MR Image Denoising and Super-Resolution Using Regularized Reverse Diffusion

Hyungjin Chung, Eun Sun Lee, and Jong Chul Ye, *Fellow, IEEE*

***Abstract*— Patient scans from MRI often suffer from noise, which hampers the diagnostic capability of such images. As a method to mitigate such artifacts, denoising is largely studied both within the medical imaging community and beyond the community as a general subject. However, recent deep neural network-based approaches mostly rely on the minimum mean squared error (MMSE) estimates, which tend to produce a blurred output. Moreover, such models suffer when deployed in real-world situations: out-of-distribution data, and complex noise distributions that deviate from the usual parametric noise models. In this work, we propose a new denoising method based on score-based reverse diffusion sampling, which overcomes all the aforementioned drawbacks. Our network, trained only with coronal knee scans, excels even on out-of-distribution *in vivo* liver MRI data, contaminated with a complex mixture of noise. Even more, we propose a method to enhance the resolution of the denoised image with the *same* network. With extensive experiments, we show that our method establishes state-of-the-art performance while having desirable properties which prior MMSE denoisers did not have: flexibly choosing the extent of denoising, and quantifying uncertainty.**

***Index Terms*— Diffusion model, stochastic contraction, denoising, MRI.**

I. Introduction

MAGNETIC resonance imaging (MRI) is a widely used noninvasive imaging modality that can produce high resolution patient scans which aid diagnosis. Nevertheless, it is often the case where images are corrupted with complex noise, which obstructs clinicians from pointing out the details. Denoising is crucial in such cases, and for that matter, several techniques [1], [2], [3], [4], [5], [6], [7] have been developed over the years.

Before the widespread popularity of deep learning, denoising methods such as spatial filtering [8], transform domain filtering [9], non local means [10], etc. were studied extensively. Nowadays, deep learning based methods are mainstream, inaugurating state-of-the-art (SOTA) over traditional methods. Denoising methods based on supervised training [11], [12] were the first to be developed, but these are far from being practical, since paired acquisition of clean and noisy images is rarely possible. To circumvent this difficulty, several unsupervised and self-supervised methods were proposed [13], [14], [15]. Although these methods do not require paired images for training, the performance degrades when the distribution of test data deviates from the training data, or when the actual noise distribution differs from the parametric model assumption (e.g. Gaussian noise). Furthermore, most denoising methods resort to the minimum mean-squared-error (MMSE) estimate, leading to outputs that are blurrier than the noisy inputs [16]. This phenomenon was theoretically studied in [17], where the authors showed that there is an inevitable trade-off between distortion (i.e. Error from the ground truth), and perception (i.e. distance between the true data distribution and the distribution of the reconstructions); MMSE denoisers will lead to outputs that minimize the distortion at the expense of degraded perception, or to put it another way, blurred image. The analysis made in [17] is mostly confined to natural images, and theoretical analysis of *perception* in the medical imaging field has not been rigorously conducted. Having said that, one of the contributions of this paper is to reveal how perception-oriented denoisers may be beneficial in terms of diagnostic capacity in the field of medical imaging.

The downsides which MMSE denoisers cause are especially relevant in the context of MRI denoising. While one can simply consider thermal noise in the k-space acquisition as Gaussian noise, in several cases the noise distribution is much more complex. One example of this can be seen when the data was acquired with high flip angle, reduced FOV scan. In such cases, MMSE denoisers (e.g. [13]) either fail to produce feasible reconstructions due to distribution shift, or produce washed out results.

Recently, diffusion models [18], [19] have shown impressive progress in image generation [18], [19], [20], outperforming

Manuscript received 3 October 2022; revised 30 October 2022; accepted 5 November 2022. Date of publication 7 November 2022; date of current version 3 April 2023. This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] through the Artificial Intelligence Graduate School Program, Korea Advanced Institute of Science and Technology (KAIST) under Grant 2019-0-00075; in part by the National Research Foundation (NRF) of Korea under Grant NRF-2020R1A2B5B03001980; and in part by the KAIST Key Research Institute (Interdisciplinary Research Group) Project. *(Corresponding authors: Eun Sun Lee; Jong Chul Ye.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Chung-Ang University Hospital, Korea, under IRB No. 2112-011-19394.

Hyungjin Chung is with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea (e-mail: hj.chung@kaist.ac.kr).

Eun Sun Lee is with the Department of Radiology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Dongjak-gu, Seoul 06973, Republic of Korea (e-mail: seraph377@gmail.com).

Jong Chul Ye is with the Department of Mathematical Sciences, Department of Bio and Brain Engineering, Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea (e-mail: jong.ye@kaist.ac.kr).

Digital Object Identifier 10.1109/TMI.2022.3220681

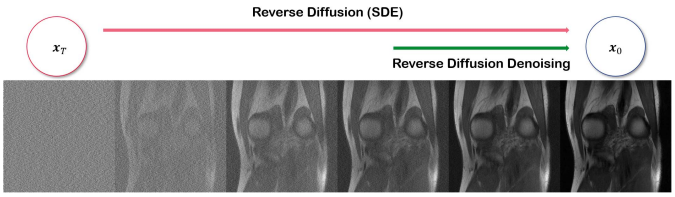

Fig. 1. Overview of the proposed reverse diffusion denoising scheme. *Hijacking* the reverse diffusion process leads to a strong denoiser. Our model is trained on an open-source knee dataset [27], yet is scaled to in vivo liver MRI scans.

even the best-in-class generative adversarial networks (GAN). While diffusion models were first developed as generative models, these are now also being adopted to inverse problems including compressed sensing MRI [21], [22], [23], CT reconstruction [23], super-resolution [24], [25], [26], and much more. Two very appealing properties of diffusion models are as follows: 1) One can acquire results from posterior sampling, rather than a single MMSE estimate. 2) They are robust to distribution shifts and tend to generalize even to heavily out-of-distribution (OOD) test data [21].

