Multi-Resolution Data Fusion for Super Resolution Imaging

Emma J. Reid, *Student Member, IEEE*, Lawrence F. Drummy, Charles A. Bouman, *Fellow, IEEE*, and Gregery T. Buzzard, *Senior Member, IEEE*

***Abstract*—Applications in materials and biological imaging are limited by the ability to collect high-resolution data over large areas in practical amounts of time. One solution to this problem is to collect low-resolution data and interpolate to produce a high-resolution image. However, most existing super-resolution algorithms are designed for natural images, often require aligned pairing of high and low-resolution training data, and may not directly incorporate a model of the imaging sensor.**

In this paper, we present a Multi-resolution Data Fusion (MDF) algorithm for accurate interpolation of low-resolution electron microscope data at multiple resolutions up to 8x. Our approach uses small quantities of unpaired high-resolution data to train a neural network prior model denoiser and then uses the Multi-Agent Consensus Equilibrium (MACE) problem formulation to balance this denoiser with a forward model agent that promotes fidelity to measured data.

A key theoretical novelty is the analysis of mismatched back-projectors, which modify typical forward model updates for computational efficiency or improved image quality. We use MACE to prove that using a mismatched back-projector is equivalent to using a standard back-projector and an appropriately modified prior model. We present electron microscopy results at 4x and 8x interpolation factors that exhibit reduced artifacts relative to existing methods while maintaining fidelity to acquired data and accurately resolving sub-pixel-scale features.

***Index Terms*—Data fusion, MACE, Multi-Agent consensus equilibrium, MDF, Multi-Resolution data fusion, super resolution, Plug-and-Play.**

I. Introduction

MANY important material science problems require the collection of high resolution (HR) data over large fields of view (FoV). For example, high resolution images are needed to extract detailed features, such as the 4 nm curli fibers structures in *E. coli*, which are fundamental in the formation of bacterial biofilms, or the 10-20 nm structures in gold nanorods materials, which are of interest due to their near-infrared light tunability and biological inertness [1]. Also, a large FoV is typically required to collect the representative volumes (RV) of materials that are needed to determine macroscopic properties such as material toughness and fracture strength [2]. However, imaging multiple, large FoVs at high resolution is difficult under realistic constraints. For example, raster scanning a 1 mm $\times$ 1 mm FoV at a resolution of 10 nm requires the acquisition of approximately 10 G-pixels of data, which requires roughly 17 hours under conditions described in [3]. One approach to overcoming this barrier is to acquire a large FoV at low resolution (LR) and interpolate it to obtain a higher resolution image of sufficient quality. Ideally, 4x interpolation in each direction leads to a 16x decrease in acquisition time, while 8x interpolation leads to a 64x decrease.

Traditional interpolation methods such as splines [4] do not offer sufficient quality, but recent advances using deep neural networks (DNNs) have produced a number of methods for high quality interpolation of natural images. For example, [5] and [6] use end-to-end DNNs trained on HR/LR paired images. This approach is improved in SRGAN [7] through adversarial training and perceptual loss. Another approach is EnhanceNet [8], which uses automated texture synthesis and perceptual loss but focuses on creating realistic textures rather than reproducing ground truth images. ESRGAN [9] introduces architectural improvements to SRGAN, and ESRGAN+ [10] adds further refinements. DPSR [11] uses a form of Plug-and-Play (PnP) as described later, but still requires a CNN super-resolver trained on HR/LR pairs. And finally, DPSRGAN [11] attempts to fuse these Plug-and-Play methods with the realistic textures generated by GANs.

An impediment to applying these DNN methods to material imaging problems is that they are typically trained using a large set of aligned HR/LR patch pairs. However, in practice it is difficult to acquire large quantities of accurately aligned HR/LR data [12]. More recently, zero-shot learning has been proposed as a method that does not require aligned HR/LR data for training. In zero-shot learning, the LR image is further downsampled to produce paired data to train a small, image-specific CNN that is used to upsample the original [13]. While this allows for quick and accurate reconstructions, it does not make use of high-resolution data in its training.

Manuscript received May 24, 2021; revised October 28, 2021, December 13, 2021, and December 21, 2021; accepted January 1, 2022. Date of publication January 5, 2022; date of current version January 27, 2022. This work was supported in part by UES Inc. under Prime Contract FA8650-15- D-5405 and in part by NSF under Grant CCF-1763896. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Se Young Chun. *(Corresponding author: Emma J. Reid.)*

Emma J. Reid and Gregery T. Buzzard are with the Department of Mathematics, Purdue University, West Lafayette, IN 47907 USA (e-mail: emma-justinereid@gmail.com; buzzard@purdue.edu).

Lawrence F. Drummy is with Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, West Lafayette, OH 45433 USA (e-mail: lawrence.drummy.1@afresearchlab.com).

Charles A. Bouman is with the Departments of Electrical and Computer Engineering, and Biomedical Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: bouman@purdue.edu).

Digital Object Identifier 10.1109/TCI.2022.3140551

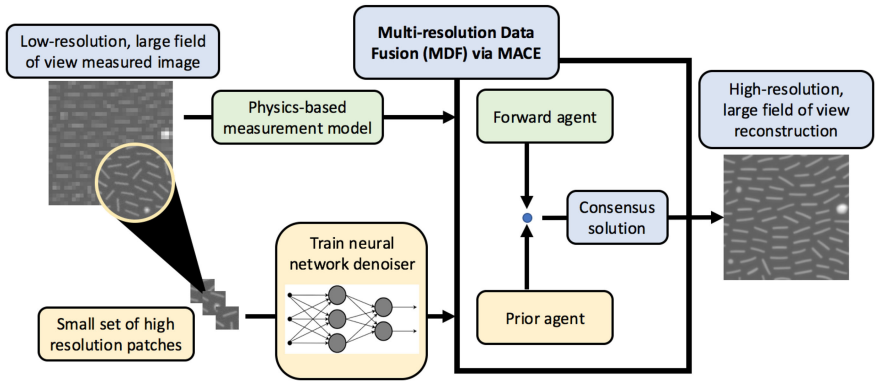

Fig. 1. Overview of the MDF pipeline. A small set of high-resolution data (unpaired with low-resolution data) is used to tune a CNN denoiser. This tuned denoiser is used in the MACE framework with a microscopy-based forward model to produce high-resolution images from low-resolution data.

In this paper, we update the MDF algorithm of [14] to use a prior model based on deep learning and give a theoretical justification for the use of mismatched backprojectors. Our method yields accurate 4x and 8x interpolation of large FoV low-resolution EM images using selected unpaired patches of HR data. Moreover, our method can perform interpolation by a range of factors without retraining of the prior model.

More specifically, the novel contributions of this work include:

- An update of the MDF algorithm from [14] to incorporate recent advances in deep learning and a description of this algorithm in the MACE problem formulation.
- The use of the MACE formulation to show that the use of a mismatched or relaxed adjoint projector (RAP) is provably equivalent to using the standard back projector with an appropriately modified prior. See Section III for more details on RAP.
- Experimental results indicating that interpolation factors of 4x to 8x are possible with realistic transmission and scanning electron microscopy data sets.

Our MDF-RAP approach can be applied to a number of imaging processing tasks and modalities, and its modularity allows for changes in forward models, resolution level, and regularization strength without retraining. Our analysis of the Relaxed Adjoint Projection interprets this particular change to the forward model to a corresponding change to the prior.

In Fig. 1, we provide a visual representation of the MDF approach. Using a LR base image, we train a denoiser on a sparse set of HR patches of the same (or similar) specimen. This allows the denoiser to learn the underlying manifold of the HR data while simultaneously being trained to remove additive white Gaussian noise (AWGN). This denoiser is then applied in the MACE framework to achieve a super-resolution reconstruction of the low-resolution base image. This approach does not require registered pairs of HR and LR data, allowing for flexible levels of super resolution and simple generalization to other problems. Additionally it allows for use of known forward models and has a single parameter that can be used to control the weight of data fidelity relative to strength of regularization.

Code for this project is available at https://www.github.com/emmajreid/MDF.

II. Related Work & Problem Formulation

Data fusion, or the combination of multiple sources of data, has been used successfully in domains such as medical imaging, remote sensing, and others for many years [15], [16]. In medical imaging, scientists combine scans from multiple sources such as PET, CT, and MRI to reconstruct images that better show internal body properties [17]. Recent work on data fusion in remote sensing [18] provides a framework for multi-modal and cross-modal deep learning methods for pixel-wise classification and spatial information modeling. In a separate direction, [19] applies super-resolution in the spectral domain to images that are spatially high-resolution but spectrally low-resolution; this is done by learning a joint sparse and low-rank dictionary of partially overlapping hyperspectral and multispectral images and their sparse representations. In [20] this is done with an unsupervised hyperspectral super-resolution network that blends spatial information from multispectral sensors and spectral information from hyperspectral sensors.

The PnP algorithm [21], [22] and MACE problem formulation [23] provide a framework in which multiple models can be combined with little modification, thus providing a natural approach to data fusion. A PnP approach was used in [24] to combine CT and MRI modalities across sensor, data, and image priors, but all at a single resolution. In [25], PnP was used with the Nonlocally Centralized Sparse Representation algorithm to combine sparse coding and dictionary learning to turn a denoiser into a super-resolver. However this method begins to break down in the presence of noise, which is prevalent in microscopy images.

In work closely related to the current paper, [14] uses PnP to develop MDF for microscopy. In that work, they used a fixed library of domain-specific HR patches as the basis for a non-local means denoiser and then coupled this with a data-fitting forward model to achieve super-resolution from LR image data over a large field of view. However, the resulting denoiser is computationally expensive relative to a neural network approach. Moreover, that paper did not provide theoretical foundation to guide modifications to the algorithm.

As noted above, we extend the work in [14] to use a neural network prior and describe our method and the use of RAP in

terms of the MACE formulation. In Section II-A to II-C, we examine the modeling of the microscope's point spread function, detail the forward model derivation from [14], motivate RAP, and describe the MACE framework developed in [23] and the relationship between MACE and PnP.

A. Microscopy Forward Model

Our goal is to interpolate a rasterized image $y \in \mathbb{R}^M$ to a HR version $x \in \mathbb{R}^N$. Super-resolution by a factor of L implies scaling by L in each direction, so that $N = L^2\, M$. For such a problem, the forward model is typically

$$y = \Psi x + \epsilon, \tag{1}$$

where $\Psi \in \mathbb{R}^{M \times N}$ represents the point spread function of the microscope and $\epsilon \sim N(0, \sigma_w^2)$ is an M dimensional vector of AWGN. For our application, ψ is closely modeled by averaging over $L \times L$ patches. The data-fitting cost function is then $\frac{1}{2}\|y - \Psi x\|^2$, which is embedded in a proximal map and balanced by the action of a prior model. For application purposes, we now need to approximate Ψ.

In bright-field transmission electron microscopy, a parallel beam of electrons illuminates a thin material sample, and the resulting transmitted beam is formed into an image using the microscope's objective lens. This is then further magnified using the microscope's projection lens system. Using this configuration, the transmission electron microscope can yield a magnification ranging from 1,000X to over 1,000,000X. The magnified beam is sampled at the image plane using a pixel array detector such as a direct electron detector or CCD camera. Since the electron-optical magnification can be controlled and the image is detected using a pixel array detector, a block-averaging forward model is a good approximation for this acquisition modality. In this forward model, a square region in the high-resolution image is averaged to produce a single pixel value in the low-resolution image.

