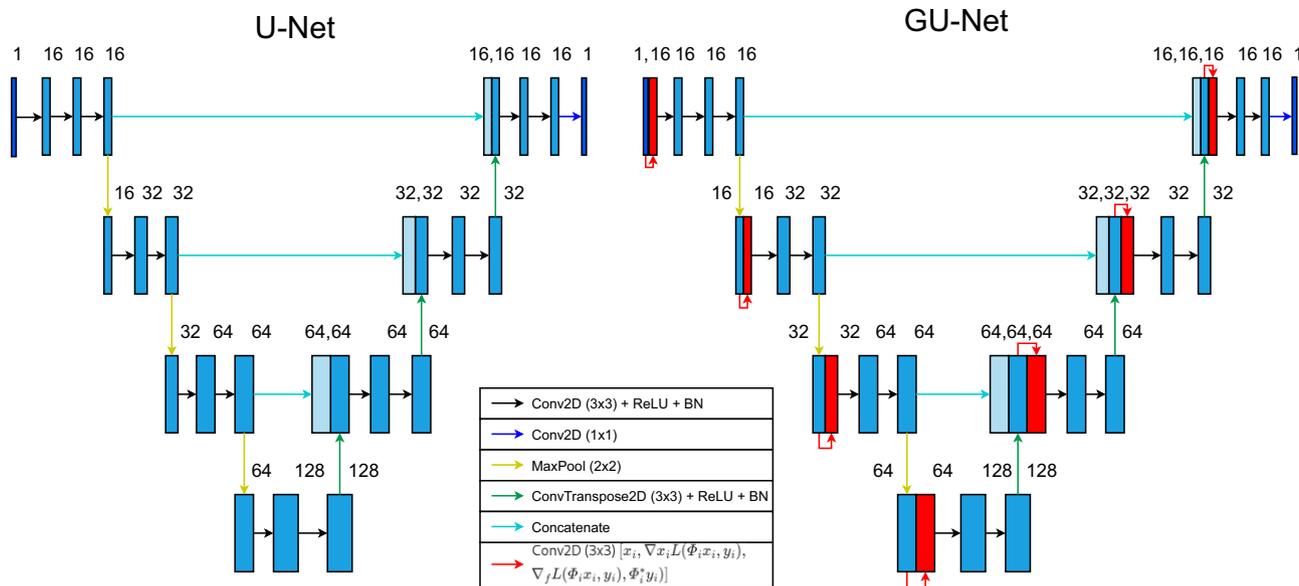


Figure 7. Left panel: The neural network architecture for the U-Net model used for learned post-processing of the initial reconstruction estimated by the pseudo-inverse of the measurement operator applied to the noisy measurements. Right panel: The neural network architecture for the GU-Net model for the learned iterative approach used to reconstruct an image from interferometric measurements. The input image of the network is the pseudo-inverse of the measurement operator applied to the noisy measurements. The model-based operation in this network takes the first channel on each of the scales and calculates the gradient (equation 32) and filtered gradient (equation 33) of the data fidelity, as well as a scale-restricted dirty image (equation 34). These are combined using a CNN layer (equation 35), concatenated to the original channels, and passed on through the rest of the network.

Table 3. Number of full-scale measurement operator calls, reconstruction time averaged over 1000 images of the COCO data set and training time, for the pseudo-inverse, our learned methods, and a variational regularization approach. The reconstruction times for our learned methods are sufficiently low to enable real-time imaging for the SPIDER instrument.

Name	Operator evaluations	Average reconstruction time (ms)	Training time (min)
Pseudo-inverse (1 GPU)	1	5.50	–
U-Net (1 GPU)	1	10.7	~30
GU-Net (1 GPU)	11 ^a	42.1	~100
Primal-dual (300 its, 1 CPU)	600	4.7×10^4	–

Note.

^aRefers to operator evaluation at the finest scale, which dominates the computational time of the GU-Net.

6.1 Data sets

We consider three data sets with different characteristics to evaluate the reconstruction performance of the imaging methods. The three data sets cover different potential use cases of the SPIDER instrument, including standard imaging, astronomical imaging, and Earth observation. All images are converted to greyscale and cropped to a size of 256×256 . For each epoch of training the images are rotated and flipped randomly, and measurements are simulated and contaminated with random Gaussian noise as described in Section 6.2. Note that the networks could also be trained at larger image sizes; however, since the SPIDER instrument only sparsely samples the Fourier domain (SPIDER acquires at a sparsity of $\frac{4440}{256 \times 256} \approx 7$ per cent), reconstructing at a larger image size would be challenging.

Since these methods are fully convolutional networks they can be adapted to reconstruct at larger image sizes, if the measurement operator is adapted accordingly. Note however that the SPIDER instrument only sparsely samples the Fourier domain (SPIDER acquires at a sparsity of $\frac{4440}{256 \times 256} \approx 7$ per cent), because of this sparsity reconstructing at a larger image size will not necessarily yield reconstructions at higher resolution.

The Common Objects in COntext (COCO, Lin et al. 2014) data set is a large and diverse set of natural images. We selected a subset of 3000 images split into 2000 training and 1000 test images. This data set provides a large and diverse data set of natural images for training the networks. Instruments similar to SPIDER could in future be considered as standard imaging devices for natural images. Furthermore, large data sets of readily available natural images such as this can be used for transfer learning, as considered subsequently. This data set is initially used to demonstrate the general reconstruction performance of the reconstruction methods when trained on a large set of diverse images. All the images in this data set have dimensions larger than our network input size, and during training, a random region of 256×256 is cropped out of the images at every epoch.

Next, we consider a small domain-specific data set of 450 simulated galaxy images, split into 300 training and 150 test images. The galaxy images are obtained from the IllustrisTNG simulations (Nelson et al. 2019b). The H-alpha column densities from the TNG50 simulation of the IllustrisTNG project (Nelson et al. 2019a; Pillepich et al. 2019) are binned to a 256×256 grid to obtain images of simulated galaxy structures. Pixels with no simulation data are inpainted using a primal-dual method in an unconstrained setting using the