Leveraging such intriguing properties, we propose to use a score-based diffusion model [19] to solve the task of denoising. More specifically, we propose to *hijack* the generative process of diffusion models, not starting from pure Gaussian noise, but starting from the distribution of the noisy images. Illustrated in Fig. 1, this amounts to using the last few steps of the reverse diffusion, where the number of iterations can be approximated by parameter estimation methods [28]. Furthermore, to control the denoising process, such that fine structures are preserved, we propose two novel regularizers. First, low-frequency constraint is introduced, which naturally connects to the recent theory of stochastic contraction in diffusion models [26]. Second, we propose to use the manifold constraint [29], which allows us to leverage the best of both worlds: MMSE denoiser, and iterative diffusion denoiser.

Going further, we propose a method to super-resolve the denoised image with the *same* score function that was used to denoise the images. This immediately leads to sharper images which retain the high frequency information, which has not been reported before with any of the widely used self-supervised denoising methods.

The closest work to our proposed method is Noise2Score [15], where the authors utilize the score function to perform denoising in a single step by utilizing Tweedie's formula. Contrarily, our method *iteratively* refines the noisy images by solving the reverse SDE with the score function, thereby enabling fine-grained control, and realistic results. Moreover, since our method is essentially performing posterior sampling, we can obtain multiple samples to quantify uncertainty, and to also obtain posterior mean of the distribution, as pointed out in the earlier work of CS-MRI [21].

II. Background

A. Diffusion Models: A Bird-Eye View

Diffusion models, original introduced in Shol-Dickstein et al. [30] and extended in a seminal work [18], first defines a forward process where the data structure is gradually destroyed (i.e. corrupted with noise). The reverse diffusion process defines the *generative* process, in which the data is gradually synthesized (i.e. denoised) starting from completely unstructured noise. Notably, one can *train* a neural network that learns this reverse process defined as a Markov chain, such that the trained neural network can be used as an iterative denoiser. By constructing such forward-reverse diffusion process, one can model an arbitrary target distribution, which is achieved by converting a tractable prior distribution (e.g. isotropic Gaussian) by iterative diffusion steps. We illustrate the typical forward-reverse diffusion steps in Fig. 1.

One can view the networks that govern the iterative reverse diffusion process as (Stein) score functions [31] (i.e. $\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})$), the vector field that guides the process to higher density modes when Markov chain Monte Carlo (MCMC) sampling is performed. Furthermore, when analyzed from a continuous time scale, it was shown that the diffusion processes are in fact instantiations of stochastic differential equations (SDE) [19]. Hence, sampling from the modeled data distribution amounts to numerically solving the reverse SDE with the trained score function.

It should also be noted that the reverse diffusion process is versatile enough that one can modify the sampling strategy to sample from the conditional distribution given some measurement $\boldsymbol{y}$, by imposing some simple measurement constraints [19], [26]. This would in turn amount to sampling from the conditional (posterior) distribution. Naturally, one can leverage such methods to solve inverse problems in an unsupervised fashion, with proper modifications to the inference procedure. We formalize the high-level intuition and propose a novel solver tailored for MRI denoising in the following sections.

B. Score-Based Diffusion Model

We consider the usual construction of diffusion process indexed with time $t \in [0, 1]$, which is given by $\{\boldsymbol{x}(t)\}_{t=0}^{1}$, with $\boldsymbol{x}(t) \in \mathbb{R}^D$. Here, $\boldsymbol{x}(0) \sim p_0$, where $p_0 = p_{\text{data}}$, and p_1 approximates isotropic Gaussian distribution. In other words, the data distribution is transformed to a tractable Gaussian as $t \mapsto 1$. To formally represent this, we first introduce the following forward SDE

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + g(t)d\boldsymbol{w}, \tag{1}$$

where $\boldsymbol{f}$ is a linear drift coefficient, g is a scalar diffusion coefficient, and $\boldsymbol{w}$ is a standard D-dimensional Wiener process. In our case, we define $\boldsymbol{f}, g$ to be

$$\boldsymbol{f} = 0, \quad g = \sqrt{\frac{d[\sigma^2(t)]}{dt}}, \tag{2}$$

with

$$\sigma(t) = \sigma_{\min}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^t. \tag{3}$$

This choice is called variance exploding SDE (VE-SDE) [19], and is widely used due to its impressive sample quality.

By Anderson's theorem [32], we subsequently have the following reverse SDE

$$d\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x},t) - g(t)^2 \underbrace{\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})}_{\text{score function}}]dt + g(t)d\bar{\boldsymbol{w}}$$
$$= \frac{d[\sigma^2(t)]}{dt} \underbrace{\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})}_{\text{score function}} dt + \sqrt{\frac{d[\sigma^2(t)]}{dt}} d\bar{\boldsymbol{w}}, \quad (4)$$

where the differential dt indicate time running backwards, and $\bar{\boldsymbol{w}}$ is again a standard D-dimensional Wiener process running backwards in time.