For scanning electron microscopy, an electron beam is focused to a small diameter probe which is then raster scanned across the surface of the sample using beam deflectors. As the probe strikes the sample, secondary electrons are ejected from the sample and collected with an integrating detector, which sums the total number of electrons scattered at each point on the surface of the sample. The raster array dimensions can be controlled to give LR and HR data, so again a block-averaging forward model is a good approximation to the imaging system. The image acquisition process for transmission and scanning electron microscopy is depicted in Fig. 2.

B. PnP Formulation

Plug-and-play (PnP) [21], [22] is an algorithm for solving inverse problems that replaces the probabilistic encoding of Bayesian prior information with an algorithmic encoding; a candidate reconstruction is denoised to produce a more plausible reconstruction without using a cost function or likelihood. This prior update is combined iteratively with a forward update until a fixed point is reached. Using the description in Section II-A, we approximate the forward model Ψ by block-averaging every non-overlapping neighborhood of $L \times L$ pixels in the HR image to obtain the LR image [14]. For notational convenience, we let A represent summation over $L \times L$ blocks, in which case $\Psi \approx A/L^2$ in (1). Also, A^T is given by replicating each pixel into an $L \times L$ block, so $AA^T = L^2 I$. The negative log-likelihood is then

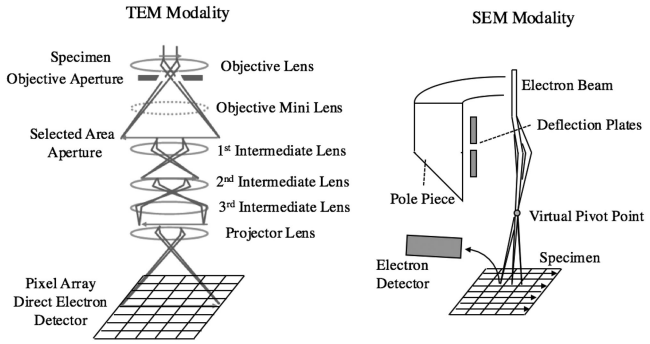

Fig. 2. Illustration of the transmission and scanning electron microscopy image acquisition process. We approximate both image acquisition processes by block-averaging a HR image as described in Section II-B. The scanning electron microscopy image shown is adapted from [26].

$$l(x) = \frac{1}{2\sigma_w^2}\left\| y - \frac{1}{L^2} Ax \right\|^2 + \frac{M}{2}\log(2\pi\sigma_w^2). \tag{2}$$

As part of the Plug-and-Play algorithm, Sreehari *et al.* used the proximal map of $l(x)$, given by

$$F(x;\sigma_\lambda) = \underset{\hat{x}\geq 0}{\operatorname{argmin}} \left[l(\hat{x}) + \frac{1}{2\sigma_\lambda^2}||x - \hat{x}||_2^2 \right]. \tag{3}$$

As in the proof of Lemma 1 in the Appendix, we use the fact that $AA^T = L^2 I$ to reparameterize the solution of (3) in a simpler form. We also use clipping to impose the physical constraint of nonnegativity as in the constrained minimization of (3). This yields

$$F(x;\sigma_\lambda) = \left[x + \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + L^2\sigma_w^2} A^T \left(y - \frac{1}{L^2} Ax \right) \right]_+ . \tag{4}$$

The block replication inherent in A^T can lead to blocky artifacts in the final reconstruction, which motivates our introduction of the Relaxed Adjoint Projector in Section III. Sreehari *et al.* used a library-based Non-Local Means (NLM) prior model to incorporate HR data, which led to slow reconstruction times due to the computational cost of NLM [27]. For this reason, we opted to use a deep neural network denoiser as our prior model. PnP has been used with similar denoisers as prior models in [28]–[30]. However, limitations of this previous work include little ability to control regularization strength and the use of generically trained neural networks as opposed to domain-specific denoisers.

C. MACE Reconstruction Framework

In PnP, which is derived from the ADMM solution to a Bayesian inversion problem, forward and prior models can be designed separately, thus promoting modularity and separating sensor modeling from image domain modeling. However, the

resulting fixed point of the PnP algorithm no longer has the probabilistic interpretation given in the Bayesian approach.

Multi-Agent Consensus Equilibrium (MACE) [23] provides a formulation and interpretation of the problem solved by the PnP algorithm and leads naturally to generalizations and analysis of PnP. MACE describes the fixed point of PnP in terms of an equilibrium condition and extends the algorithmic denoisers of PnP to a framework for the fusion of multiple forward and prior models; MACE also provides parametric control of regularization. MACE itself is not an algorithm, but the solution of the MACE equations can be found with PnP or with other algorithms for solving systems of equations.

The motivating problem for MACE is to minimize the function

$$f(x) = \sum_{i=1}^{K} \mu_i f_i(x_i) \text{ s.t. } x_i = x,\ i = 1, \ldots, K, \tag{5}$$

with $x, x_i \in \mathbb{R}^N$ and weights $\mu_i > 0$ with $\sum_{i=1}^{K} \mu_i = 1$.

Buzzard *et al.* [23] describe a framework that generalizes (5) to include algorithmic priors and forward maps (sometimes called agents). For K vector valued maps, $F_i : \mathbb{R}^N \to \mathbb{R}^N$, $i = 1, \ldots, K$, define the consensus equilibrium for these maps to be any solution $(x^*, \mathbf{u}^*) \in \mathbb{R}^N \times \mathbb{R}^{NK}$ such that

$$F_i(x^* + \mathbf{u}_i^*) = x^*,\ i = 1, \ldots K \quad \text{and} \quad \overline{\mathbf{u}}_\mu^* = 0 \tag{6}$$

where $\mathbf{u}$ is a vector in $\mathbb{R}^{NK}$ constructed by stacking vectors $u_1, \ldots u_K$, and $\overline{\mathbf{u}}_\mu$ is the weighted average $\overline{\mathbf{u}}_\mu = \sum_{i=1}^{K} \mu_i u_i$. In the case of (5), the maps F_i are chosen to be proximal maps associated with the f_i.

The conditions in (6) are equivalent to a related system of equations.Namely for $\mathbf{v} \in \mathbb{R}^{NK}$ with $\mathbf{v} = (v_1^T, \ldots, v_K^T)$, $v_i \in \mathbb{R}^N\ \forall i$, define $\mathbf{F}(\mathbf{v})$ by stacking the vectors $F_j(v_j)$, and define $\mathbf{G}_\mu(\mathbf{v})$ by stacking K copies of $\overline{\mathbf{v}}_\mu$. Additionally for $x \in \mathbb{R}^N$, define $\hat{x}$ to be the vector obtained by stacking N copies of x. Then $(x^*, \mathbf{u}^*)$ satisfies (6) if and only if $\mathbf{v}^* = \hat{x}^* + \mathbf{u}^*$ satisfies $\overline{\mathbf{v}}_\mu^* = x^*$ and

$$\mathbf{F}(\mathbf{v}^*) = \mathbf{G}_\mu(\mathbf{v}^*). \tag{7}$$

As in PnP, the F_i may be replaced by more general operators such as CNN denoisers, in which case the solution of (7) is the fixed point of a generalized PnP algorithm. This is solved with Mann iterations in [31] as shown in Algorithm 1, which we use for our results.

In the case of 2 agents as in our results below, this algorithm is equivalent to the PnP algorithm of [14], but the MACE formulation is central to the analysis of RAP presented below.

Algorithm 1: Algorithm to Find MACE Solutions.

Input: Initial Reconstruction $x^{(0)} \in \mathbb{R}^N$
Output: Final Reconstruction x^*
$\mathbf{x} \leftarrow \mathbf{v} \leftarrow \begin{pmatrix} x^{(0)} \\ \vdots \\ x^{(0)} \end{pmatrix}$
while *unconverged* **do**
  $\mathbf{x} \leftarrow \mathbf{F}(\mathbf{v})$
  $\mathbf{z} \leftarrow \mathbf{G}_\mu(2\mathbf{x} - \mathbf{v})$
  $\mathbf{v} \leftarrow \mathbf{v} + 2\rho(\mathbf{z} - \mathbf{x})$
end
$x^* \leftarrow \overline{\mathbf{v}}_\mu$

III. RAP AND MDF

A key strength of the MACE framework is the ability to incorporate algorithmic denoisers and other operators. In this section, we leverage this observation in two ways.

First, we show that using a mismatched backprojector in (4) is equivalent to using the original forward operator in (4) with a modified prior agent. We call the mismatched backprojector a Relaxed Adjoint Projector (RAP). More precisely, we modify the forward operator in (4) by replacing A^T with a related matrix B. This change means that the forward operator is no longer the proximal map for $l(x)$. However, we show using the MACE framework that this formulation is equivalent to a formulation using the original forward operator in (4) but with an alternative prior that depends on B. This provides important intuition for the use of RAP and allows for the application of existing convergence results.

Second, we describe our method of MDF, in which representative HR patches are collected independently of the LR scan. These patches are then used to train a denoising prior model, which intrinsically learns the underlying distribution of the high-resolution modality. We then perform LR scans over a large area and fuse the two modalities using the Multi-Agent Consensus Equilibrium framework to produce a HR image encompassing the full FoV.

A. Relaxed Adjoint Projection

The matrix A^T in (4) is often called a back-projector because it takes the error between measurements y and the forward model and projects them back to image space. In practice, a different matrix may be used in place of A^T; this is known in some contexts as using a mismatched backprojector. This may be done for computational simplicity, as in [32], or because the mismatch leads to improved results, as in earlier work on iterative filtered backprojection [33]–[35].

In the context of (4), we provide an interpretation of this mismatch, which we term Relaxed Adjoint Projection (RAP), as equivalent to the standard backprojector with an appropriately modified prior. That is, the equilibrium problem obtained with a given prior update map and a forward update with a mismatched backprojector has the same solutions as a modified prior map and the standard backprojector update.

For MDF-RAP, we consider the RAP update map

$$\tilde{F}(x; \sigma_\lambda) = \left[x + \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + L^2 \sigma_w^2} B \left(y - \frac{1}{L^2} Ax \right) \right]_+, \tag{8}$$

where the matrix B represents bicubic upsampling by a factor of L and replaces the block replication operator A^T. We note that while bicubic backprojection is used here to avoid blocky artifacts observed when using A^T, other interpolants can be used.

We first describe the equilibrium problem associated with RAP, then prove that the solution of this problem arises from 3 different formulations:

- **RAP forward update**, standard prior, equal weight averaging
- standard forward update, standard prior, **modified averaging**
- standard forward update, **modified prior**, equal weight averaging.

Thus, replacing A^T with B is equivalent to using A^T with a modified prior that can be described in terms of B.

In general, the update in (8) is not a proximal map for any function when the matrix BA is not symmetric since a symmetric Jacobian is a property of all proximal maps. By converting the RAP update into a modified prior, we recover the ability to use standard convergence results while maintaining the benefits associated with mismatched backprojection. We note that results in [36] prove convergence for an adjoint mismatch in the Proximal Gradient Algorithm but do not address the equivalence described here.

To describe RAP further, note that the first-order optimality condition for a solution of (5) when all μ_j are equal is

$$\nabla f_1(x^*) + \cdots + \nabla f_K(x^*) = 0. \tag{9}$$

Also, the proximal map for a convex and differentiable f_j is given by $F_j(v_j) = x_j = v_j - \nabla f_j(x_j)$; i.e, the update can be regarded as an implicit gradient descent step, with the gradient evaluated at the output point $F_j(v_j) = x_j$.