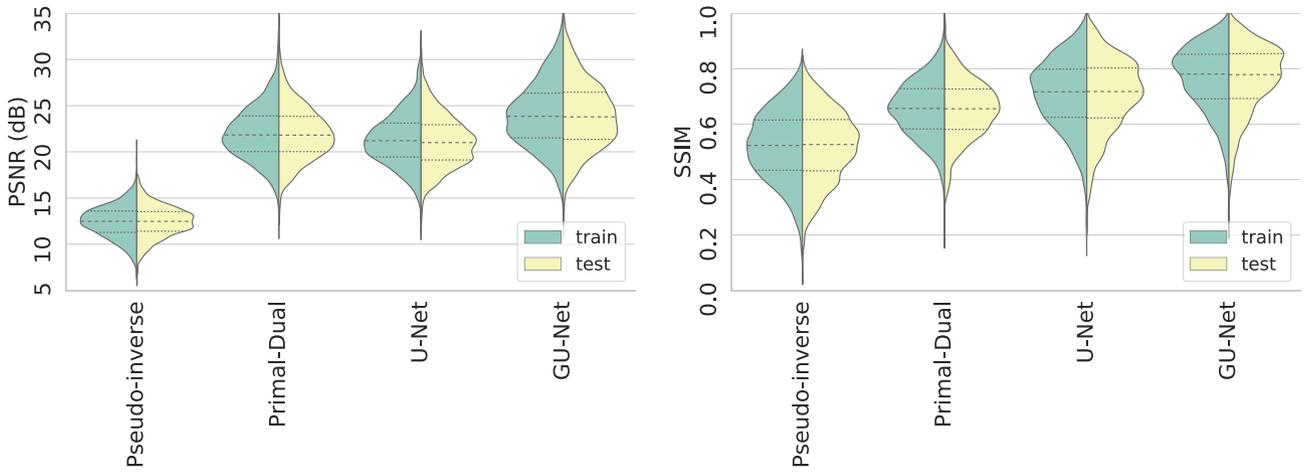


Figure 8. Distribution of reconstruction quality (PSNR and SSIM) for the different reconstruction methods on both the train and the test sets from the COCO data set of natural images for measurements with an ISNR of 30 dB. The reconstructions are made using the pseudo-inverse of the measurement operator, a primal-dual optimization scheme representing the state-of-the-art, our learned post-processing approach (U-Net), and our learned unrolled iterative approach (GU-Net). The dashed and dotted lines inside the distributions indicate the mean and quartiles of the distributions.

ℓ_1 -norm of a dictionary of wavelets containing the Dirac basis and the first eight Daubechies wavelets (Db1–Db8) as the regularization functional. We add a small amount of noise to the simulation data [input signal-to-noise ratio (ISNR) = 30 dB] and use Bayesian inference to estimate the regularization parameter for this inpainting problem from the simulation data as described in Pereyra, Bioucas-Dias & Figueiredo (2015). These steps are followed merely to process the IllustrisTNG particle simulations in order to provide a data set of ground-truth galaxy images suitable for our imaging experiments.

The final data set we consider is another small domain-specific data set of 450 satellite Earth observations taken from the *Deep Globe* satellite challenge (Demir et al. 2018), split into 300 training and 150 test images.

6.2 Simulating measurements

For a SPIDER instrument with parameters as defined in Table 1, we simulate measurements using the NUFFT operator (Section 4.1) with upsampling factor 2 per dimension, resulting in upsampling to 512×512 images, and a 6×6 KB kernel, which in this configuration results in $y \in \mathbb{C}^{440}$ measurements.

We contaminate the measurements with complex Gaussian noise, $\Re(\mathbf{n}), \Im(\mathbf{n}) \in \mathcal{N}(0, \sigma/\sqrt{2})$ with 0 mean and standard deviation of the real and imaginary components defined by an ISNR of 30 dB:

$$\sigma = \frac{\|\Phi \mathbf{x}\|_{\ell_2}}{\sqrt{M}} \cdot 10^{-\frac{\text{ISNR}}{20}}. \quad (36)$$

6.3 Training and transfer learning

We train the networks on pairs (y_i, x_i) of synthetic noisy SPIDER Fourier measurements and images with the mean squared error (MSE) cost

$$\mathcal{C}(y_i, x_i) = \frac{\sum_{i=1}^N \|\Phi_{\theta}^{\dagger} y_i - x_i\|^2}{N}, \quad (37)$$

where $\Phi_{\theta}^{\dagger} : \mathbb{C}^M \rightarrow \mathbb{R}^N$ is the learned pseudo-inverse operator (i.e. the networks defined in Section 5), such that $\Phi_{\theta}^{\dagger} y$ gives the learned reconstructions of the noisy measurements y .

The networks are trained using the ADAM optimizer (Kingma & Ba 2014) with a learning rate of 0.001 and a batch size of 5. The

training data for each epoch are precomputed to save time. Networks are trained for 200 epochs.

6.4 Computation time

The average evaluation time required to compute reconstructed images for the different methods is shown in Table 3 for images of the COCO data set. The evaluation times include the initial application of the NUFFT pseudo-inverse. Training time is also included and does not include the time of pre-computing the augmented training data.

As expected, the U-Net model is highly computationally efficient due to the small number of operator evaluations. The GU-Net model is moderately slower than the U-Net model, but not significantly so, again as expected. We note that learned methods are trained and evaluated on the GPU while the primal-dual method is run on a CPU, precluding direct comparison of the reconstruction time. Nevertheless, the evaluation times in Table 3 are roughly proportional to the number of fine scale operator evaluations, which corroborates that these evaluations dominate the computational cost, and so a GPU implementation of the primal-dual algorithm would remain $\sim 600 \times$ slower than our U-Net model and $\sim 55 \times$ slower than our GU-Net model. The computational times required by our learned models to recover images, in the order of 10s of milliseconds, is sufficiently low that our proposed methods can indeed open up real-time imaging for the SPIDER instrument.

6.5 Reconstruction quality

We first compare the pseudo-inverse, primal-dual, and two learned reconstruction methods on the COCO data set. The distribution of the quantitative metrics (PSNR and SSIM) over the training and test sets are depicted in Fig. 8. The reconstructions for a subset of training and test images from COCO data set are illustrated in Fig. 9.