Here, we see that the reverse SDE involves the score function (i.e. gradient of the log probability). To estimate the score function, denoising score matching [33] objective is used, which sidesteps the difficulties, and makes the training easily scalable to higher dimensions [31]. More specifically, we train a time-conditional neural network $\boldsymbol{s}_\theta(\boldsymbol{x},t)$ with the following objective:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t\sim U(0,1)}\Big[\lambda(t)\mathbb{E}_{\boldsymbol{x}(0)}\mathbb{E}_{\boldsymbol{x}(t)|\boldsymbol{x}(0)} \times \Big[\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t),t) - \nabla_{\boldsymbol{x}} \log p_{0t}(\boldsymbol{x}(t)|\boldsymbol{x}(0))\|_2^2\Big]\Big], \quad (5)$$

where $U(0,1)$ denotes the uniform distribution in the interval $[0,1]$. Here, we use the fact that the perturbation kernel $\nabla_{\boldsymbol{x}} \log p_{0t}(\boldsymbol{x}(t)|\boldsymbol{x}(0))$ is always Gaussian when $\boldsymbol{f}$ is chosen to be affine, which is trivially true in our case. Hence, the derivative with respect to $\boldsymbol{x}$ can be computed analytically, which makes the training procedure simple and scalable.

Once the network is optimized via (5), one can plug in the trained score function to numerically solve (4). This can be done by e.g. Euler-Maruyama discretization [19], or higher order methods [34]. On the other hand, one can use predictor-corrector (PC) samplers [19], which alternate between numerical SDE solver and Langevin dynamics [35].

C. Come-Closer-Diffuse-Faster (CCDF) [26]

A downside of diffusion models is that they are very slow. This is because one needs to start sampling from totally random Gaussian noise, and thus few thousand passes through the neural network is necessary to achieve optimal results. Recently, Chung et al. [26] showed that such is not necessarily the case when solving inverse problems with diffusion models. Rather than starting from Gaussian noise, one can forward diffuse the initial corrupted image (measurement), and only use the last few steps of reverse diffusion iteration to arrive at the final reconstruction, given that one uses a proper non-expansive mapping as a data consistency imposing step [36].

This intriguing behavior is originated from the key observation that the score-based reverse diffusion process itself is a stochastic contraction mapping so that as long as the data consistency imposing mapping is non-expansive, the alternating applications of the reverse diffusion and data consistency term results in a stochastic contraction to a fixed point [26]. This theory is ideally suitable for the inverse problems, as the steps which impose data fidelity can be easily cast as non-expansive mapping [26]. In Section III, we discuss ways to regularize the reverse diffusion process for our purpose by applying such non-expansive mapping.

D. Noise2Score

Having access to estimated score also endows us with other capabilities. Suppose that we have the noising process as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w},\ \boldsymbol{w} \sim \mathcal{N}(0, \eta^2 \boldsymbol{I}), \quad (6)$$

where $\boldsymbol{y} \in \mathbb{R}^D$ is the noisy measurement, $\boldsymbol{x} \in \mathbb{R}^D$ is the latent clean image, and $\boldsymbol{w} \in \mathbb{R}^D$ represents the noise vector sampled from $\mathcal{N}$. Here, $\mathcal{N}(0, \eta^2 \boldsymbol{I})$ denotes zero-mean Gaussian with variance η^2. From this model, in order to estimate the clean image $\boldsymbol{x}$ from the noisy image $\boldsymbol{y}$, Tweedie's formula states that the posterior mean of $\boldsymbol{x}$ given $\boldsymbol{y}$ is [37]

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}] = \boldsymbol{y} + \eta^2 \nabla_{\boldsymbol{y}} \log p(\boldsymbol{y}). \quad (7)$$

In another word, we can find the minimum mean square error (MMSE) estimator, given the correct score function of $\boldsymbol{y}$. It was shown by the authors of [37] that one can apply similar procedures to the general noise distributions that follow the exponential family (e.g. Poisson, Gamma).

III. Main Contributions

A. Score Function as Denoiser

Consider a noisy image $\boldsymbol{x}(t)$, defined by the forward diffusion process, which essentially samples from $p(\boldsymbol{x}(t)|\boldsymbol{x}(0))$:

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \boldsymbol{w},\ \boldsymbol{w} \sim \mathcal{N}(0, \sigma(t)^2 \boldsymbol{I}), \quad (8)$$

where $\sigma(t)$ is defined in Eq. (3). We see that this is identical to (6) when $\eta = \sigma(t)$. Moreover, suppose that $\boldsymbol{s}_\theta$ is expressive enough and trained to optimality such that $\boldsymbol{s}_{\theta^*} = \nabla_{\boldsymbol{x}_t} p_t$, i.e.

$$\boldsymbol{s}_{\theta^*}(\boldsymbol{x}(t),t) = \nabla_{\boldsymbol{x}_t}\left(-\frac{\|\boldsymbol{x}(t)-\boldsymbol{x}(0)\|^2}{2\sigma(t)^2}\right) = -\frac{\boldsymbol{x}(t)-\boldsymbol{x}(0)}{\sigma(t)^2}.$$

Subsequently, using Tweedie's formula in (7) gives us

$$\mathbb{E}[\boldsymbol{x}(0)|\boldsymbol{x}(t)] = \boldsymbol{x}(t) + \sigma(t)^2 \boldsymbol{s}_{\theta^*}(\boldsymbol{x}(t),t)$$
$$= \boldsymbol{x}(t) + \sigma(t)^2\left(-\frac{\boldsymbol{x}(t)-\boldsymbol{x}(0)}{\sigma(t)^2}\right) = \boldsymbol{x}(0). \quad (9)$$

With (9), we again confirm that it is possible to use the estimated score function trained under the score-SDE framework [19] to estimate the denoised image of $\boldsymbol{x}(t)$. The only requirements are that the noise variance $\sigma(t)$ lies within the bound of the noise schedule for training and that we can correctly estimate $\sigma(t)$. In this case, Eq. (9) provides us with a method to denoise $\boldsymbol{x}(t)$ in a *single* step.