In the case of a mismatched backprojector, we assume for the moment that $F_j^{R_j}$ is given by $F_j^{R_j}(v_j) = x_j = v_j - R_j \nabla f_j(x_j)$ for some matrix R_j, which we think of as close to the identity (precise conditions on R are given in the theorem statements in the appendix). In the case of (4) and (8), this corresponds to $RA^T = B$, and if B is close to A^T, then R can be chosen to be close to the identity. Using this F^R and all $\mu = 1/K$ in the equilibrium condition (7), we have

$$v_j^* - R_j \nabla f_j(x_j^*) = \frac{1}{K} \sum_k v_k^*. \tag{10}$$

Since the left hand side is x_j^* for each j and the right hand side is independent of j, we have $x_j^* = x^*$ is independent of j. Summing these equations over j and subtracting the sum of the v_j^* from both sides and taking the negative gives

$$R_1 \nabla f_1(x^*) + \cdots + R_K \nabla f_K(x^*) = 0. \tag{11}$$

This is the corresponding equilibrium condition for the operators $F_j^{R_j}$ and equal weight averaging.

Theorem 1 states that the set of solutions of the equilibrium condition with RAP are the same as those obtained using standard back projections but an alternative averaging operator $\mathbf{G}^R$, given by a matrix-weighted average using the R_j as matrix weights. As before, we stack the operators $\mathbf{F}_j^{R_j}$ to obtain $\mathbf{F}^R$. The details of the notation, hypotheses, and the proof are given in the appendix.

Theorem 1: Under appropriate hypotheses (see appendix) on $\mathbf{F}$ and R, there is a map from solutions $\mathbf{v}^*$ of

$$\mathbf{F}^R(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$$

to solutions $\hat{\mathbf{v}}^*$ of

$$\mathbf{F}(\hat{\mathbf{v}}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$$

such that for each such pair, $\mathbf{G}(\mathbf{v}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$. There is also such a map from $\hat{\mathbf{v}}^*$ to $\mathbf{v}^*$.

The map $\mathbf{G}(\mathbf{v}^*)$ is obtained by stacking the solution x^*, so this theorem implies that these two formulations have the same set of possible reconstructions. The following corollary applies this to give the equivalence between the use of mismatched backprojection in the data-fitting operator and the use of standard backprojection with an alternative prior.

Theorem 2: With appropriate assumptions (see appendix) on $B = RA^T$ and the denoiser H, and with equal weighting $\mu_j = 1/2$, the following two choices lead to the same MACE solution in (7):

- $F_1 = \tilde{F}$ is the RAP update in (8) and $F_2 = H$ is a given prior operator;
- $F_1 = F$ is the standard update in (4) and $F_2 = H \circ \Phi_R$, where Φ_R is a function dependent on the matrix R.

This makes precise the idea that the mismatch in RAP is equivalent to a corresponding modification to the update step of the prior operator. By using the mismatch, which is often more efficiently implemented in the form of RAP than with alternative averaging, we gain the ability to more closely match the prior to the observed structure of the data without changing the algorithmic implementation of the prior.

B. Convergence of Relaxed Adjoint Projection (RAP)

From [23], Algorithm 1 is known to converge when the map $\mathbf{v} \mapsto 2\mathbf{F}(\mathbf{v}) - \mathbf{v}$ is nonexpansive, and this condition is satisfied when each F_j is the proximal map for a convex function. When the RAP update is used, then F_j is typically no longer a proximal map and may not be nonexpansive.

In Theorem 3 we give conditions under which F_j using RAP is a proximal map after an appropriate change of coordinates. The key idea, related to work in [37], is to consider operators of the form $F(x) = Wx + q$. When F is a proximal map, then W is symmetric and positive-definite. A change of variables allows us to recover this property even for some non-symmetric W. This allows us to prove convergence of Algorithm 1 when using the RAP forward model and denoiser H as in Theorem 2. The proof is in the appendix.

Theorem 3: Let $F(x) = Wx + q$ where $W = V\Lambda V^{-1}$ is an $N \times N$ matrix, Λ is diagonal with eigenvalues in (0,1], and $q \in \mathbb{R}^N$, and let H be a denoiser such that $V^{-1}HV$ is nonexpansive. Then the Algorithm 1 converges using the operators $F_1 = F$ and $F_2 = H$.

C. Multi-Resolution Data Fusion

The MACE formulation with either the standard or RAP forward operator provides a natural way to incorporate low-resolution data and maintain fidelity of the reconstruction to this

data. For MDF, we use a neural network denoiser as a prior agent to incorporate selected high-resolution data, either from the image under reconstruction or from related images.

The theoretical foundation of Plug-and-Play implies that the prior operator should be a denoiser for images in the target distribution perturbed by AWGN, in principle independent of the noise present in the data itself [38]. However, as seen in [21] the prior operator can play a large role in the quality of the final reconstruction. In the context of learned denoisers such as CNNs, this means that the CNN must denoise well on the images in the distribution under consideration. Since we use an iterative algorithm, the CNN must also denoise well on neighboring images in order to converge to a high-quality reconstruction. Ideally, the denoiser should be able to take any image in the reconstruction space and move it closer to an image in the target distribution, but in practice we must settle for an approximation based on a sparsely sampled set of images near the distribution.

We note here that increasing the super-resolution factor L necessarily increases the set of data-consistent reconstructions – increasing L increases the dimension of the kernel of A. In particular, given two reconstructions that both fit the data equally well and that are both equally well-denoised by the denoiser (more precisely, both are equilibrium solutions), there is no reason to favor one over the other. This means first that the importance of the prior operator increases with L and second that larger L gives any reconstruction algorithm more opportunity to "hallucinate" detail that may or may not be present in the true image. In the context of scientific and medical applications where the reconstruction can influence important decisions, it can be detrimental to push the limits of super-resolution and/or regularization beyond reasonable expectations [39].

IV. Methods and Results

We apply the MDF-RAP method on 3 microscopy datasets with pronounced differences in data distribution: gold nanorods, an *E. coli* biofilm, and pentacene crystals. The gold nanorod images are composed of non-overlapping, nearly linear segments at various angles with nearly circular impurities. The *E. coli* biofilm images contain a wide variety of shapes and textures as well as large regions of nearly empty space. The pentacene crystal images are typically composed of large regions of relatively constant intensity with sharply defined edges. The nanorod and pentacene images were obtained using scanning electron microscopy, while the *E. coli* images were obtained using transmission electron microscopy.

We present comparisons of MDF-RAP with a variety of alternatives for both 4x and 8x interpolation. At 4x we compare with bicubic interpolation, DPSR, DPSRGAN [11], and PnP with DnCNN trained on natural images. However, at 8x, we do not compare with DPSRGAN since it is not available for this interpolation rate.

A. Computational Methods

We use a MACE formulation as in (7) with two agents, one for data fidelity and one to incorporate prior information in the form of a denoiser. With 2 agents, the relative weights μ_i simplify to one weight μ for the forward agent and the complementary weight $1-\mu$ for the prior agent. Since the prior agent operates in the reconstruction space of HR images, we train a CNN denoiser to remove 10% AWGN from high-resolution target images.

We use Algorithm 1 to determine the corresponding reconstructions. Based on the MACE equation $\mathbf{F}(\mathbf{v}) = \mathbf{G}(\mathbf{v})$, we define a measure of convergence error as

$$\text{Convergence Error} = \frac{||\mathbf{G}(\mathbf{v}) - \mathbf{F}(\mathbf{v})||_2}{\sigma_n ||\mathbf{G}(\mathbf{v})||_2}, \tag{12}$$

where σ_n is the noise level used to train the prior model. We display a plot of convergence behavior in Section IV-C.

For a baseline case (labeled PnP below), we use the standard backprojector as in equation (4) and a DnCNN denoiser [40] trained on *natural images.* We use $\mu = 0.5$ for all results with this approach. Algorithm 1 in this context is equivalent to Plug-and-Play [23], hence the label PnP. This case is comparable to other methods with no domain-specific training.

For the data fusion case (labeled MDF-RAP below), we use the RAP backprojector as in equation (8) and a DnCNN denoiser trained on *representative high-resolution microscopy images.* Although this involves two changes relative to the PnP baseline, Theorem 2 implies that using RAP is equivalent to using the standard backprojector with a further modified prior. In order to disambiguate RAP from the domain-specific prior, we illustrate the effect of RAP separately in Fig. 7. In the case of MDF-RAP, the use of RAP changes the effect of the prior, so we adjust the regularization strength using $\mu = 0.8$ for synthetic data and $\mu = 0.2$ for real data.

The DnCNN architecture consists of 17 total layers with the following structure: (i) Conv+ReLU for the first layer with 64 filters of size 3x3x1. (ii) Conv+BatchNorm+ReLU for layers 2-16 with 64 filters of size 3x3x64. (iii) Conv for the last layer with 1 filter of size 3x3x64 [40]. The network uses a residual mapping D to learn the structure of the noise in its training pairs $(x_{\text{clean}}, x_{\text{noisy}})$. A forward pass through the model is thus given by $\hat{x}_{\text{clean}} = x_{\text{noisy}} - D(x_{\text{noisy}})$. We used code adapted from https://github.com/cszn/KAIR to implement DnCNN.

To train each image-tuned prior for MDF-RAP, we randomly extracted 400 180x180 patches from a high-resolution training image. This represents a naive sampling of the high-resolution data. We employed a 80/10/10 split for training, validation, and testing data. The amount of HR data used for training for each respective dataset is given in Table III. Using these training patches, we generate corresponding noisy patches by adding AWGN with standard deviation $\sigma = 0.1$. These patch pairs are then passed through the network for training (note that there is no pairing of high-resolution images with low-resolution images). We used an increase in validation loss as a stopping criterion to avoid overfitting. Each of our MDF-RAP networks trained for 1-2 hours using 1 Nvidia V100 GPU.

B. Data Generation

In order to provide quantitative accuracy metrics and demonstrate real-world behavior, we present two experimental approaches: (i) partially simulated and (ii) fully real data.

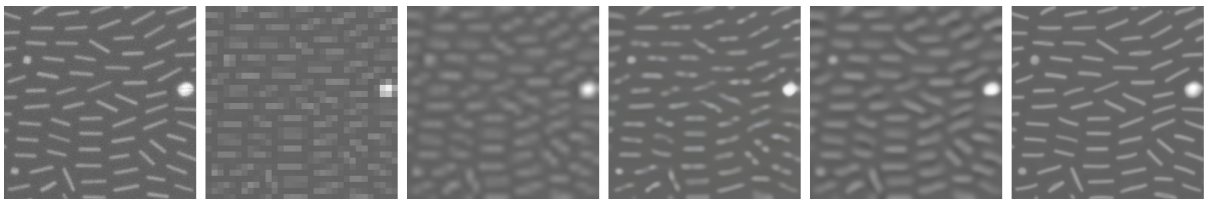

Fig. 3. 8x super-resolution reconstructions of a simulated LR EM image of gold nanorods with $\mu = 0.8$ for MDF-RAP. Each shows a field of view 569 nm wide. MDF-RAP produces nanorods with clear edges and the proper shape, while each of the other methods includes significant blurring and/or incorrect shapes.