Our U-Net model performs similarly to the primal-dual algorithm, which represents the state-of-the-art variational regularization approach, with marginally lower PSNR but marginally greater SSIM. However, recall that imaging with the U-Net model is orders of magnitude faster than the primal-dual algorithm. Our GU-Net model outperforms both the U-Net and primal-dual approaches in

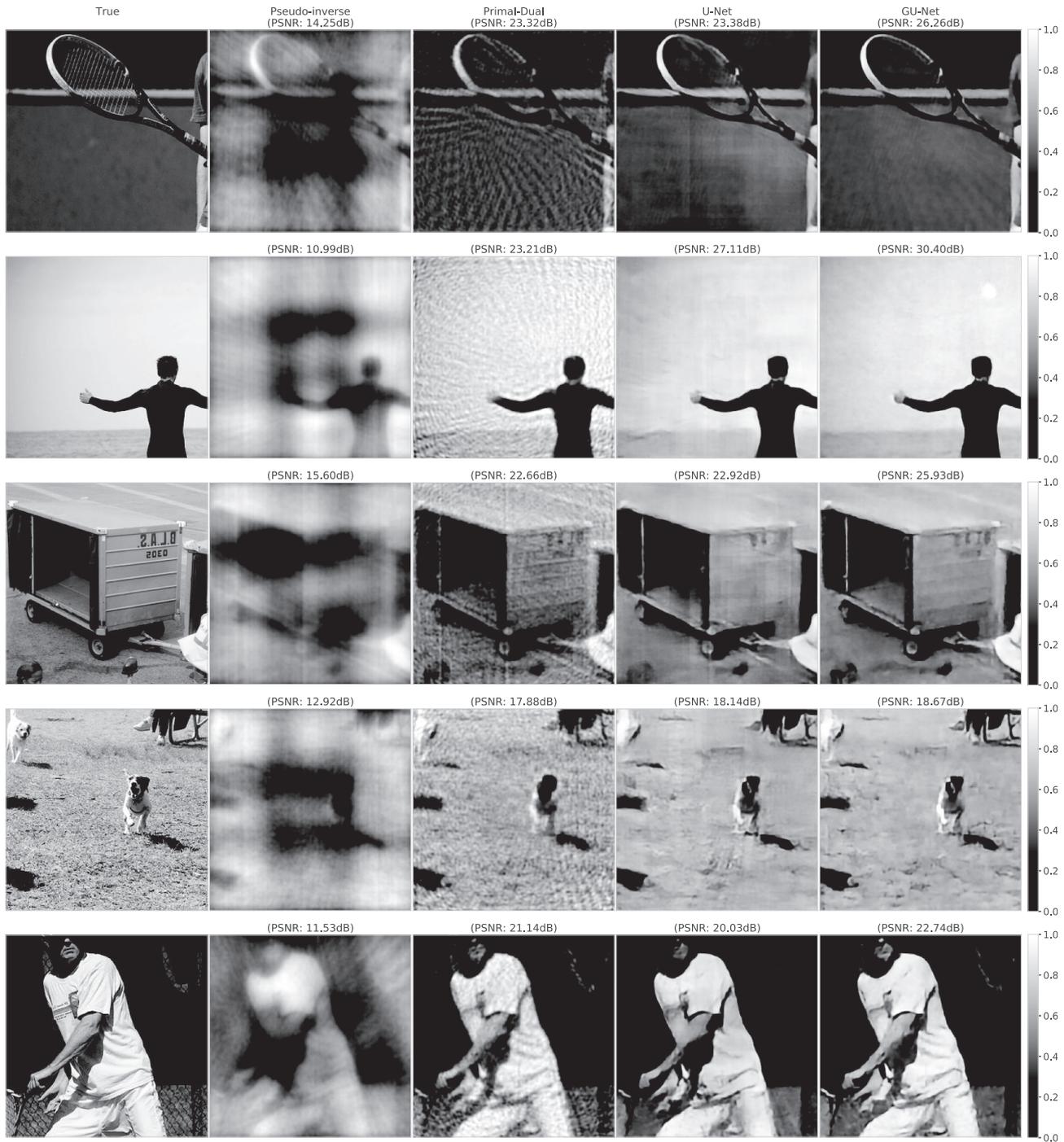


Figure 9. Example reconstructed images for the pseudo-inverse, the primal-dual algorithm, our learned post-processing approach (U-Net), and our learned unrolled iterative approach (GU-Net) on five images from the COCO test set. The first column shows the true image followed by the four different reconstructions from noisy measurements with an ISNR of 30 dB.

both PSNR and SSIM, as expected since it makes greater use of knowledge of the measurement operator, while introducing only a moderate increase in computation time compared with the U-Net approach. All methods outperform the pseudo-inverse, which is to be expected.

We note the near mirror symmetry of the plots between training and test set, which indicates only small difference in distribution of the metrics between the train and test sets (this is also the case for the primal-dual method which is data set agnostic since it does not

include any training). This suggests that the trained models generalize well to unseen data in the same domain.

From the examples in Fig. 9, it is apparent that the pseudo-inverse yields noisy reconstructions with aliasing artefacts, the primal-dual algorithm provides an improvement but still yields reconstructions with considerable artefacts, the U-Net model generates good reconstructions that are sometimes oversmoothed, losing details, while the GU-Net model does best at recovering details and suppressing artefacts, resulting in the overall best reconstructions.

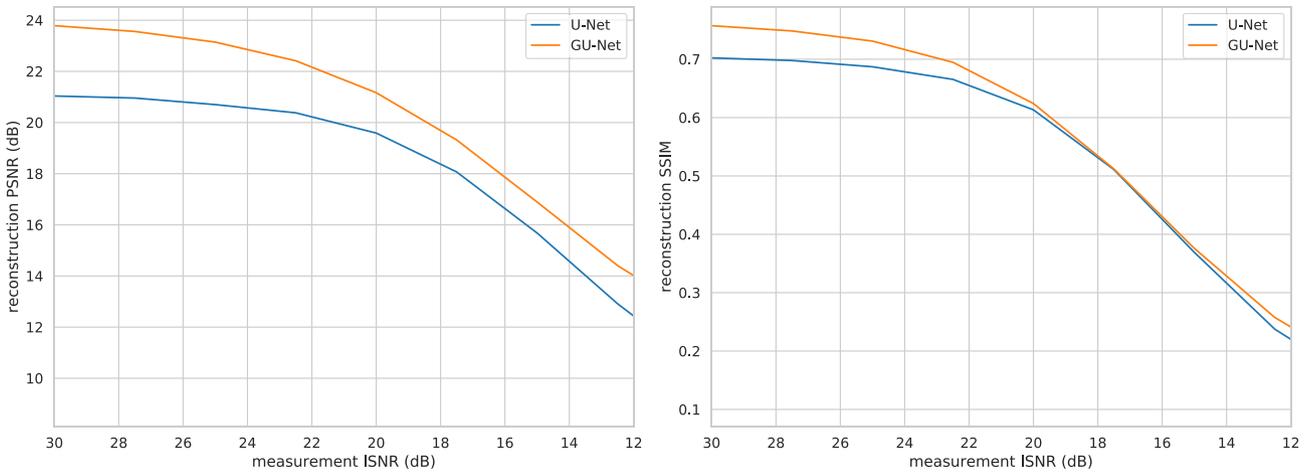


Figure 10. Robustness of learned models (trained for ISNR = 30 dB) to increased input noise level: mean PSNR (left panel) and SSIM (right panel) averaged over a set of 100 images from the COCO test set as a function of ISNR in the noise model (as described in Section 6.2). Our learned models exhibit sufficient robustness to variations in noise, such that a moderate underestimation of the noise level will not result in a significant loss of reconstruction quality.