However, the single-step denoising via Noise2Score often produces blurred results, which is undesirable since this may raise the difficulty of clinical decision-making. The situation gets severer as the degree of noise scale increases. Indeed, this is a problem not only for denoising with the score function, but also for most of the recent self-supervised denoisers [13], [14], [37], [38] relying on a single-step estimate.

This suggests that fine-grained control and posterior sampling of the denoising estimate would be beneficial for obtaining sharp and high-fidelity denoised images. Fortunately, given the generative reverse SDE in (4), we can deal with the difficulty, which we explore in the following section.

B. Reverse Diffusion as Denoiser

Recall that the reverse SDE in (4) can be numerically solved (integrated) by methods such as Euler-Maruyama discretization. Specifically, we discretize the time index t into N intervals uniformly. In this case, our noise schedule of $\sigma(t)$ in (3) and the discretized sample become

$$\sigma_i := \sigma_{\min}\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^{\frac{i-1}{N-1}}, \quad \boldsymbol{x}_i := \boldsymbol{x}(t)|_{t=\frac{i-1}{N-1}}. \tag{10}$$

In the case of solving reverse SDE for VE-SDE, this amounts to iteratively applying

$$\text{Solver}(\boldsymbol{x}_{i+1}, z) := \boldsymbol{x}_{i+1} + (\sigma_{i+1}^2 - \sigma_i^2)\boldsymbol{s}_\theta(\boldsymbol{x}_{i+1}, \sigma_{i+1}) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2}z, \tag{11}$$

as shown in Algorithm 1. Namely, Euler-Maruyama discretization [39] can be seen analagous to the classic Euler integration in the context of ordinary differential equations (ODE), where one simply applies the first order approximation to the non-linear equation. Applying (11) will produce $\boldsymbol{x}_i$, which will become the next starting point for applying Solver$(\boldsymbol{x}_i, z)$.

Here, we have written down a single Euler-Maruyama step [39] for simplicity, but application of other solvers is straightforward. The solver in Algorithm 1 can be thought of as generating a *sample path* [39], given some initial random noise sample $\boldsymbol{x}_N \sim p_t(\boldsymbol{x})$. By doing so, one would be able to achieve a sample from $p_0(\boldsymbol{x})$, given that the score function was properly trained.

Algorithm 1 Numerical Solving of (4)

Require: $\{\sigma_i\}_i$
1: $\boldsymbol{x}_N \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\max}^2 \boldsymbol{I})$
2: **for** $i = N-1:0$ **do**
3: $\quad z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
4: $\quad \boldsymbol{x}_i \leftarrow \text{Solver}(\boldsymbol{x}_{i+1}, z)$
5: **end for**
6: **return** $\boldsymbol{x}_0$

On the contrary, let us define a noisy image $\boldsymbol{x}_{N'} := \boldsymbol{x}(t')$, where $N' := Nt'$, and $t' \in [0, 1]$. In this case, we can consider that the given noisy image is a sample from $p_{t'}(\boldsymbol{x})$. When we know the noise level σ_{true} of $\boldsymbol{x}(t')$, we can specify the value of t' by $t' = \sigma^{-1}(\sigma_{\text{true}})$. Then, we can generate a sample path starting from the initial value $\boldsymbol{x}(t')$, which would amount to iterating the solver for N' steps, rather than N.

Although in the easier case when we know a priori the amount of noise, we can simply choose t' such that $\sigma(t')$ matches the known noise variance, in most practical situations, this is not possible and one should resort to some surrogate method to estimate the noise parameters. For that matter, we propose to use a non-parametric estimation method based on eigenvalue analysis of the covariance matrix [28], which we denote in Algorithm 2 with E. Note that this is simply a design choice, and one can use other estimators as well [40], [41], [42].

One should note that the noise distribution may deviate from the Gaussian distribution. For the case of single-step estimation as in Noise2Score [15], one should be more careful about how to use the score function and Tweedie's formula to denoise a given image. In contrast, as the proposed method solves the reverse SDE, a single discrete denoising step amounts to denoising with score function, and then adding additive Gaussian noise (i.e. diffusion term). With the iterative application of such a procedure, the noise distribution becomes closer and closer to the Gaussian distribution by design. Coupling this with the inherent generalization capacity of diffusion models [21], we can expect that the proposed method will generalize to other noise distributions as well. We back up our argument with a simulation study on denoising Rician noise in section. V-A.

Algorithm 2 Reverse Diffusion Denoiser

Require: $\{\sigma_i\}_i$, $\boldsymbol{x}_{N'}$, E, α
1: $\sigma_{est} = E(\boldsymbol{x}_{N'})$ ▷ Noise level estimation
2: $t' = \sigma^{-1}(\sigma_{est})$
3: $N' = \alpha t' N$
4: **for** $i = N'-1:0$ **do**
5: $\quad z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
6: $\quad \boldsymbol{x}_i \leftarrow \text{Solver}(\boldsymbol{x}_{i+1}, z)$
7: **end for**
8: **return** $\boldsymbol{x}_0$

Another hyper-parameter to endow further flexibility is α in Algorithm 2. According to the circumstances, a clinician might either have to apply denoising aggressively such that one can maximally reduce the noise level, or apply it mildly so that one can best preserve the original details. Hence, one can decide on the value α, or simply apply Algorithm 2 with different values of α multiple times, to choose the best-performing value. We discuss this interesting aspect of the proposed method further in Section V-C.