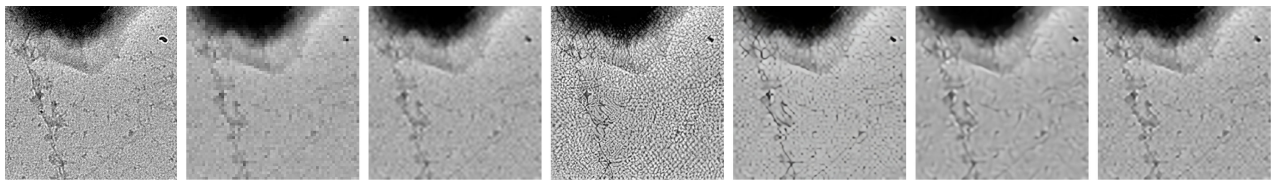

Fig. 4. 4x super-resolution reconstructions of a simulated LR EM image of E coli with $\mu = 0.8$ for MDF-RAP. Each shows a field of view 251 nm wide. MDF-RAP provides the best visual compromise between clarity and fit to data as shown by its reconstruction of distinct curli fibers and background preservation.

TABLE I
ACQUISITION PARAMETERS FOR EXPERIMENTAL DATASETS. WE OMIT LR PIXEL SPACING FOR *E. COLI* AS WE DO NOT PERFORM SUPER RESOLUTION ON MEASURED DATA IN THIS CASE. THE 2.2 NM NANORODS SAMPLE WAS USED FOR THE RESULTS SHOWN IN FIG. 3

Material	HR Pixel Spacing	LR Pixel Spacing	Data Modality
Nanorods	1.1, 2.2 nm	5.5 nm	Scanning
E. coli	0.98 nm	N/A	Transmission
Pentacene	41.7 nm	168.9 nm	Scanning

In the partially simulated case, we use actual HR microscopy data and then apply the forward model $x \mapsto Ax/L^2$ to obtain simulated LR data. Each algorithm is applied to the LR data and the result is compared to the paired HR data using PSNR (and FRC in select cases) to provide quantitative accuracy measures.

In the case of fully real data, both the HR and the LR data are obtained experimentally and used as in the previous case with the exception that we do not have aligned pairs and so cannot quantitatively compare the super-resolved output of an algorithm to a reference HR image. These results on real data allow for qualitative assessment under practical application conditions.

For MDF-RAP, we use domain-specific HR images to train the CNN denoiser, but these are not required to be aligned with corresponding LR images, and the LR images are not used for training. In all cases, we need LR data to illustrate super-resolution. All data used in testing the algorithm (both LR and HR) was taken from the test set and therefore excluded from MDF-RAP training data.

Our HR data was collected using two separate microscopy methods to yield three data sets. Table I describes the experimental parameters used for data acquisition. Scanning electron microscopy images for gold nanorods and pentacene were obtained on an FEI XL30 at 5 kV with a secondary electron detector and a Zeiss Gemini at 5 kV using an in-lens secondary electron detector. The interaction volume of the focused electron beam was on the order of the size of the resulting pixel size in the image. The Scanning Electron Microscopy modality raster scanned a focused beam across the sample with a pixel dwell time of 50 nanoseconds. TEM images for *E. coli* were obtained on a 60-300 Thermo Fisher Titan operating at 300 kV in bright field mode. Images were collected on a 4 k by 4 k Gatan K2 Direct Electron Detector using Serial EM at various electron optical magnifications. LR overview images were first collected, followed by automated HR image montages. The bright field imaging modality uses a wide illumination that covers the entire imaging array.

C. Results on Partially Synthetic Data

In Figs. 3–5, we display a HR ground-truth image, the corresponding simulated LR data, the output of bicubic upsampling (as a baseline), DPSRGAN (when possible), DPSR, PnP (using a CNN trained on natural images as the prior agent), and MDF-RAP (using RAP and domain-specific training). Note that MDF-RAP is the only method that incorporates microscopy data in its training. None of Bicubic Interpolation, DPSRGAN, DPSR, and PnP have seen the microscopy data; we include these comparison points to illustrate the limits of domain-independent methods. The comparisons with MDF-RAP are meant to characterize improvements obtained by incorporating domain-specific data fusion in the prior model.

Fig. 3 illustrates 8x super-resolution on gold nanorods, which are highly structured. Although 8x super-resolution is significantly underconstrained, MDF-RAP leverages the data-fidelity

Fig. 5. 4x super-resolution reconstructions of a simulated LR EM image of pentacene crystals with $\mu = 0.8$ for MDF-RAP. Each shows a zoomed field of view 5.34 μ wide. MDF-RAP performs well in terms of PSNR (and FRC in Fig. 8) while DPSR produces edges that are even sharper than the edges in the HR GT.

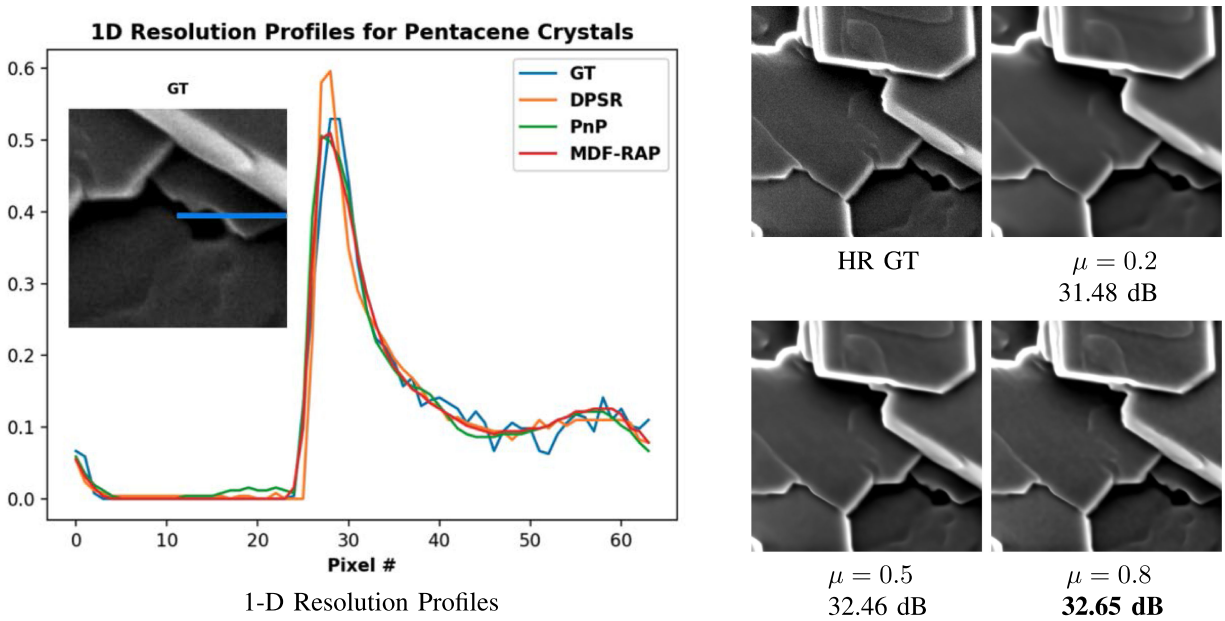

Fig. 6. *Left:* Plot of intensities along the blue segment for selected reconstructions from Fig 5. DPSR produces the sharpest edge boundary, while PnP and MDR-RAP better approximate GT image on crystal edges. *Right:* Comparison of 4x super-resolution reconstructions of a simulated LR EM image of pentacene crystals at varying regularization levels. Note as μ increases, more detail is present in the final reconstruction. However this can introduce textural artifacts from the forward model into the final reconstruction.

operator and a domain-specific CNN to produce a high-quality reconstruction. Here the competing reconstructions each have significant shortcomings. Additionally note that no retraining was required to apply PnP or MDF-RAP at 8x, since changes to the resolution factor requires only a change of L in the forward model rather than retraining the prior model.

In Fig. 4, with 4x super resolution on data with less regularity and more fine detail, MDF-RAP produces the highest PSNR, although not by a significant margin. However, MDF-RAP arguably provides the best visual compromise between clarity and fit to data among the competing methods. While DPSR produces lines and background structures with sharper edges than MDF-RAP, these sharp features do not always align with structures in the HR GT.

In Fig. 5, we show 4x super resolution reconstructions of pentacene crystals. MDF-RAP outperforms the other methods in terms of PSNR, with PnP a close second in this metric. Visually, DPSR produces an image with edges that are sharper than in the HR GT image, which may account for the lower PSNR of DPSR relative to MDF-RAP. Such sharpness is clear in the 1D resolution profile shown in Fig. 6, as the peak of DPSR's plot is significantly higher than that of the GT. To further investigate the the accuracy of MDF-RAP versus DPSR, we use Fourier Ring Correlation to compare these reconstructions to the HR GT (discussion below and plots in Fig. 8).

TABLE II
AVERAGE PSNRS OVER 100 IMAGES FOR *E. COLI*, PENTACENE, AND NANORODS DATASETS AT 4X INTERPOLATION. ON AVERAGE, TECHNIQUES INCORPORATING MDF OUTPERFORM ALL OTHER METHODS

Material	Bicubic	DPSR-GAN	DPSR	PnP	MDF	**MDF-RAP**
Nanorods	32.43	29.87	34.55	34.21	**34.97**	34.95
E. coli	19.98	14.27	19.69	19.95	**20.16**	20.05
Pentacene	30.00	28.84	32.65	33.78	34.09	**34.29**

In Table II, we show average results for reconstructions of synthetic LR data at 4x interpolation. For each dataset, we extracted 100 256x256 images and created corresponding simulated LR data. These images were then passed into each interpolation method for reconstruction. Finally we collected the PSNRs

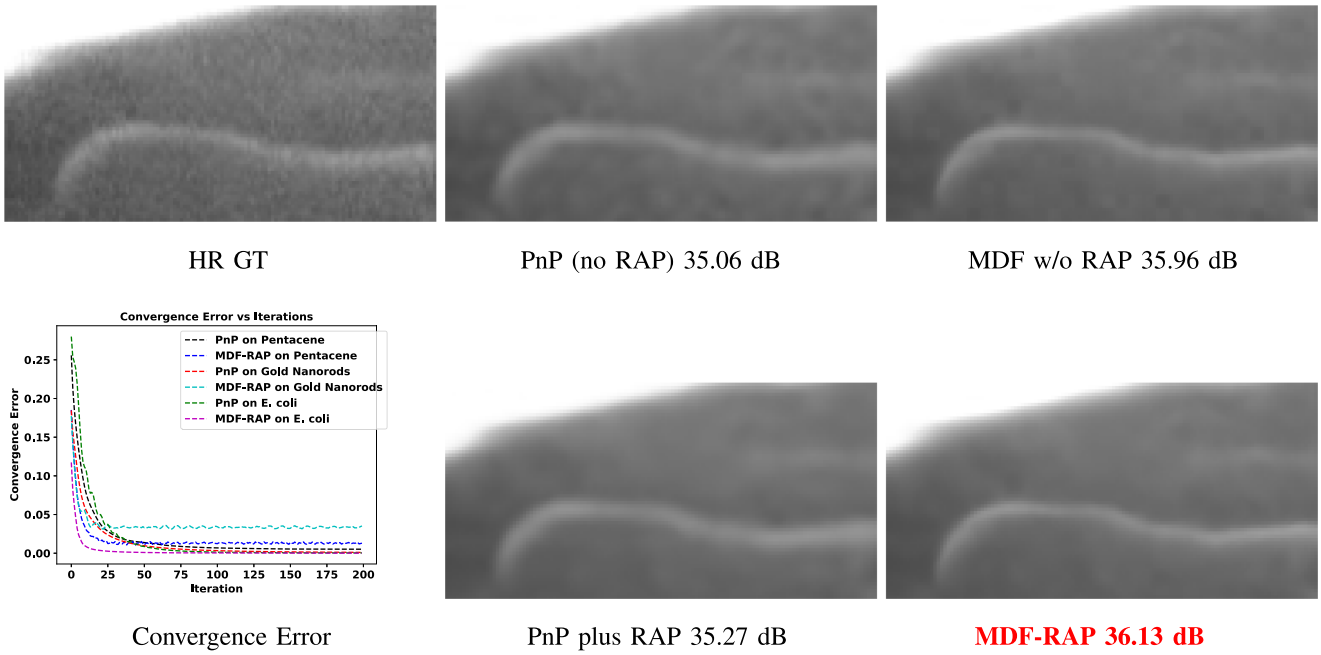

Fig. 7. *Top row and bottom right:* 4x super-resolution reconstructions of a simulated LR image of pentacene crystals as in Figs. 5 and 6. Here PnP (with or without RAP) uses a DnCNN denoiser trained only on natural images, while MDF (with or without RAP) uses a domain-specific denoiser. PnP (no RAP) has a herringbone texture, while MDF w/o RAP has 4x4 blocky artifacts, neither of which appears in the HR GT or the reconstructions with RAP. MDF with or without RAP produces a clearer edge than the PnP reconstructions. *Bottom left:* Convergence Error (equation (12)) plots for 4x PnP and 4x MDF-RAP on pentacene, gold nanorods, and *E. coli* datasets. All methods converge with under 5% error, with fastest convergence on the *E. coli* dataset. .

relative to the original high-resolution image and averaged these across the 100 images to generate the values shown. We also include a comparison to using an image-tuned prior model with the forward model in (4), labeled as MDF. As shown in Table II, methods using an image-tuned prior outperform all other interpolation methods tested.