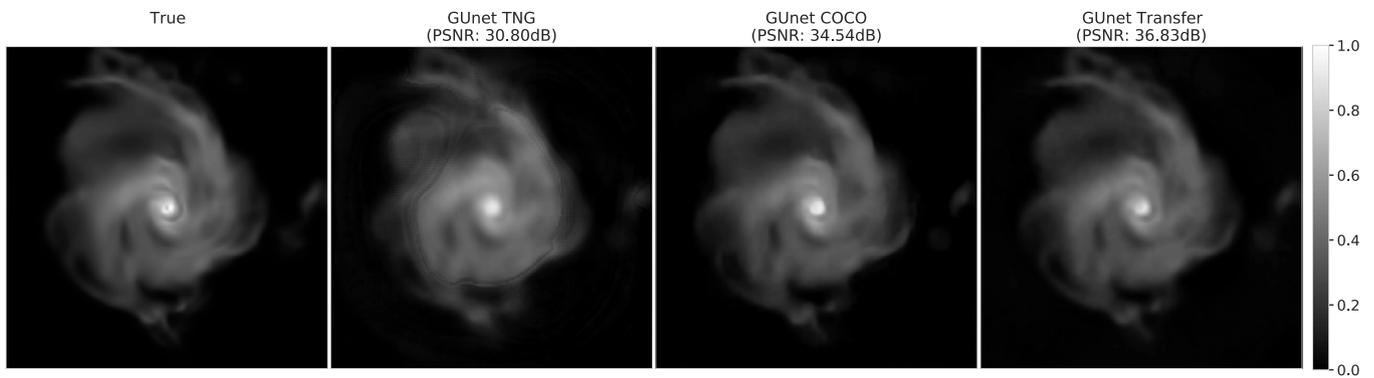


Figure 11. Comparison of the three different approaches for training a model for smaller, domain-specific data sets (as described in Section 6.3). From left to right: the original image, a reconstruction from a network trained only on the small galaxy data set, a reconstruction using a network trained only on natural images from the COCO data set, and reconstruction made with a network trained first on natural images and then fine-tuned using transfer learning to the galaxy data set. The transfer learning approach recovers images with the highest reconstruction quality and the least amount of artefacts.

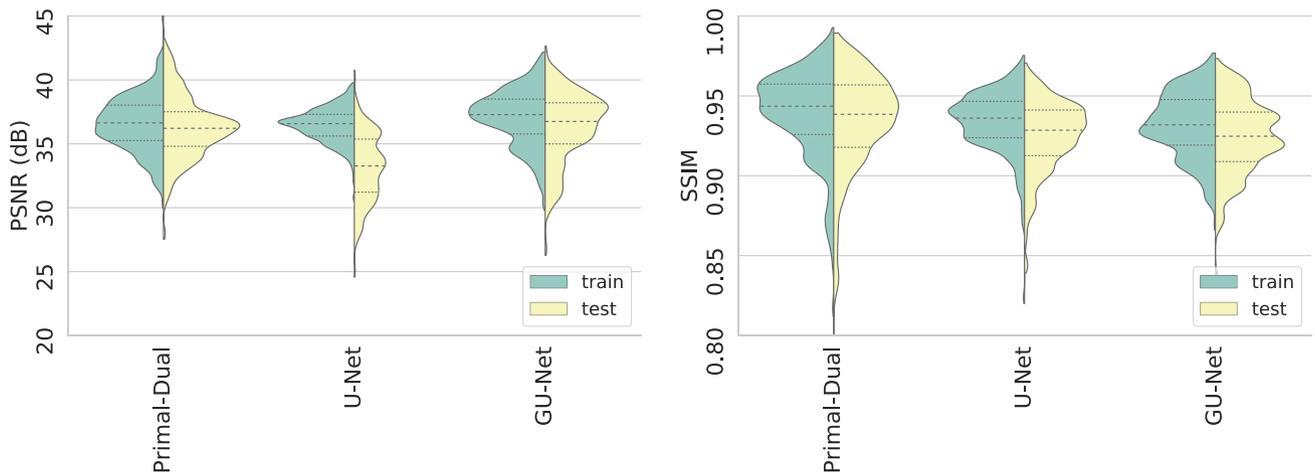


Figure 12. Distribution of quantitative imaging metrics (PSNR and SSIM) over the train and test sets of the galaxy data set. The dashed and dotted lines indicate the mean and quartiles of the distributions. The learned approaches are trained using images of the COCO data set first, and are then adapted to the galaxy data set by transfer learning.