C. Low Frequency (LF) Regularizer

Ideally, one may use Algorithm 2 and achieve denoised results. However, due to the inherent stochasticity, simply applying Algorithm 2 may result in altered structure. Especially in medical imaging, this could lead to catastrophic outcome. In order to prevent such outcome, we propose a low frequency regularization scheme.

Specifically, we focus on the fact that noise components are mostly focused on the high frequency part of the k-space of the given image. Hence, it would be reasonable to keep the low frequency component intact, and focus mainly on correcting the high frequency component. Formally, the regularization step can be written as

$$\boldsymbol{x}_{i+1} = \lambda \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega \boldsymbol{F}\boldsymbol{x}_{N'} + (1-\lambda)\boldsymbol{x}_i, \tag{12}$$

where $\boldsymbol{F} \in \mathbb{R}^{D\times D}$ denotes the discrete Fourier transform matrix, $\boldsymbol{P}_\Omega \in \mathbb{R}^{D\times D}$ is a diagonal sub-sampling matrix with ones at the low frequency region Ω, and $\lambda \in [0, 1]$ is a hyper-parameter which puts emphasis on the regularization. Note that (12) can be written as

$$\boldsymbol{x}_{i-1} = \boldsymbol{T}(\boldsymbol{x}_i) := \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b},$$

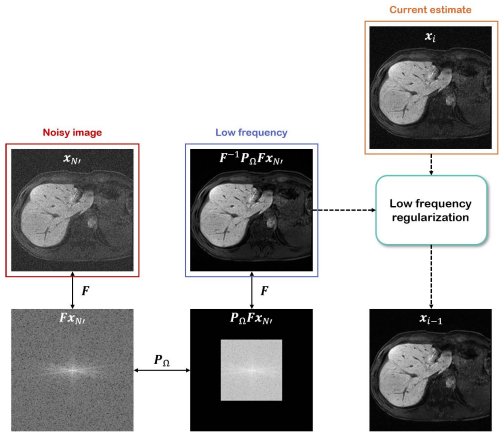

Fig. 2. Illustration of low frequency regularization.

where $\boldsymbol{A} = (1-\lambda)\boldsymbol{I}$ and $\boldsymbol{b} = \lambda \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega \boldsymbol{F}\boldsymbol{x}_{N'}$. Hence, $\boldsymbol{T}$ is nonexpansive since we have $\|\boldsymbol{A}\| = \|(1-\lambda)\boldsymbol{I}\| \leq 1$. Therefore, we can see that (12) not only regularizes the structure to be unaltered, but it also accelerates the contraction to a feasible solution, as proved in [26].

An illustration of the regularization is depicted in Fig. 2. Essentially, the low frequency part from the initial noisy image $\boldsymbol{x}_{N'}$ is softly injected at every iteration in order to keep the important structure intact.

D. Manifold Constraint Regularization

Recall that Noise2Score could acquire the MMSE estimate through a single call of score function as discussed in Section III-A, which we refer to as $\hat{\boldsymbol{x}}_0$. Another regularization that one can impose is to constrain the denoising path such that we minimize the deviation from the MMSE estimate so that we can obtain the following correction for the gradient descent:

$$\hat{\boldsymbol{x}}_0 = \boldsymbol{x}_{i+1} + \sigma_{i+1}^2 s_\theta(\boldsymbol{x}_{i+1}, \sigma_{i+1}) \tag{13}$$

$$\boldsymbol{x}_i'' = \boldsymbol{x}_i' - \gamma \nabla_{\boldsymbol{x}_{i+1}} \|\boldsymbol{x}_i' - \hat{\boldsymbol{x}}_0\|_2^2, \tag{14}$$

where γ is some step size, and $\boldsymbol{x}_i'$ is the previous sampling from the reverse diffusion. By such regularization, we can combine the robustness of a one-step MMSE denoiser, and the detail-preserving property of an iterative diffusion denoiser. Moreover, it was shown in [29] that such regularization imposes the generative path to stay closer to the data manifold, and hence referred to as manifold constraint gradient (MCG) regularization.

E. Post-Hoc Super-Resolution

It is often the case that an image that we wish to see not only is corrupted with noise, but also has low resolution. In such cases, it would be ideal if we could apply super-resolution to the denoised image, since now the noise component would not be amplified. Unfortunately, conventional denoisers are devoid of such ability. In contrast, it has been studied that a single score function can be used to solve multiple inverse problems, when the right consistency term is applied [26].