We investigate the effect of the prior model on image reconstruction in two ways. First, the weight μ in the MACE formulation can be used to tune the regularization strength of the prior. As μ increases, the forward model is weighted more and thus the amount of regularization is reduced. In Fig. 6, we compare reconstructions as μ varies and note the increased smoothing that occurs as μ is decreased. We used $\mu = 0.8$ for the reconstructions in Figs. 3–5 to preserve image detail in the final reconstructions.

Second, we investigate domain-specific training and RAP one at a time in Fig. 7. While the use of RAP is explicitly a change in the forward agent, it is equivalent to a change in the prior. The top 2 reconstructions show PnP and MDF without RAP, both of which yield unnatural artifacts in texture. PnP has a herringbone pattern, while MDF without RAP has 4x4 blocks that do not match the surrounding background. The results with RAP in the bottom two images do not have these artifacts. The MDF-RAP reconstruction provides the best clarity and PSNR.

A possible explanation for the artifacts in the reconstructions without RAP is that the equilibrium solution in (6) requires an update from the forward model that is counterbalanced by an equal but opposite update from the prior model. In the case of the standard backprojector, the update in (4) has the form of adding a blocky image in the range of A^T to the existing reconstruction. This must be balanced by an update from the neural network that results in the same output image. A blocky update from a neural network is unlikely unless the underlying image itself has some structure that promotes such an update. We hypothesize that the artifacts seen in the no-RAP reconstructions arise from this need to balance the forward and prior updates and that the use of the smoother RAP update reduces the need for such artificial structures in the final reconstruction.

In the bottom left of Fig. 7, we plot the average convergence error in (12) as a function of iteration over these 100 256x256 images. For each dataset, MDF-RAP converged with under 5% error. The gold nanorods dataset at 8x superresolution leads to the highest convergence error, which is likely due to the existence of multiple data-consistent reconstructions at this level of superresolution.

In Fig. 8, we use Fourier Ring Correlation (FRC) [41] to estimate image resolution. FRC measures the correlation between the frequency spectra of 2 images, with the frequency domain subdivided into concentric rings to define spectral correlation as a function of spatial frequency. The effective image resolution is determined by the intersection of the FRC curve with a given threshold curve, with higher FRC values indicating closer match to the target image. We use code from [42] to generate the plots in Fig. 8.

The plots in Fig. 8 show that MDF-RAP recovers significantly more of the high frequency components of the 8x gold nanorods image than DPSR, which in turn captures more than the LR image. This is consistent with the visual results of Fig. 3. MDF-RAP, DPSR, and the LR image all have very similar FRC curves for the *E. coli* images, again consistent with Fig. 4.

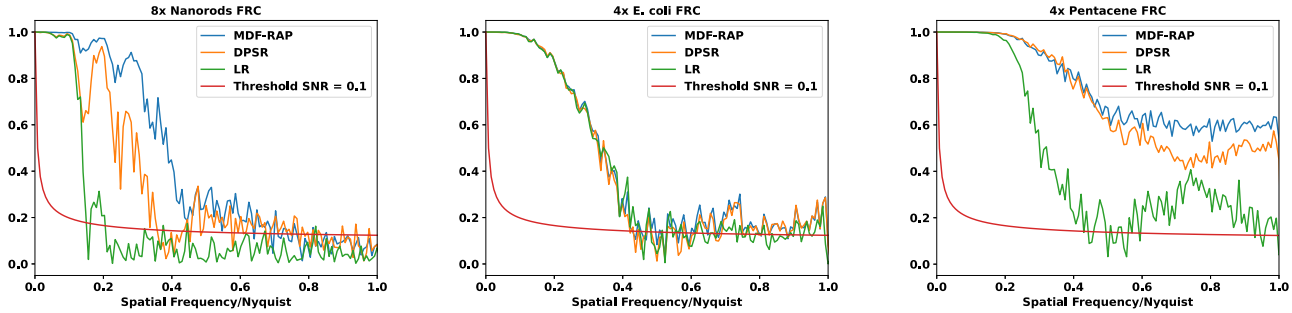

Fig. 8. Fourier Ring Correlation plots for gold nanorods, *E. coli*, and pentacene. FRC describes the correlation in frequency domain between two images; higher values indicate closer correlation. The green lines are from the simulated LR images, with sharp decay near the sampling rates of $1/8$ for nanorods and $1/4$ for *E. coli* and pentacene. MDF-RAP outperforms DPSR on nanorods, but neither method improves significantly over the LR image for *E. coli*. Consistent with the PSNR results in Fig. 3, MDF-RAP outperforms DPSR on the pentacene example, despite the visual appeal of the DPSR result. .

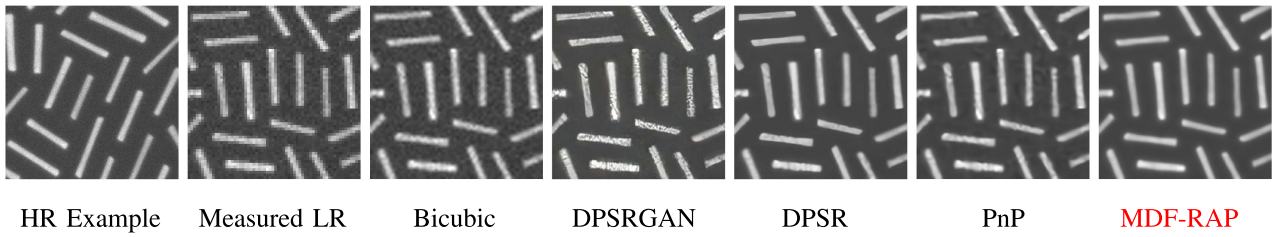

Fig. 9. 4x super-resolution reconstructions of a measured LR EM image of gold nanorods. In this case there is no aligned HR image; the left panel is for qualitative comparison. Each panel shows a field of view 352 nm wide, and MDF-RAP uses $\mu = 0.2$. While all methods reconstruct the nanorods' shape, only MDF-RAP reconstructs the nanorods without internal gaps and maintains the slightly curved ends characteristic of gold nanorods.

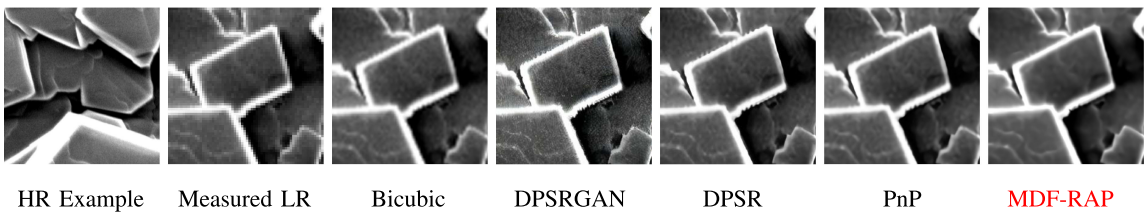

Fig. 10. 4x super-resolution reconstructions of a measured LR EM image of pentacene crystals. In this case there is no aligned HR image; the left panel is for qualitative comparison. Each panel shows a field of view 10.67 μ wide, and MDF-RAP uses $\mu = 0.2$. Note the significant aliasing artifacts along the crystal edges in all non-MDF reconstructions.

In the case of pentacene, the FRC curve for MDF-RAP stays noticeably higher than DPSR in the upper frequency range. This is consistent with the PSNR in Fig. 5 despite the visual appeal of the DPSR result. The intersection of the FRC curves with the threshold SNR curve indicates that DPSR achieves an effective super-resolution of roughly 2x on gold nanorods, while MDF-RAP achieves a super-resolution in the range of 3x to 5x. Neither method provides a significant improvement on the *E. coli* image, and both achieve something close to the target 4x super-resolution on the pentacene image.

D. Results on Measured Data

In Figs. 9–10, we show results using measured LR data as input. In this case there is no paired HR data for quantitative comparison, so we provide a measured HR image of a similar specimen for visual comparison.

In Fig. 9, with 4x superresolution of gold nanorod images, the data is not severely undersampled, so each method is able to reconstruct the shape of the nanorods. However, relative to the other methods, MDF-RAP provides a more faithful reconstruction of the nanorod interiors and ends while removing background noise.

In Fig. 10, the majority of the methods produce aliasing artifacts along the edge of the crystal. MDF-RAP minimizes these artifacts relative to the other methods and again provides a good balance between clarity and realistic texture.

E. Speed-Up and Computational Complexity

In Table III, we examine the speed-up in acquisition time by using MDF-RAP. The speed-up is calculated by taking the ratio of the pixels necessary for a HR FoV to the sum of the pixels in the LR FoV and the pixels in the HR training data.

$$\text{Speed-Up} = \frac{\text{HR Reconstruction Pixels}}{\text{Acquired LR pixels + HR Training Pixels}}$$

TABLE III
DATA ACQUISITION SPEEDUP WITH MDF-RAP. THE SPEEDUP FACTOR IS CALCULATED BY TAKING THE RATIO OF THE NUMBER OF HR RECONSTRUCTED PIXELS TO THE SUM OF THE ACQUIRED LR PIXELS AND HR TRAINING PIXELS

Material	Interpolation Factor	LR data acquired	HR training data	Reconstructed HR data	Speed-Up
Nanorods	5x	2048 x 1388	1232 x 1367	10240 x 6940	15.70x
Nanorods	8x	2048 x 1388	1232 x 1367	16384 x 11104	40.19x
E. coli	4x	7404 x 7666	5049 x 9827	29616 x 30664	8.54x
Pentacene	4x	1280 x 755	1280 x 755	5120 x 3020	10.05x

In the ideal case, in which a domain-specific CNN denoiser is already trained, the acquisition speed-up for Lx interpolation is L^2. In the cases shown in Table III we include the HR data acquisition needed for CNN training, so the actual speed-up ranges from roughly 50% to 62% of the ideal.