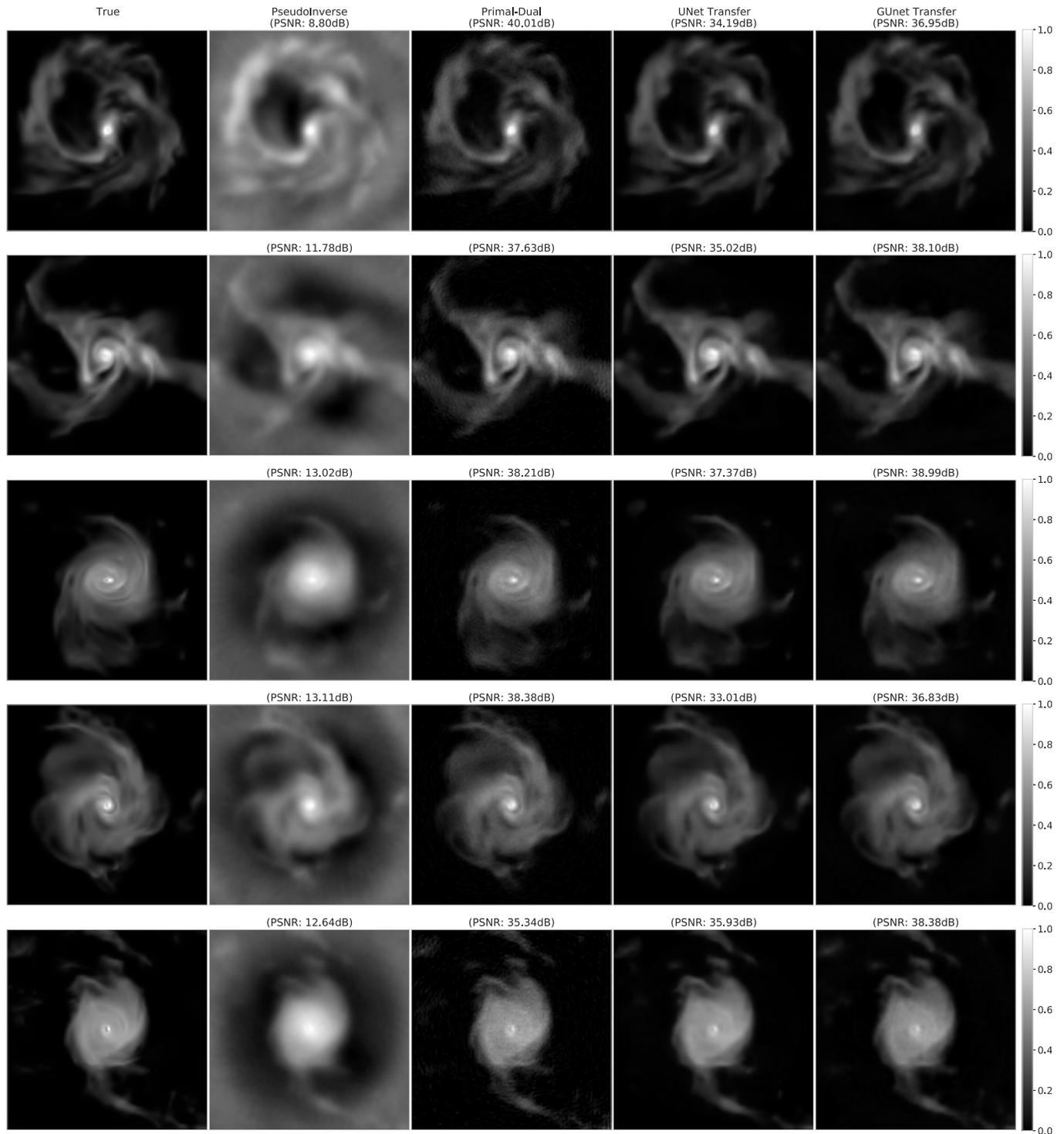


Figure 13. Reconstruction performance of the pseudo-inverse, a state-of-the-art reconstruction (primal-dual), our learned post-processing approach (U-Net), and our learned unrolled iterative approach (GU-Net) on five images from the galaxy images test set. The first column shows the true image followed by the four different reconstructions from noisy measurements with an ISNR of 30 dB. The learned methods are trained on images from the COCO data set first and are then adapted to the galaxy data set by transfer learning.

6.6 Robustness to noise

Next, we evaluate the robustness of our trained models with respect to increased levels of additive Gaussian noise in the input. All models are trained with $\text{ISNR} = 30$ dB; however, we vary the ISNR from 30

to 12.5 dB in test images. The averages of PSNR and SSIM over a set of COCO images are shown in Fig. 10.

The plots in Fig. 10 both show a relative plateau followed by a transition to linear decay at around $\text{ISNR} = 20$ dB indicating

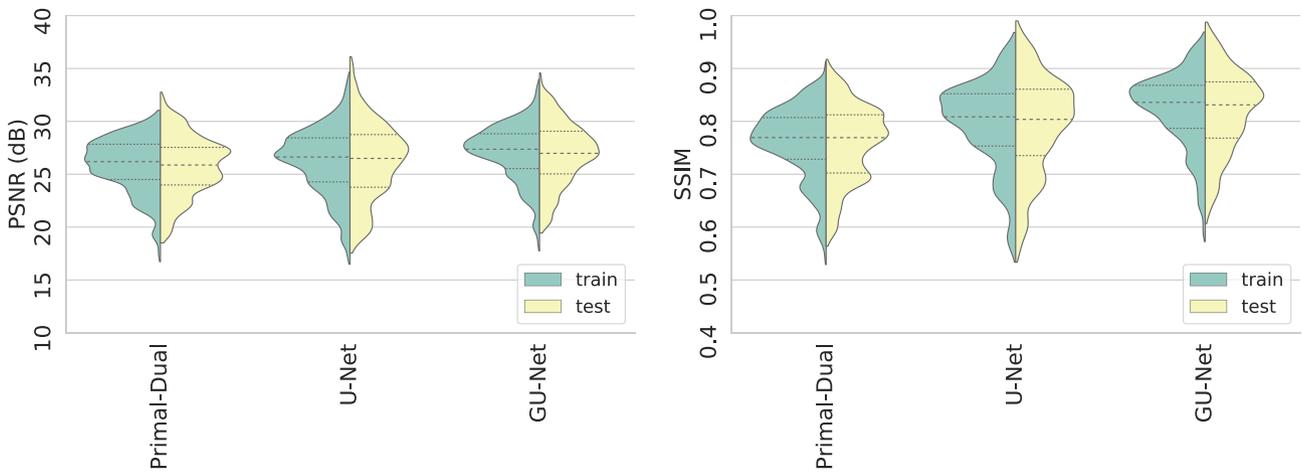


Figure 14. Distribution of quantitative imaging metrics (PSNR and SSIM) over the train and test sets of the *Deep Globe* satellite image data set. The dashed and dotted lines indicate the mean and quartiles of the distributions. The learned approaches are trained using images of the COCO data set first, and are then adapted to the satellite image data set by transfer learning.

robustness to a moderate increase in noise level, with a linear deprecation for higher noise levels. The PSNR GU-Net curve remains above the U-Net curve as noise increases, while the SSIM curves are essentially on top of one another for higher noise levels. This suggests that the advantage of the GU-Net in terms of SSIM is marginal for higher noise levels, while we maintain some PSNR advantage even for higher noise levels. Overall, Fig. 10 suggest sufficient robustness of the trained models with respect to varying input noise levels, such that a moderate underestimation of the noise level will not result in a significant loss of reconstruction quality.

6.7 Generalization to other data sets

Besides assessing the performance of the models on natural images, where we have a reasonably large number of training instances, we also consider other data sets where data availability is limited, as is the case for the galaxy and satellite data sets described in Section 6.1.

We consider three approaches for training networks for the two smaller, domain-specific data sets: (i) training on the small data set from scratch; (ii) reconstruction of test images using the network trained on the COCO data set without retraining; and (iii) transfer learning via initializing with the COCO pre-trained networks and fine-tuning on images from the small data set for 100 epochs.

As a first test, we compare these three approaches for the smaller, domain-specific galaxy data set. Fig. 11 shows GU-Net reconstructions for a test galaxy image in these three scenarios. The network trained on just the galaxy images produces artefacts in its reconstruction, due to the small volume of training data, while the model trained solely on images from the COCO data set appears blurred, as the galaxy images do not have sharp edges. The transfer learning approach, which leverages both the COCO data set and the smaller data set of galaxy images, recovers the image with the highest PSNR. Therefore, we restrict further comparisons to the transfer learning scenario.