Algorithm 3 Regularized Reverse Diffusion Denoiser + SR

Require: $\{\sigma_i\}_i$, $\boldsymbol{x}_{N'}$, E, M, α, γ

$\sigma_{est} = E(\boldsymbol{x}_{N'})$ ▷Noise level estimation
$t' = \sigma^{-1}(\sigma_{est})$
$N' = \alpha t' N$
for $i = N'-1 : 0$ **do**
  $z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
  $\boldsymbol{x}_i' \leftarrow \text{Solver}(\boldsymbol{x}_{i+1}, z)$
  $\hat{\boldsymbol{x}}_0 \leftarrow \boldsymbol{x}_{i+1} + \sigma_{i+1}^2 s_\theta(\boldsymbol{x}_{i+1}, \sigma_{i+1})$
  $\boldsymbol{x}_i'' \leftarrow \boldsymbol{x}_i' - \gamma \nabla_{\boldsymbol{x}_{i+1}} \|\boldsymbol{x}_i' - \hat{\boldsymbol{x}}_0\|$
  $\boldsymbol{x}_i \leftarrow \lambda \boldsymbol{F}^{-1}\boldsymbol{P}_\Omega \boldsymbol{F}\boldsymbol{x}_{N'} + (1-\lambda)\boldsymbol{x}_i''$
  } Denoiser
end for
$z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
$\hat{\boldsymbol{x}}_M \leftarrow \boldsymbol{x}_0 + \sigma_M z$
for $j = M-1 : 0$ **do**
  $z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
  $\hat{\boldsymbol{x}}_j \leftarrow \text{Solver}(\hat{\boldsymbol{x}}_{j+1}, z)$
  $\hat{\boldsymbol{x}}_j = (\boldsymbol{I} - \boldsymbol{\Phi})\hat{\boldsymbol{x}}_j' + \boldsymbol{x}_0$
  } SR
end for
return $\hat{\boldsymbol{x}}_0$

Here, we show that we can design a super-resolution algorithm that is applied once the denoising is complete. Specifically, we define a blur kernel $\boldsymbol{h}_D$ which is defined by successive applications of the downsampling filter by a factor D, and upsampling filter by a factor D. This can be represented as a matrix multiplication: $\boldsymbol{P}\boldsymbol{x}' := \boldsymbol{h}_D * \boldsymbol{x}'$, where $\boldsymbol{x}'$ denotes intermediate estimate from the reverse diffusion. Then, we use the following data consistency iteration:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{T}(\hat{\boldsymbol{x}}_i') := (\boldsymbol{I} - \boldsymbol{P})\hat{\boldsymbol{x}}_i' + \boldsymbol{x}_0, \tag{15}$$

where $\hat{\boldsymbol{x}}_i'$ is the current estimate from the diffusion, and $\boldsymbol{x}_0$ is the denoised image from the denoising step. As shown in [26], we can see that $\boldsymbol{T}$ in (15) is non-expansive for the normalized filter $\boldsymbol{h}_D$. Therefore, one can alternately apply the reverse diffusion and data consistency in a recursive manner to achieve a fixed point through stochastic contraction:

$$\hat{\boldsymbol{x}}_i' = \hat{\boldsymbol{x}}_{i+1} + (\sigma_{i+1}^2 - \sigma_i^2) s_\theta(\hat{\boldsymbol{x}}_{i+1}, \sigma_{i+1}) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2}\, z$$
$$\hat{\boldsymbol{x}}_i = (\boldsymbol{I} - \boldsymbol{P})\hat{\boldsymbol{x}}_i' + \boldsymbol{x}_0.$$

Summing everything up, we arrive at Algorithm 3, which we call R2D2+.[1] Note that we have initialized the SR process with a forward diffusion. i.e.

$$\hat{\boldsymbol{x}}_M = \boldsymbol{x}_0 + \sigma_M z. \tag{16}$$

Hence, rather than starting from random Gaussian noise as in [24], one can start from $\hat{\boldsymbol{x}}_M$, and use a small number of iterations to achieve reconstruction, as introduced as CCDF strategy in [26]. Accordingly, both the denoising and the SR steps of R2D2+ require a few tens of iterations, as opposed to other diffusion models which require a few thousand steps of iterations [18], [19], [24].

[1] Short for **R**egularized **R**everse **D**iffusion **D**enoiser + SR.

IV. Methods

A. Experimental Data

We train our score function on the open-sourced fastMRI [27] knee datset.[2] Specifically, we train the network as advised in [21], using 320×320 sized fully-sampled single coil MRI magnitude images. Since the data is a simulated single-coil measurement, the ground truth images itself are contaminated with various sources of noise. However, this is not a problem, since it was shown that one can also train a score function with noisy images [15]. We conduct two experiments with our trained score function - simulation study, and in vivo study.

For the case of denoising simulation study, we test two different noise distributions - Gaussian noise, and Rician noise [43]. Gaussian noise was simulated by adding noise to the magnitude image, whereas Rician noise was simulated by adding the same level of noise to the real, and the imaginary part of the image, and then taking the absolute value. While thermal noise is presented in the measurement domain (i.e. $k-$space), it is standard practice to simulate noise in the image domain, as Gaussian noise remains Gaussian after inverse Fourier transform [2], [6], [44], and in the magnitude domain it becomes Rician noise. Accordingly, we did simulate Rician noise exactly the same way by adding noise in the image domain. Also note that we have included Gaussian denoising, as adding noise directly to the image domain is common practice and this is the case that best matches our theory. We additionally perform a simulation study where the forward model consists of downsampling $320 \times 320 \rightarrow 128 \times 128$ ($\times 2.5$) operator and additive Gaussian noise ($\sigma = 13.0$), and the objective is to both denoise and super-resolve the images.

For the in vivo study, the data for testing was collected from Chungang University Medical center, with the Siemens Skyra scanner. The testing data consists of 28 3-D volumes, acquired via high flip angle CAIPIRINHA [45],[3] with reduced FOV and thin slice thickness (1.5 mm) to capture small lesions. Specific parameters for the dataset are listed in TABLE I.