A limitation of our method is reconstruction time. Running 20 iterations of MDF-RAP on a 256 x 256 GT image takes 32.6 seconds on a CPU, while DPSR takes 12.2 seconds. This ratio is preserved on a GPU, where MDF-RAP takes 18.8 seconds and DPSR takes 7 seconds. The longer running time of MDF-RAP is likely caused by repeated application of the prior model. One could accelerate the MDF-RAP code by incorporating parallelization and better managing memory, which we leave as future work. In contrast, an advantage of MDF-RAP over DPSR is that the CNN for MDF-RAP is trained on high-resolution images, so does not require retraining for new super-resolution levels or point spread functions.

V. CONCLUSION

We introduced a Multi-Resolution Data Fusion framework with a Relaxed Adjoint Projection and a domain-specific neural network prior operator. RAP leads to reduced textural artifacts, is easy to implement, and is shown to be equivalent to using the standard data fitting operator with a modified prior. The domain-specific neural network prior operator is trained on a limited set of high-resolution images that do not require pairing with low-resolution images.

In a set of experiments, MDF-RAP improves quality relative to existing methods while maintaining data fidelity, accurately resolving sub-pixel-scale features, and providing speed-ups of 8 to 20 times relative to imaging a full high-resolution image. MDF-RAP can be used at multiple super-resolution factors without additional training, and by changing the forward model, it can be used for multiple image acquisition models. This modularity is an important strength in that each component can be used for multiple applications.

More philosophically, the goal of MDF-RAP (and of regularized inversion generally) is to bias the reconstruction to compensate for a set of imperfect measurements. The comparisons between the MDF-RAP and PnP reconstructions show that the results are improved by using domain-specific training data. This bias-variance tradeoff can be pushed too far in the case of many reasonable reconstructions that are consistent with measured data, so any use of super-resolution methods requires thoughtful use and some form of validation. Even so, speed-ups of 8 to 20 times relative to imaging a full high-resolution image justify further investigation and use.

APPENDIX

In this appendix, we prove Theorem 1, which shows that the equilibrium solutions of RAP are identical to those using a standard backprojector with an alternative averaging operator. We then provide a series of lemmas that culminate in Theorem 2, which shows that the use of RAP with a given denoiser is equivalent to using the standard backprojector and an alternative prior. Finally, we prove Theorem 3, which shows that Algorithm 1 with RAP converges to a fixed point.

We begin with a proposition necessary to define several maps.

Proposition 1: Let ϕ be maximally monotone. Then $(I+\phi)^{-1}$ is globally defined, single-valued, and nonexpansive.

Proof: See [43, section 6]. ■

In the theorem below, we use maps $\mathbf{F}_i$ that are implicitly defined in the sense that the map ϕ_i is evaluated at the output of the corresponding map $\mathbf{F}_i$. This is a generalization of the condition satisfied by a proximal map and is equivalently written as $(I+\phi_i)^{-1}$, which is known as the resolvent of ϕ_i, as in Theorem 1.

Theorem 1: Assume that each of ϕ_i and $R_i\phi_i$ are maximal monotone functions from $\mathbb{R}^n$ to itself for each $i=1,\dots,K$, where each R_i is an $n\times n$ matrix with $\sum_i R_i$ invertible. Define

- $\mathbf{F}_i(v_i) = v_i - \phi_i(\mathbf{F}_i(v_i)) = (I+\phi_i)^{-1}(v_i)$
- $\mathbf{F}_i^R(v_i) = v_i - R_i\phi_i(\mathbf{F}_i^R(v_i)) = (I+R_i\phi_i)^{-1}(v_i)$
- $\mathbf{G}_i(\mathbf{v}) = \frac{1}{K}\sum_i v_i$
- $\mathbf{G}_i^R(\mathbf{v}) = (\sum_i R_i)^{-1}(\sum_i R_i v_i)$

Then there is a map from solutions $\mathbf{v}^*$ of

$$\mathbf{F}^R(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$$

to solutions $\hat{\mathbf{v}}^*$ of

$$\mathbf{F}(\hat{\mathbf{v}}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$$

such that for each such pair, $\mathbf{G}(\mathbf{v}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$. There is also such a map from $\hat{\mathbf{v}}^*$ to $\mathbf{v}^*$. Moreover, the common value x^* in the stacked vector $\mathbf{G}(\mathbf{v}^*)$ satisfies $\sum_i R_i\nabla f_i(x^*) = 0$.

Note that $\mathbf{G}(\mathbf{v}^*)$ is formed by stacking copies of the consensus solution x^*, so this theorem says that the two formulations in (i) and (ii) have exactly the same set of consensus solutions.

Proof: Assume that $\mathbf{F}^R(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$, and define x^* to be the identical entries of $\mathbf{G}(\mathbf{v}^*)$, so that $F_i(v_i^*) = x^*$ for each i. Applying $\mathbf{G}$ to both sides yields $\mathbf{G}(\mathbf{F}^R(\mathbf{v}^*)) = \mathbf{G}^2(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$. Expanding using the definitions of $\mathbf{F}_i^R$ and $\mathbf{G}_i$ yields

$$\frac{1}{K}\sum_i (v_i^* - R_i\phi_i(\mathbf{F}_i(v_i^*))) = \frac{1}{K}\sum_i v_i^*. \tag{13}$$

Multiplying by K and canceling the sum of the v_i^* gives

$$\sum_i R_i \phi_i(x^*) = 0. \tag{14}$$

Conversely given x^* such that $\sum_i R_i\phi_i(x^*) = 0$, define $v_i^* = x^* + R_i\phi_i(x^*)$ for all i. We will show that $\mathbf{F}^R(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$. Since $\mathbf{F}_i^R(v_i) = (I + R_i\phi_i)^{-1}(v_i)$, we have

$$\mathbf{F}_i^R(v_i^*) = (I + R_i\phi_i)^{-1}(x^* + R_i\phi(x^*)) = x^*. \tag{15}$$

Also,

$$\mathbf{G}(\mathbf{v}^*)_i = \frac{1}{N}\sum_i (x^* + R_i\phi(x^*)) \tag{16}$$

$$= x^* + \frac{1}{N}\sum_i R_i\phi_i(x^*). \tag{17}$$

Note that the second term in this sum is 0 by assumption, so $\mathbf{G}(\mathbf{v}^*)_i = x^*$ for all i and hence $\mathbf{F}^R(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*)$.

Assume now that $\mathbf{F}(\hat{\mathbf{v}}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$, and let $\hat{x}^*$ be the identical entries of $\mathbf{F}(\hat{\mathbf{v}}^*)$. As before, $\mathbf{G}^R(\mathbf{F}(\hat{\mathbf{v}}^*)) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$ and this with the definitions yields

$$\left(\sum R_i\right)^{-1} \sum R_i(\hat{v}_i^* - \phi_i(\mathbf{F}_i(\hat{v}_i^*))) = \left(\sum R_i\right)^{-1}\left(\sum R_i\hat{v}_i^*\right). \tag{18}$$

Applying $(\sum R_i)$ to both sides, canceling $\sum R_i\hat{v}_i^*$, and using $\mathbf{F}_i(\hat{v}_i^*) = \hat{x}^*$ gives $\sum_i R_i\phi_i(\hat{x}^*) = 0$.

Conversely given $\hat{x}^*$ such that $\sum R_i\phi_i(x^*) = 0$, define $\hat{v}_i^* = \hat{x}^* + \phi_i(x^*)$ for all i. A calculation similar to the previous case shows that $\mathbf{F}(\hat{\mathbf{v}}^*) = \mathbf{G}^R(\hat{\mathbf{v}}^*)$.

Hence for each $\mathbf{v}^*$ satisfying (i), the identical entries x^* of $\mathbf{G}(\mathbf{v}^*)$ satisfy (14). This condition then determines $\hat{\mathbf{v}}^*$ satisfying (ii) and so that x^* is the common entry in $\mathbf{G}^R(\mathbf{v}^*)$. The map from $\hat{\mathbf{v}}^*$ to $\mathbf{v}^*$ is the same in reverse. ■

Lemma 1: The map F in (4) can be expressed as either

1) $F(x) = (I + \nabla f)^{-1}(x)$
2) $F(x) = (I - r\nabla f)(x)$,

where $\sigma^2 = \frac{\sigma_\lambda^2}{\sigma_w^2}$, $f = \frac{\sigma^2}{2}\|y - Ax\|_2^2$ and $r = 1/(1 + \sigma^2 L^2)$.

Proof: We'll first establish the form in 1. Note that the first order optimality condition for $F(x) = \text{argmin}_v\{f(v) + \frac{1}{2}\|v - x\|^2\}$ is $\nabla f(v) + v - x = 0$. Solving for v gives $v^* = (I + \nabla f)^{-1}(x)$. Since f is a positive, semi-definite quadratic penalty, its subdifferential is maximal monotone, hence this inverse is well-defined by Proposition 1.

Using $\nabla f(v) = \sigma^2 A^T(Av - y)$ in the first order optimality condition above and isolating v^* gives

$$v^* = x - \sigma^2 A^T(Av^* - y). \tag{19}$$

Therefore, v^* is of the form $v^* = x + A^T z$ for some z. Using this form of v^* in (19) gives

$$x + A^T z = x - \sigma^2 A^T(A(x + A^T z) - y). \tag{20}$$

Some algebra and $AA^T = L^2 I$ gives

$$A^T(y - Ax) = (1 + L^2\sigma^2)A^T z, \tag{21}$$

with a solution of $z = \frac{\sigma^2}{1+\sigma^2 L^2}(y - Ax)$. We substitute this into $v^* = x + A^T z$ to obtain $F(x) = v^* = x + r\sigma^2 A^T(y - Ax)$, or $(I - r\nabla f)(x)$, where $r = \frac{1}{1+\sigma^2 L^2}$. ■

Lemma 2: Suppose $\text{Lip}(\psi) \le \alpha < 1$. Then $\Psi = I + \psi$ is invertible and

$$\text{Lip}(\Psi^{-1}) \le \frac{1}{1-\alpha} \tag{22}$$

$$\text{Lip}(I - \Psi^{-1}) \le \frac{\alpha}{1-\alpha} \tag{23}$$

Proof: If ψ is Lipschitz with constant α, then $\Psi = I + \psi$ is also Lipschitz. Consequently Ψ will be differentiable almost everywhere by Rademacher's theorem. The forward and reverse triangle inequalities imply

$$(1-\alpha)\|x - z\| \le \|\Psi(x) - \Psi(z)\| \le (1+\alpha)\|x - z\|. \tag{24}$$

These bounds show that the singular values of Ψ are bounded away from 0, which implies its Jacobian is of full rank. The Lipschitz Inverse Function Theorem [44] implies that Ψ is invertible. Defining $w = \Psi(x)$ and $v = \Psi(z)$ transforms the bounds on $\|\Psi(x) - \Psi(z)\|$ to

$$(1-\alpha)\|\Psi^{-1}(w) - \Psi^{-1}(v)\| \le \|w - v\| \le (1+\alpha)\|\Psi^{-1}(w) - \Psi^{-1}(v)\|. \tag{25}$$

Dividing the left hand side of the inequality by $1 - \alpha$ yields $\text{Lip}(\Psi^{-1}) \le 1/(1-\alpha)$.