The distribution of the quantitative metrics over the training and test sets of the galaxy data set for the transfer learning scenario can be seen in Fig. 12. Example reconstructions can be seen in Fig. 13. From Fig. 12, it is evident that the U-Net has been overfitted to the training data as its performance metrics on the test data are substantially worse. Nevertheless, the example reconstructions of both learned models in Fig. 13 are visually very similar. We

suppose this is due to the nature of the galaxy data set for which neither PSNR nor SSIM is a particularly well suited metric. Indeed, visually the galaxy images reconstructed by the learned models are arguably slightly more appealing than those recovered by the primal-dual algorithm, while the quantitative metrics are similar or arguably marginally favour the primal-dual case. The close match between the train and test distributions for GU-Net indicates that incorporating the physical model into the learned reconstruction yields models with superior generalization performance to unseen data.

The same experiments are performed for the satellite image data set, with distributions of metrics shown in Fig. 14 and example reconstructions shown in Fig. 15. Both our learned approaches outperform the primal-dual approach, in terms of metrics and visual inspection (the primal-dual algorithm yields more grainy, noisy reconstructions). Moreover, the GU-Net models exhibits a slight improvement over the U-Net model, particularly for images with clear edges.

In summary, by adopting a transfer learning approach our learned imaging methods achieve similar or superior reconstruction quality to a state-of-the-art variational regularization approach even for data sets for which limited training data is available, in a fraction of the computational time.

7. CONCLUSIONS

We have proposed two new learned approaches to reconstruct images for the SPIDER instrument, which compared with the classical state-of-the-art proximal optimization algorithms improve reconstruction quality, while dramatically reducing the computational time required to recover images.

Our first learned method adopts a learned post-processing approach to clearly separate the physical measurement model from the learned imaging via a U-Net model. A consequence of this configuration is that the pseudo-inverse of the physical measurement model need only be applied once, resulting in orders of magnitude reduction in imaging time compared with the benchmark traditional optimization algorithm. Reconstruction quality is similar to the benchmark traditional algorithm.

Our second learned method, the GU-Net, uses a similar architecture but enhanced with multiscale evaluations of the gradient of the data fidelity term, interweaving the measurement model into

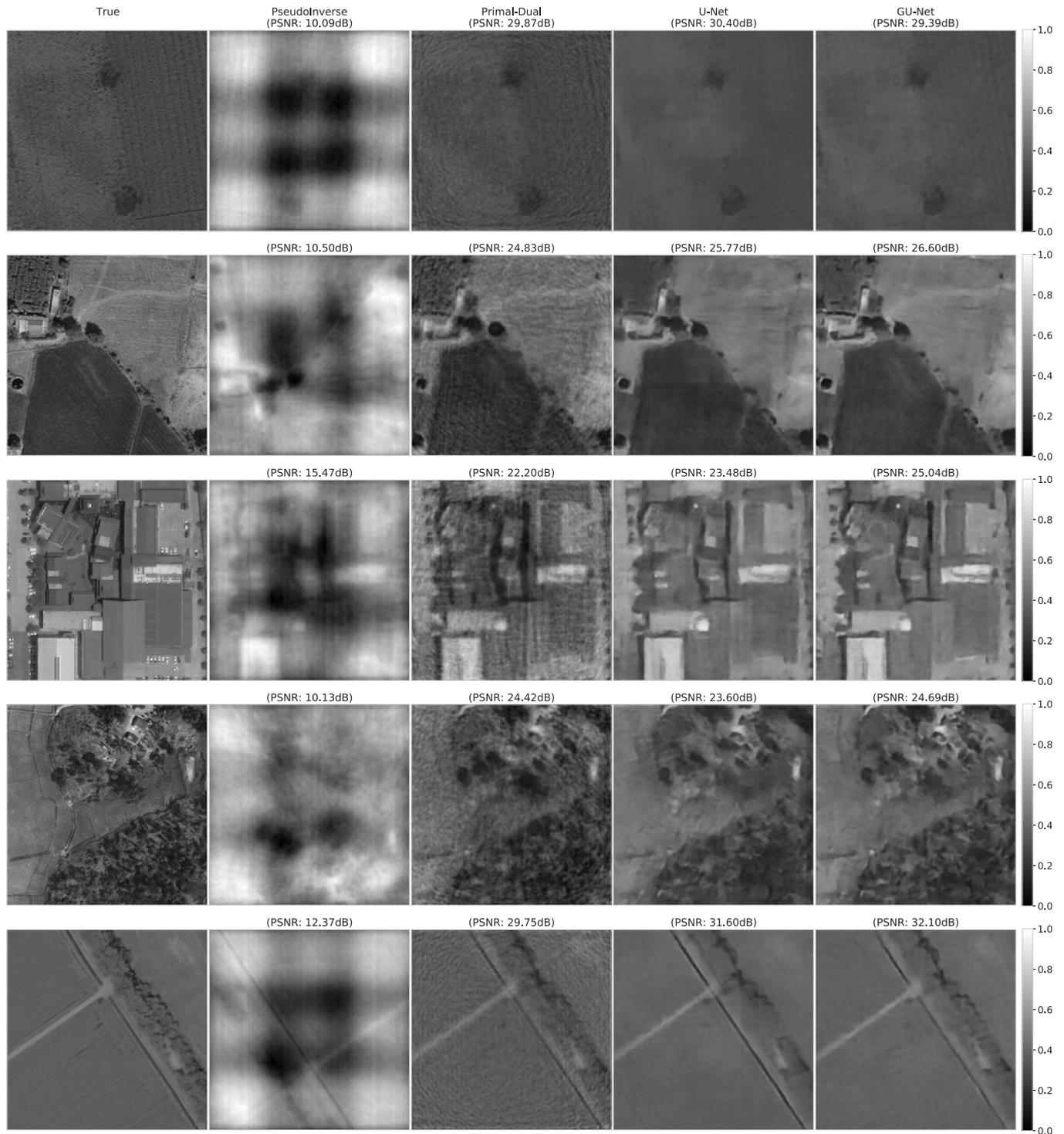


Figure 15. Reconstruction performance of the pseudo-inverse, a state-of-the-art reconstruction (primal-dual), our learned post-processing approach (U-Net), and our learned unrolled iterative approach (GU-Net) on five images from the satellite images test set. The first column shows the true image followed by the four different reconstructions from noisy measurements with an ISNR of 30 dB. The learned methods are trained on images from the COCO data set first and are then adapted to the satellite image data set by transfer learning.

the U-Net architecture. While this results in a moderate increase in the time required to recover images compared with the U-Net model, reconstruction quality is improved by making greater use of knowledge of the physical measurement model.