TABLE I
Specifications for the Liver MRI Test Dataset

Parameter	value
Scanner	Siemens 3T Skyra scanner
Contrast	T1W fast suppressed (FS)
Sequence	Gradient Echo (GRE)
Time of Repetition(TR) [ms]	6.42
Time of Echo (TE) [ms]	2.46
Echo Train Length (ETL)	2
Matrix size	256×256
Resolution [mm^3]	$1.0 \times 1.0 \times 1.5$

B. Details of Implementation

We train our score function $\boldsymbol{s}_\theta$ based on VE-SDE cost [19], formally written as

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t\sim U(\epsilon,1)} \mathbb{E}_{\boldsymbol{x}(0)\sim p_0} \mathbb{E}_{\boldsymbol{x}(t)\sim \mathcal{N}(\boldsymbol{x}(0),\sigma^2(t)\boldsymbol{I})} \times \left[\lambda(t) \left\| \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t) - \frac{\boldsymbol{x}(t) - \boldsymbol{x}(0)}{\sigma^2(t)} \right\|_2^2 \right], \tag{17}$$

with $\lambda = \sigma^2(t)$, such that one achieves the maximum likelihood training objective of [46]. To avoid numerical issues from the unbounded values of score [37], we set $\epsilon = 10^{-5}$.

In the function $\sigma(t)$, used in the diffusion function of the VE-SDE, we need to specify $\sigma_{\min}$ and $\sigma_{\max}$. Following the parameters of [21], we set $\sigma_{\min} = 0.01$, $\sigma_{\max} = 378$. We use a batch size of 2, and train the network using Adam [47] optimizer with a linear warm-up schedule. Specifically, we linearly increase the learning rate for the first 5000 steps of optimization, reaching `2e-4` at the end of warmup. Learning rate remains static after the warmup. Exponential moving average of 0.9999 is applied to the parameters, and the training took about 3 weeks with 2×RTX 2080Ti GPUs. We use `ncsnpp` [19] as the neural network architecture for modeling $\boldsymbol{s}_\theta$, which conditions the network on t with Fourier features [48]. For the numerical SDE solver, we use the predictor-corrector (PC) sampler proposed in [19]. The parameter for LF regularization is set to $\lambda = 0.05$. We set $\alpha = 0.1$ for the simulation study, and $\alpha = 0.2$ for the in-vivo study unless specified otherwise. Static step size of $\gamma = 1.0$ is set for the manifold constraint. We use the official implementation of [28], with the default parameter (patch size of 8) for the estimation of noise variance.

C. Comparison Methods

We compare the proposed method with the state-of-the-art (SOTA) methods. First, to compare against a representative non-deep learning method, we take random sampling non-local means (SNLM) [49], and Block Matching & 3D filtering (BM3D). In order to denoise an image, both SNLM and BM3D require the level of noise variance, which we estimate using the same method that was used for the noise estimation of the proposed method [28].

For deep learning methods, we first compare with self-supervised learning approaches, such as Noise2Noise (N2N) [13], Neighbor2Neighbor (Nei2Nei) [38], and Noise2Score (N2Score) [15]. In order to train both N2N and Nei2Nei methods, we use random Gaussian noise of scale uniformly ranging from $\sigma = 5.0$ to $\sigma = 50.0$, with the default settings as advised in the original works. For N2Score, we only need a pre-trained score function, and we use the same score function that was used for the proposed method. We include the comparison with stochastic image denoising (SID) [50] by using the same score function as ours and using the sampling process as introduced in [50]. We also compare against RicianNet [51], which is a fully supervised method trained to denoise Rician noise specifically.

Note that for the denoising + SR simulation study and the in-vivo study, we perform a post-hoc super-resolution method as proposed in Section III-E. In our efforts to conduct experiments as fair as possible, we trained the state-of-the-art SR model SwinIR [52] on the fastMRI dataset to perform $\times 2.5$ SR, and applied it to all other methods except for the

[2] https://fastmri.org/

[3] **C**ontrolled **A**liasing **I**n **P**arallel **I**maging **R**esults **I**n **H**igher **A**cceleration.