Now fix v_1 and v_2, and let $w_j = \Psi(v_j) = v_j + \psi(v_j)$, so that $(I - \Psi^{-1})(w_j) = \psi(v_j)$. Using the Lipschitz constants for ψ and Ψ^{-1} gives

$$\|(I - \Psi^{-1})(w_1) - (I - \Psi^{-1})(w_2)\| \tag{26}$$

$$= \|\psi(v_1) - \psi(v_2)\| \tag{27}$$

$$\le \alpha\|v_1 - v_2\| \tag{28}$$

$$= \alpha\|\Psi^{-1}(w_1) - \Psi^{-1}(w_2)\| \tag{29}$$

$$\le \frac{\alpha}{1-\alpha}\|w_1 - w_2\|. \tag{30}$$

■

Lemma 3: Assume $\sigma^2 < 1/L^2$ and let f and r be as in Lemma 1 with this σ^2. Then there exist constants $\delta, C_1 > 0$ such that if R is a matrix satisfying $\|R - I\| < \delta$, then there exists a matrix $\tilde{R}$ depending on R so that

- $\|\tilde{R} - I\| \le C_1\|R - I\|$
- $\tilde{R}\nabla f$ is maximal monotone

and so that

$$x - rR\nabla f(x) = (I + \tilde{R}\nabla f)^{-1}(x). \tag{31}$$

Proof: Define $W = \sigma^2 A^T A$. Note that $A^T A$ can be factored as a projection followed by scaling by L^2, hence $\|rW\| = \sigma^2 L^2/(1 + \sigma^2 L^2) < 1/(1 + \sigma^2 L^2)$ by assumption. Thus there exists $\delta_1 > 0$ so that if $\|R - I\| < \delta_1$, then $\|rWR\| < 1$, hence $(I - rWR)$ is invertible by Lemma 2. As motivated below, let

$$\tilde{R}x = \begin{cases} rR(I - rWR)^{-1}x & \text{for } x \in \text{range}(A^T) \\ Rx & \text{for } x \in \text{null}(A). \end{cases} \tag{32}$$

Since the orthogonal complement of range(A^T) is null(A), we can extend by linearity to all of $\mathbb{R}^n$.

Since $AA^T = L^2 I$, induction shows that $W^k A^T = (\sigma^2 L^2)^k A^T$. Since $\sigma^2 L^2 < 1$, we can expand the first part of (32) at $R = I$ using a convergent power series to get

$$r(I - rW)^{-1} A^T = r \sum_{k=0}^{\infty} r^k W^k A^T \tag{33}$$

$$= r \sum_{k=0}^{\infty} (r\sigma^2 L^2)^k A^T \tag{34}$$

$$= \frac{r}{1 - r\sigma^2 L^2} A^T. \tag{35}$$

Since $r = 1/(1 + \sigma^2 L^2)$, this is A^T.

The same idea shows that for R sufficiently close to the identity, say $\|R - I\| < \delta_2$ for some $\delta_2 \in (0, 1)$, $\tilde{R} = \tilde{R}(R)$ restricted to range(A^T) can be written as a power series in R and satisfies $\tilde{R}(I) = I$. Expanding this power series about $R = I$ gives constants $c_1, c_2 > 0$ so that

$$\|\tilde{R} - I\| \leq (c_1 + c_2 \|R - I\|)\|R - I\|. \tag{36}$$

Define $C_1 = \max(1, c_1 + c_2)$. Since $\|R - I\| \leq 1$, we have

$$||\tilde{R} - I|| \leq C_1 ||R - I|| \tag{37}$$

on range(A^T), hence on all of $\mathbb{R}^n$ since $\tilde{R} = R$ on null(A).

We now show that $\tilde{R}\nabla f$ is maximal monotone. Note that

$$\tilde{R}\nabla f(x) - \tilde{R}\nabla f(w) = \sigma^2 \tilde{R} A^T A(x - w), \tag{38}$$

so $\tilde{R}\nabla f$ is maximal monotone when $\tilde{R}A^T A$ is positive semi-definite. Note that if $x \in$ null(A), then $x^T \tilde{R} A^T A x = 0$. If $x \in$ range(A^T), then $x = A^T z$ for some $z \in \mathbb{R}^n$. Substituting this for the right-side x gives

$$x^T \tilde{R} A^T A x = x^T \tilde{R} A^T A A^T z. \tag{39}$$

Since $AA^T = L^2 I$, this is $L^2 x^T \tilde{R} x$. By reducing δ_2 if needed, (37) implies that $\tilde{R}$ is positive definite. Extending by linearity shows that $\tilde{R}A^T A$ is positive semi-definite, so $\tilde{R}\nabla f$ is maximal monotone. Let $\delta = \min\{\delta_1, \delta_2\}$.

Finally, let $\tilde{F}(x) = x - rR\nabla f(x)$, so $\tilde{F}(x) = (I - rR\nabla f)(x)$. Then $\tilde{F}(x) = (I + \tilde{R}\nabla f)^{-1}(x)$ exactly when

$$(I + \tilde{R}\nabla f) \circ (I - rR\nabla f)(x) = x \tag{40}$$

Expanding and rearranging gives

$$\tilde{R}\nabla f \circ (I - rR\nabla f) = rR\nabla f. \tag{41}$$

Let $p = \sigma^2 A^T y$, so that $\nabla f(x) = Wx - p$. Using this in (41) gives

$$\tilde{R}(Wx - rWR(Wx - p) - p) = rR(Wx - p), \tag{42}$$

and collecting terms gives

$$\tilde{R}(I - rWR)(Wx - p) = rR(Wx - p). \tag{43}$$

Since $Wx - p$ maps $\mathbb{R}^n$ onto range(A^T), this is equivalent to $\tilde{R}(I - rWR) = rR$ on range(A^T), which is consistent with (32). Hence $\tilde{R}$ as defined in (32) satisfies (31), thus completing the proof. ■

Lemma 4: Let f be as in Lemma 1 and let ϕ be a Lipschitz and strongly maximal monotone function with Lipschitz constant k. Then there exists a constant $\delta > 0$ such that if R is a matrix satisfying $\|R - I\| < \delta$, then $\tilde{R}^{-1}\phi$ is maximal monotone, where the matrix $\tilde{R}$ is from Lemma 3. Moreover, there exists a function Φ_R and constant C such that

$$\text{Lip}(\Phi_R - I) \leq C\|R - I\|$$

and so that

$$(I + \phi)^{-1} \circ \Phi_R = (I + \tilde{R}^{-1}\phi)^{-1}.$$

Proof: It suffices to show $\Phi_R(x) = (I + \phi)(I + \tilde{R}^{-1}\phi)^{-1}(x)$ has the desired properties. We add and subtract ϕ and factor out $(I + \phi)^{-1}$ to obtain

$$\Phi_R = (I + \phi)(I + \phi + (\tilde{R}^{-1} - I)\phi)^{-1} \tag{44}$$

$$= (I + \phi)[(I + (\tilde{R}^{-1} - I)\phi(I + \phi)^{-1})(I + \phi)]^{-1} \tag{45}$$

$$= (I + (\tilde{R}^{-1} - I)\phi(I + \phi)^{-1})^{-1}. \tag{46}$$

Hence $\Phi_R = (I + \psi)^{-1}$ with $\psi = (\tilde{R}^{-1} - I)\phi(I + \phi)^{-1}$. By Lemma 3, $\|\tilde{R} - I\| \leq C_1\|R - I\|$. Restrict R so that this is less than 1/2, so by Lemma 2 with $\Psi = \tilde{R}$,

$$\|\tilde{R}^{-1} - I\| \leq 2C_1\|R - I\|. \tag{47}$$

Let $d_1 = Lip((I + \phi)^{-1})$, which is at most 1 since the resolvent of a monotone operator is nonexpansive [43]. Then

$$Lip(\psi) \leq 2C_1\, k d_1 \|R - I\|. \tag{48}$$

Hence there exists $\delta_1 > 0$ such that if $\|R - I\| < \delta_1$, then $Lip(\psi) < 1$, in which case Lemma 2 implies $\Phi_R = (I + \psi)^{-1}$ is well-defined with $Lip(\Phi_R) \leq 1/(1 - Lip(\psi))$. Lemma 2 with $\Psi = I + \psi = \Phi_R^{-1}$ implies

$$Lip(I - \Phi_R) = Lip(I - \Psi^{-1}) \leq \frac{Lip(\psi)}{1 - Lip(\psi)}. \tag{49}$$

By (48), we can choose $\delta > 0$ so that if $\|R - I\| < \delta$, then $Lip(I - \Phi_R) \leq C\|R - I\|$, where $C = 4C_1\, k d_1$.

Recall that ϕ strongly monotone means that there exists $m > 0$ so that for all x, v, $(x - v)^T(\phi(x) - \phi(v)) \geq m||x - v||_2^2$. Let $\eta = \tilde{R}^{-1} - I$. By (47), we can reduce δ to get $\|\eta\| < \frac{m}{2k}$. To show that $\tilde{R}^{-1}\phi$ is maximal monotone, note that

$$(x - v)^T(\tilde{R}^{-1}\phi(x) - \tilde{R}^{-1}\phi(v))$$

$$= (x - v)^T(\phi(x) - \phi(v)) - (x - v)^T(\eta(\phi(x) - \phi(v))$$

By assumption, the first term is bounded below by $m||x - v||_2^2$. Additionally $\|\phi(x) - \phi(v)\| \leq k\|x - v\|$ by assumption, so

$$(x - v)^T(\tilde{R}^{-1}\phi(x) - \tilde{R}^{-1}\phi(v)) \tag{50}$$

$$\geq m||x - v||_2^2 - k\|\eta\|\|x - v\|_2^2, \tag{51}$$

which gives a lower bound of $\frac{m}{2}\|x - v\|_2^2$. Hence $\tilde{R}^{-1}\phi$ is strongly monotone, and since $\tilde{R}^{-1}$ is linear, $\tilde{R}^{-1}\phi$ is maximal monotone. ■

Theorem 2: Suppose ϕ is a Lipschitz and strongly maximal monotone function and let $H = (I + \phi)^{-1}$ and $\mu_1 = \mu_2 = 1/2$. Assume $\sigma^2 < 1/L^2$. Then there exist $\alpha > 0$ and $C > 0$ such that if R is a matrix satisfying $RA^T = B$ and $\|R - I\| \leq \alpha < 1$, there exists Φ_R a Lipschitz map depending on H and R with $\text{Lip}(\Phi_R - I) \leq C\|R - I\|$ such that the following two choices lead to the same set of solutions x^* in (7):

- $\mathbf{F}_1 = \tilde{F}$ is the RAP update in (8) and $\mathbf{F}_2 = H$;
- $\mathbf{F}_1 = F$ is the standard update in (4) and $\mathbf{F}_2 = H \circ \Phi_R$.

This theorem says that the effect of RAP with a denoiser H can be explained by using the standard data-fitting term together with a modified denoiser defined by a Lipschitz transformation of the image domain followed by the original denoiser.

Proof: In order to apply Theorem 1, we verify that each $\mathbf{F}_j$ is of the form $(I + \omega)^{-1}$ for ω a maximal monotone function. By Lemma 1, we have that $F = (I + \nabla f)^{-1}$ and ∇f is maximal monotone. As ϕ is maximal monotone, H is of the desired form, and H is well-defined by Proposition 1. Since $\sigma^2 < 1/L^2$, Lemma 3 with the same f, r as defined in Lemma 1 implies that $\tilde{F} = (I + \tilde{R}\nabla f)^{-1}$. Lemma 3 also gives that $\tilde{R}\nabla f$ is maximal monotone. Finally as ϕ is assumed to be a Lipschitz and strongly maximal function, Lemma 4 gives that $H \circ \Phi_R = (I + \tilde{R}^{-1}\phi)^{-1}$ is well-defined and that $\tilde{R}^{-1}\phi$ is maximal monotone. Thus we may apply the results of Theorem 1 to each pair $(\tilde{F}, H)$ and $(F, H \circ \Phi_R)$.