Overall, our learned methods achieve similar or superior reconstruction quality compared with traditional approaches, while realizing a dramatic reduction in the time required to recover images, to the extent that real-time imaging with the SPIDER

instrument becomes possible for the first time, opening up many new use-cases.

While our learned methods are trained using data with a specific noise level, their performance on measurements with increased noise levels remains similar for a moderate increase in noise level. Furthermore, in scenarios where only a limited volume of training data is available, as common in domain-specific problems, the performance of the models can be increased by transfer learning from a domain with sufficient training data. While for actual observations one should try to construct as representative training as possible, our learned methods show that it is possible to achieve similar or superior reconstruction quality to the traditional primal-dual approach in settings where limited train data are available, and at a fraction of the computational cost.

In addition, we have presented two approaches to modelling the measurement process of the SPIDER instrument. One method is based on the NUFFT, which is widely used in radio interferometry and is applicable to any arbitrary sampling distribution. We also present a new modelling approach based on the Radon transform, an NU-Radon method, that applies specifically when the sampling distribution has radial spokes of measurements, as is the case for SPIDER, and show similarities to works in the medical imaging domain. This latter approach is computationally more efficient when the number of measurements is large and the calculation of the measurement operator is dominated by the (de)gridding of measurements or when the number of pixels is large for a modest number of spokes. Both modelling methods and all imaging techniques have been implemented in Python and are available for use in the LEIA⁸ code.

While the methods presented in this work can be used with any arbitrary sampling distribution (when using the NUFFT), the sampling distribution we have considered is fixed. For radio interferometric imaging the sampling distribution typically varies for each observation as it depends on the location of the object in the sky as well as the length of time of the observation. Looking to the future, in order to extend the learned methods presented in this paper to radio interferometry, the networks need to be trained in such a way as to support varying sampling patterns. This is the focus of ongoing research, for which preliminary results are encouraging. Variants of the methods presented in this paper are therefore likely to be of use not only for SPIDER imaging but for radio interferometric imaging more generally.

ACKNOWLEDGEMENTS

We thank Luke Pratley for general discussions and for his work on the OptimusPrimal code, and Matthew Price for providing the galaxy images created from the IllustrisTNG data and for general discussions. MM is supported by the UCL Centre for Doctoral Training in Data Intensive Science (STFC Training grant ST/P006736/1). This work is also partially supported by EPSRC (grant number EP/W007673/1). The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work. This work also used computing equipment funded by the Research Capital Investment Fund (RCIF) provided by UKRI, and partially funded by the UCL Cosmoparticle Initiative.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Adler J., Öktem O., 2017, *Inverse Probl.*, 33, 124007
 Adler J., Öktem O., 2018, *IEEE Trans. Med. Imaging*, 37, 1322
 Allam T., Jr, 2016, MSc thesis, Univ. Coll. London
 Arridge S., Maass P., Öktem O., Schönlieb C.-B., 2019, *Acta Numer.*, 28, 1
 Badham K. et al., 2017, Proc. 2017 Conf. Lasers Electro-Opt. Pac. Rim (CLEO-PR), Photonic Integrated Circuit-based Imaging System for SPIDER. IEEE, Singapore, p. 1
 Bobin J., Starck J.-L., Fadili J., Moudden Y., Donoho D., 2007, *IEEE Trans. Image Process.*, 16, 2675
 Boink Y. E., Manohar S., Brune C., 2020, *IEEE Trans. Med. Imaging*, 39, 129
 Boyd S. P., Vandenberghe L., 2004, *Convex Optimization*. Cambridge Univ. Press, Cambridge
 Cai X., Pereyra M., McEwen J. D., 2018a, *MNRAS*, 480, 4170
 Cai X., Pereyra M., McEwen J. D., 2018b, *MNRAS*, 480, 4154
 Cai X., Pratley L., McEwen J. D., 2019, *MNRAS*, 485, 4559
 Cai X., McEwen J. D., Pereyra M., 2022, *Stat Comput*, 32, 87
 Candès E., Demanet L., Donoho D., Ying L., 2006, *Multiscale Model. Simul.*, 5, 861
 Carrillo R. E., McEwen J. D., Wiaux Y., 2012, *MNRAS*, 426, 1223
 Carrillo R. E., McEwen J. D., Van De Ville D., Thiran J.-P., Wiaux Y., 2013, *IEEE Signal Process. Lett.*, 20, 591
 Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *MNRAS*, 439, 3591
 Chamolle A., Pock T., 2011, *J. Math. Imaging Vis.*, 40, 120
 Chen H., Zhang Y., Kalra M. K., Lin F., Chen Y., Liao P., Zhou J., Wang G., 2017, *IEEE Tans. Med. Imaging*, 36, 2524
 Chu Q., Shen Y., Yuan M., Gong M., 2017, *Opt. Commun.*, 405, 288
 Combettes P. L., Pesquet J.-C., 2011, in Bauschke H. H., Burachik R. S., Combettes P. L., Elser V., Luke D. R., Wolkowicz H., eds, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Vol. 49. Springer, New York, p. 185
 Connor L., Bouman K. L., Ravi V., Hallinan G., 2022, *MNRAS*, 514, 2614
 Daubechies I., 1992, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, USA
 Demir I. et al., 2018, Proc. 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. IEEE, Salt Lake City, UT, p. 172
 Dokmanić I., Bruna J., Mallat S., de Hoop M., 2016, preprint (arXiv:1609.05502)
 Dong C., Loy C. C., He K., Tang X., 2016, *IEEE Trans. Pattern Anal. Mach. Intell.*, 38, 295
 Duijndam A., Schonewille M., 1997, SEG Technical Program Expanded Abstracts 1997, Society of Exploration Geophysicists, USA, p. 1135
 Duncan A. et al., 2015, Proc. Adv. Maui Opt. Space Surveillance Technol. Conf., SPIDER: Next Generation Chip Scale Imaging Sensor. The Maui Economic Development Board, Maui, Hawaii, p. 27
 Fessler J., Sutton B., 2003, *IEEE Trans. Signal Process.*, 51, 560
 Gheller C., Vazza F., 2021, *MNRAS*, 509, 990
 Gregor K., LeCun Y., 2010, Proc. 27th Int. Conf. Mach. Learn. (ICML'10), Learning Fast Approximations of Sparse Coding. Omnipress, Madison, WI, p. 399
 Gribonval R., Nielsen M., 2003, *IEEE Trans. Inform. Theor.*, 49, 3320
 Hauptmann A., Adler J., Arridge S., Öktem O., 2020, *IEEE Trans. Comput. Imaging*, 6, 843
 Hu H., Liu C., Zhang Y., Feng Q., Liu S., 2021, *Opt. Express*, 29, 24300
 Jackson J., Meyer C., Nishimura D., Macovski A., 1991, *IEEE Trans. Med. Imaging*, 10, 473
 Jin K. H., McCann M. T., Froustey E., Unser M., 2017, *IEEE Trans. Image Process.*, 26, 4509