From the proof of Theorem 1, since $\tilde{F} = (I + \tilde{R}\nabla f)^{-1}$ and $H = (I + \phi)^{-1}$ we see that $\mathbf{v}^* = (v_1^*, v_2^*)$ is a solution of

$$\begin{bmatrix} \tilde{F}(v_1^*) \\ H(v_2^*) \end{bmatrix} = \mathbf{G}(\mathbf{v}^*)$$

if and only if

$$\tilde{R}\nabla f(x^*) + \phi(x^*) = 0,$$

where x^* is the common entry of the stacked vector $\mathbf{G}(\mathbf{v}^*)$. Since $\tilde{R}$ is invertible, this is equivalent to

$$\nabla f(x^*) + \tilde{R}^{-1}\phi(x^*) = 0.$$

Since $F = (I + \nabla f)^{-1}$ and $H \circ \Phi_R = (I + \tilde{R}^{-1}\phi)^{-1}$, again the proof of Theorem 1 implies that this is true if and only if $\tilde{\mathbf{v}}^* = (\tilde{v}_1^*, \tilde{v}_2^*)$ is a solution of

$$\begin{bmatrix} F(\tilde{v}_1^*) \\ H(\Phi_R(\tilde{v}_2^*)) \end{bmatrix} = \mathbf{G}(\tilde{\mathbf{v}}^*)$$

with x^* the common entry of the stacked vector $\mathbf{G}(\tilde{\mathbf{v}}^*)$.

This implies that the two formulations have the same set of consensus solutions, x^*. ■

Proposition 2: If A is an $n \times n$ symmetric matrix with eigenvalues in (0,1] and $b \in \mathbb{R}^n$, then the mapping $F(x) = Ax + b$ is a proximal map.

Proof: The conditions on A imply that $A^{-1} - I$ is symmetric and positive semidefinite, so the Cholesky decomposition gives an invertible R such that $R^T R = \frac{1}{\sigma^2}(A^{-1} - I)$ (for specified $\sigma^2 > 0$). Define p so that $\sigma^2 AR^T p = b$ and consider the proximal map defined by

$$\text{argmin}_u \left\{ \frac{1}{2}||Ru - p||^2 + \frac{1}{2\sigma^2}||u - x||^2 \right\}. \tag{52}$$

The first-order optimality condition yields

$$R^T(Ru^* - p) + \frac{1}{\sigma^2}(u^* - x) = 0, \tag{53}$$

and gathering the u^* terms and multiplying by σ^2 gives

$$(I + \sigma^2 R^T R)u^* = x + \sigma^2 R^T p. \tag{54}$$

Noting that $(I + \sigma^2 R^T R) = A^{-1}$ and using the choice of p gives $u^* = Ax + b$. Hence $F(x) = Ax + b$ is a proximal map. ■

Proposition 3: Let $F(x) = Wx + q$ where $W = V\Lambda V^{-1}$ with Λ diagonal having eigenvalues in (0,1] and $q \in \mathbb{R}^N$. For $x \in \mathbb{R}^N$ define $\hat{x} = V^{-1}x$. Then $\hat{F}(\hat{x}) = V^{-1}F(V\hat{x})$ is a proximal map in the coordinates $\hat{x}$.

Proof: Expanding $\hat{F}$ using $W = V\Lambda V^{-1}$ gives $\hat{F}(\hat{x}) = \Lambda\hat{x} + V^{-1}q$. Since Λ is diagonal with eigenvalues in (0,1], the previous proposition implies that $\hat{F}(\hat{x})$ is a proximal map. ■

Proposition 3 applies with F as the RAP update in (8) and $\hat{F}$ as the standard update in (4). Theorem 3 then implies that Algorithm 1 converges to a fixed point using the RAP update.

Theorem 3: Let F and $\hat{F}$ be as in Proposition 3. Let H be a denoiser such that $\hat{H}(\hat{x}) = V^{-1}H(V\hat{x})$ is nonexpansive in the coordinates $\hat{x}$. Then Algorithm 1 converges using the operators F and H.

Proof: An expansion of $\mathbf{F}$ and $\mathbf{G}$ shows that Algorithm 1 with two operators is equivalent to the standard PnP algorithm of [21]. By Proposition 3, $\hat{F}$ is a proximal map, and by assumption, $\hat{H}$ is nonexpansive. Hence by [23], Algorithm 1 using operators $\mathbf{F}_1 = \hat{F}$ and $\mathbf{F}_2 = \hat{H}$ converges to a fixed point.

The bilinear change of variables $\hat{x} = V^{-1}x$ yields a one-to-one correspondence

$$\hat{F}(\hat{x}) = V^{-1}F(V\hat{x}) \tag{55}$$

$$\hat{H}(\hat{x}) = V^{-1}H(V\hat{x}) \tag{56}$$

Applying this to each component and map in Algorithm 1 produces a shadow sequence equivalent to running Algorithm 1 with $\mathbf{F}_1 = F$ and $\mathbf{F}_2 = H$. This shadow sequence converges by continuity of V and V^{-1}, so the PnP algorithm converges using F and H. ■

ACKNOWLEDGMENT

The authors would like to thank C. M. Hampton for assistance in data acquisition, K. Park and J. Streit for gold nanorods fabrication, and A. Mehmood for his mentorship.

REFERENCES

[1] K. Park *et al.*, "Highly concentrated seed-mediated synthesis of monodispersed gold nanorods," *ACS Appl. Mater. Interfaces*, vol. 9, no. 31, pp. 26363–26371, 2017.

[2] Y. He, J. Wu, D. Qiu, and Z. Yu, "Construction of 3-D realistic representative volume element failure prediction model of high density rigid polyurethane foam treated under complex thermal-vibration conditions," *Int. J. Mech. Sci.*, vol. 193, Mar. 2021, Art. no. 106164.

[3] H. S. Anderson, J. Ilic-Helms, B. Rohrer, J. Wheeler, and K. Larson, "Sparse imaging for fast electron microscopy," *Proc. SPIE*, vol. 8657, Feb. 2013, Art. no. 86570C.

[4] S. A. Dyer and J. S. Dyer, "Cubic-spline interpolation. 1," *IEEE Instrum. Meas. Mag.*, vol. 4, no. 1, pp. 44–46, Mar. 2001.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Eur. Conf. Comput. Vis.*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 184–199.

[6] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super resolution by deep CNN with skip connection and network in network," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 217–225.

[7] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 105–114.

[8] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4501–4510.

[9] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 63–79.

[10] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN : Further improving enhanced super-resolution generative adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 3637–3641.

[11] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1671–1681.

[12] Y. Qian, J. Xu, L. F. Drummy, and Y. Ding, "Effective super-resolution methods for paired electron microscopic images," *IEEE Trans. Image Process.*, vol. 29, pp. 7317–7330, Jun. 2020.

[13] A. Shocher, N. Cohen, and M. Irani, "Zero-shot super-resolution using deep internal learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3118–3126.

[14] S. Sreehari, S. V. Venkatakrishnan, K. L. Bouman, J. P. Simmons, L. F. Drummy, and C. A. Bouman, "Multi-resolution data fusion for super-resolution electron microscopy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 88–96.

[15] L. Marti-Bonmati, R. Sopena, P. Bartumeus, and P. Sopena, "Multimodality imaging techniques," *Contrast Media Mol. Imag.*, vol. 5, pp. 180–189, 2010.

[16] M. Moseley and G. Donnan, "Multimodality imaging," *Stroke*, vol. 35, no. 11_suppl_1, pp. 2632–2634, 2004.

[17] F. Alam and S. U. Rahman, "Challenges and solutions in multimodal medical image subregion detection and registration," *J. Med. Imag. Radiat. Sci.*, vol. 50, no. 1, pp. 24–30, 2019.

[18] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[19] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.

[20] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Springer, 2020, pp. 208–224.

[21] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.

[22] S. Sreehari *et al.*, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *Trans. Comput. Imag.*, vol. 2, pp. 408–423, 2016.

[23] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium," *SIAM J. Imag. Sci.*, vol. 11, no. 3, pp. 2001–2020, Jan. 2018.

[24] M. U. Ghani and W. C. Karl, "Data and image prior integration for image reconstruction using consensus equilibrium," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 297–308, Mar. 2021.

[25] A. Brifman, Y. Romano, and M. Elad, "Turning a denoiser into a super-resolver using plug and play priors," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1404–1408.

[26] P. Bree, C. M. M. van Lierop, and P. Bosch, "On hysteresis in magnetic lenses of electron microscopes," in *Proc. IEEE Int. Symp. Ind. Electron.*, 2010, pp. 268–273.

[27] B. Liu and J. Liu, "Overview of image noise reduction based on non-local mean algorithm," *MATEC Web Conf.*, vol. 232, 2018, Art. no. 03029.

[28] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3929–3938.

[29] K. Zhang, Y. Li, W. Zuo, L. Zhang, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," 2020, *arXiv:2008.13751*.

[30] B. Shi, Q. Lian, and H. Chang, "Deep prior-based sparse representation model for diffraction imaging: A plug-and-play method," *Signal Process.*, vol. 168, 2020, Art. no. 107350.

[31] S. Majee, T. Balke, C. A. J. Kemp, G. T. Buzzard, and C. A. Bouman, "4D X-ray CT reconstruction using multi-slice fusion," in *Proc. IEEE Int. Conf. Comput. Photography*, 2019, pp. 1–8.

[32] S. V. Venkatakrishnan, E. Cakmak, H. Billheux, P. Bingham, and R. K. Archibald, "Model-based iterative reconstruction for neutron laminography," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, 2017, pp. 1864–1869.

[33] D. S. Lalush and B. M. W. Tsui, "Improving the convergence of iterative filtered backprojection algorithms," *Med. Phys.*, vol. 21, no. 8, pp. 1283–1286, 1994.

[34] A. Ziabari *et al.*, "Model based iterative reconstruction with spatially adaptive sinogram weights for wide-cone cardiac CT," in *Proc. 5th Int. Conf. Image Formation X-Ray Computed Tomogr.*, 2018, pp. 15–18.

[35] R. Schofield *et al.*, "Image reconstruction: Part 1 - Understanding filtered back projection, noise and image acquisition," *J. Cardiovasc. Computed Tomogr.*, vol. 14, no. 3, pp. 219–225, 2020.

[36] E. Chouzenoux, J.-C. Pesquet, C. Riddell, M. Savanier and Y. Trousset "Convergence of proximal gradient algorithm in the presence of adjoint mismatch," *Inverse Problems 37*, Jan. 2021, doi: 10.1088/1361-6420/abd85c.

[37] P. Nair, R. G. Gavaskar, and K. N. Chaudhury, "Fixed-point and objective convergence of plug-and-play algorithms," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 337–348, Mar. 2021.

[38] C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play: A general approach for the fusion of sensor and machine learning models," *SIAM News*, vol. 54, no. 2, pp. 1–4, Mar. 2021.

[39] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proc. Nat. Acad. Sci.*, vol. 117, no. 48, pp. 30088–30095, Dec. 2020.

[40] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[41] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *J. Struct. Biol.*, vol. 183, no. 3, pp. 363–367, 2013.

[42] S. Ali. "Fourier ring correlation," 2018. [Online]. Available: https://github.com/s-sajid-ali/FRC

[43] E. K. Ryu and S. Boyd, "Primer on monotone operator methods," *Appl. Comput. Math.*, vol. 15, no. 1, pp. 3–43, 2016.

[44] F. H. Clarke, "On the inverse function theorem," *Pacific J. Math.*, vol. 64, no. 1, pp. 97–102, 1976.