⁸<https://github.com/astro-informatics/LeIA>

- Kendrick R. L. et al., 2013, Proc. Adv. Maui Opt. Space Surveillance Technol. Conf., Flat Panel Space Based Space Surveillance Sensor. The Maui Economic Development Board, Maui, Hawaii, p. E45
- Kingma D. P., Ba J., 2014, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Kobler E., Effland A., Kunisch K., Pock T., 2020, Proc. 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Total Deep Variation for Linear Inverse Problems. IEEE, Seattle, WA, p. 7546
- Li H., Schwab J., Antholzer S., Haltmeier M., 2020, *Inverse Probl.*, 36, 065005
- Lin T.-Y. et al., 2014, preprint ([arXiv:1405.0312](https://arxiv.org/abs/1405.0312))
- Liu G., Wen D.-S., Song Z.-X., 2018, *MNRAS*, 478, 2065
- Liu G., Wen D., Song Z., Zhang W., Li Z., Wei X., Jiang T., 2019, in Zhao Y., Barnes N., Chen B., Westermann R., Kong X., Lin C., eds, Image and Graphics, Vol. 11902. Springer International Publishing, Cham, p. 359
- Lunz S., Öktem O., Schönlieb C.-B., 2018, preprint ([arXiv:1805.11572](https://arxiv.org/abs/1805.11572))
- Lv G., Chen Y., Feng H., Xu Z., Li Q., 2020, Proc. 6th China High Resolution Earth Obs. Conf. (CHREOC 2019), Lecture Notes in Electrical Engineering. Springer, Singapore, p. 241
- Mukherjee S., Dittmer S., Shumaylov Z., Lunz S., Öktem O., Schönlieb C.-B., 2020, preprint ([arXiv:2008.02839](https://arxiv.org/abs/2008.02839))
- Mukherjee S., Carioni M., Öktem O., Schönlieb C.-B., 2021a, preprint ([arXiv:2106.03538](https://arxiv.org/abs/2106.03538))
- Mukherjee S., Öktem O., Schönlieb C.-B., 2021b, in Elmoatay A., Fadili J., Quéau Y., Rabin J., Simon L., eds, Scale Space and Variational Methods in Computer Vision, Vol. 12679. Springer International Publishing, Cham, p. 540
- Nelson D. et al., 2019a, *MNRAS*, 490, 3234
- Nelson D. et al., 2019b, *Comput. Astrophys. Cosmol.*, 6, 2
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, *MNRAS*, 462, 4314
- Pan B., Betcke M. M., 2023, *SIAM J. Imag. Sci.*, 16, 770
- Pereyra M., Bioucas-Dias J. M., Figueiredo M. A. T., 2015, Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO), Maximum-a-posteriori Estimation with Unknown Regularisation Parameters. IEEE, Nice, p. 230
- Pillepich A. et al., 2019, *MNRAS*, 490, 3196
- Pratley L., McEwen J., 2021, Proc. 9th Int. Conf. Photonics Opt. Laser Technol. SciTePress, Setubal, Portugal, p. 104
- Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, *MNRAS*, 473, 1038
- Pratley L., McEwen J. D., d'Avezac M., Cai X., Perez-Suarez D., Christidi I., Guichard R., 2019, preprint ([arXiv:1903.04502](https://arxiv.org/abs/1903.04502))
- Putzky P., Welling M., 2017, preprint ([arXiv:1706.04008](https://arxiv.org/abs/1706.04008))
- Ronneberger O., Fischer P., Brox T., 2015, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Vol. 9351. Springer International Publishing, Cham, p. 234
- Ryu E. K., Liu J., Wang S., Chen X., Wang Z., Yin W., 2019, preprint ([arXiv:1905.05406](https://arxiv.org/abs/1905.05406))
- Sault R. J., Conway J. E., 1999, ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II. Astron. Soc. Pac., San Francisco, p. 419
- Starck J.-L., Murtagh F., Fadili J. M., 2010, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity, 1st edn. Cambridge Univ. Press, Cambridge
- Su T., Scott R. P., Ogden C., Thurman S. T., Kendrick R. L., Duncan A., Yu R., Yoo S. J. B., 2017, *Opt. Express*, 25, 12653
- Su T. et al., 2018, *Opt. Express*, 26, 12801
- Taylor G. B., Carilli C. L., Perley R. A., Observatory N. R. A., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II. Astron. Soc. Pac., San Francisco
- Terris M., Abdulaziz A., Dabbech A., Jiang M., Repetti A., Pesquet J.-C., Wiaux Y., 2019, Proc. SPARS Workshop, Deep Post-Processing for Sparse Image Deconvolution. Toulouse, France, p. 3
- Terris M., Dabbech A., Tang C., Wiaux Y., 2022, preprint([arXiv:2202.12959](https://arxiv.org/abs/2202.12959))
- Trent D. B., 2020, MSc thesis, Univ. Coll. London, London
- Venkatakrisnan S. V., Bouman C. A., Wohlberg B., 2013, Proc. 2013 IEEE Glob. Conf. Signal Inf. Process., Plug-and-Play Priors for Model Based Reconstruction. IEEE, USA, p. 945
- Wang Z., Bovik A., Sheikh H., Simoncelli E., 2004, *IEEE Trans. Image Process.*, 13, 600
- Xu Q., Yu H., Mou X., Zhang L., Hsieh J., Wang G., 2012, *IEEE Trans. Med. Imaging*, 31, 1682
- Yang Y., Sun J., Li H., Xu Z., 2017, preprint ([arXiv:1705.06869](https://arxiv.org/abs/1705.06869))
- Yi X., Babyn P., 2018, *J. Digit. Imaging*, 31, 655
- Zernike F., 1938, *Physica*, 5, 785
- Zheng A., Gao H., Zhang L., Xing Y., 2020, *Phys. Med. Biol.*, 65, 245030

SUPPORTING INFORMATION

Supplementary data are available at *RASTAI* online.

learned_imaging_spider_rasti_v3.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.