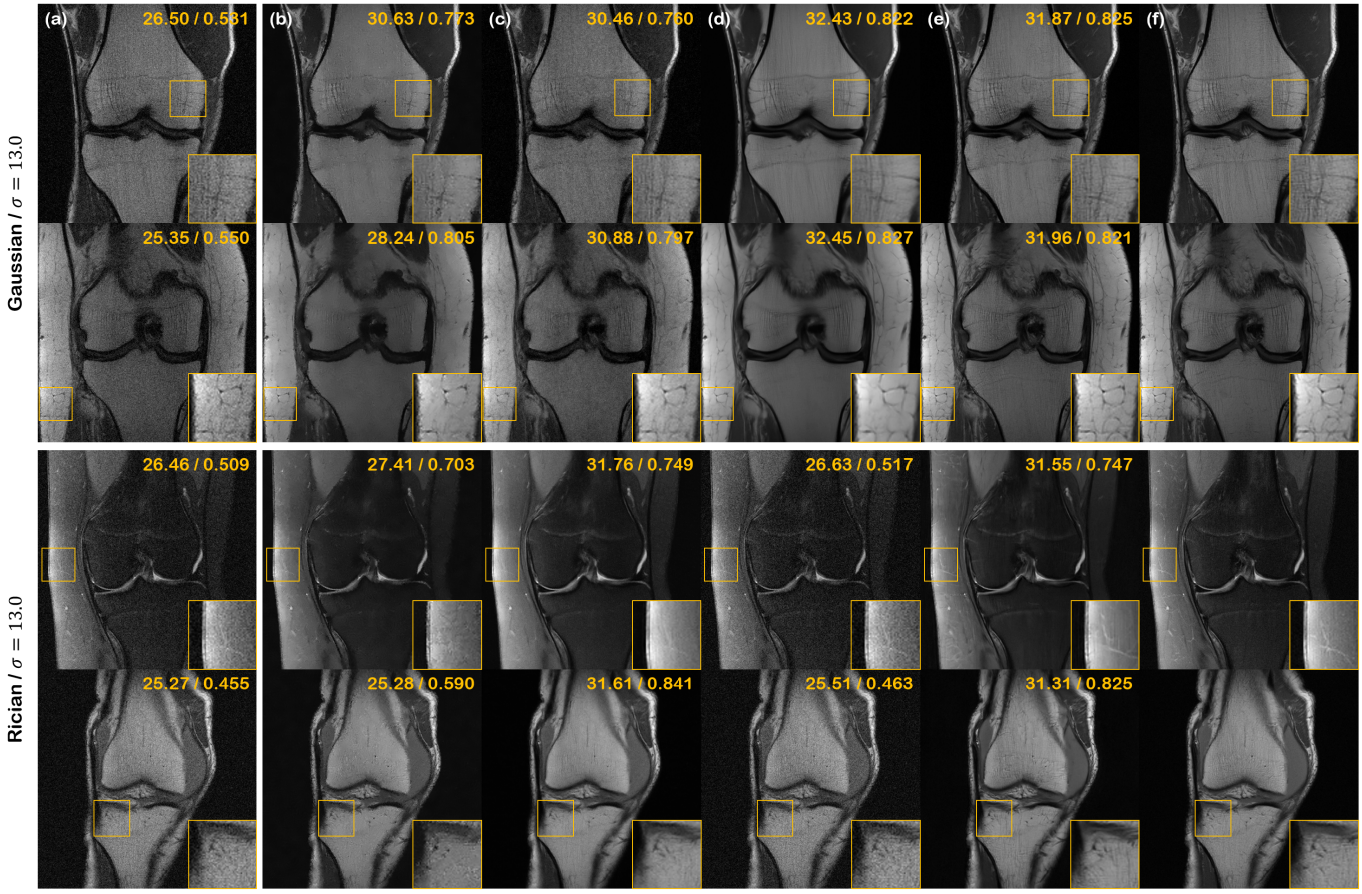

Fig. 3. Denoising results of the simulation study on fastMRI knee dataset. (a) Input noisy image, (b) SNLM [49], (c) RicianNet [51], (d) N2N [13], (e) proposed method, and (f) ground-truth. Yellow boxes show results that are magnified. First two rows: Gaussian noise $\sigma = 13.0$, Next two rows: Rician noise $\sigma = 13.0$.

proposed method. For the experiments involving denoising + SR simulation study data and the in-vivo data, we hence report on results after the SR is applied unless specified otherwise.

D. Quantitative Evaluation and Statistical Analysis

In order to thoroughly evaluate the performance of the various methods, for the simulation study where we have access to ground truth data, we use 4 different metrics: peak-signal-to-noise-ratio (PSNR); structural similarity (SSIM); learned perceptual image patch similarity (LPIPS); and mean average precision (mAP) on the downstream pathology detection task as performed in [21] with the `yolov5` model pretrained on fastMRI+ [53] dataset. Note that PSNR/SSIM are distortion metrics, measuring the exact pixel-wise deviation, whereas LPIPS/mAP are perception metrics, which focuses more on the semantic relevance [17].

For quantifying resolution in the simulation study, we used the local perturbation response (LPR) [54] metric. In essence, LPR measures how faithfully the perturbation made in the images are recovered after the reconstruction. Mathematically,

$$\text{LPR} = G(\mathcal{A}(\boldsymbol{x} + \boldsymbol{p})) - G(\mathcal{A}(\boldsymbol{x})), \tag{18}$$

where G is the reconstruction operator, $\mathcal{A}$ is the forward operator (i.e. in our case down-sampling followed by additive noise), and $\boldsymbol{p}$ refers to the image domain perturbation. We design simple perturbations designed to not throw of the perturbed data off the original data manifold. Namely, we choose two different types of perturbations that consists of different sizes. The first type consists of a medium-sized circular disk near the center of each image, as illustrated in the 3-4$^{\text{th}}$ row of Fig. 4. The second type consists of a small-sized randomly placed circular disks that are spread out near the center of the image, as illustrated in the 5-6$^{\text{th}}$ row of Fig. 4, in order to focus on measuring more of the high-frequency response. In order to quantify, we compute the PSNR of the LPR, which we denote as PSNR-LPR.

The primary objective of the proposed method is to increase the signal-to-noise ratio while also boosting the resolution of the images. Hence, for the in vivo study, where we do not have ground truth images to compute full reference image quality metrics, we use two standard metrics: signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR). Specifically, we calculate the metrics with the following equation:

$$SNR = \frac{\mu(\boldsymbol{x}_{[s]})}{\varsigma(\boldsymbol{x}_{[s]})}, \quad CNR = \frac{\mu(|\boldsymbol{x}_{[s]} - \boldsymbol{x}_{[b]}|)}{\varsigma(\boldsymbol{x}_{[b]})} \tag{19}$$

where $\boldsymbol{x}_{[s]}, \boldsymbol{x}_{[b]}$ refers to the circular masked region of interest (ROI) where the signal is dominant, and the circular masked

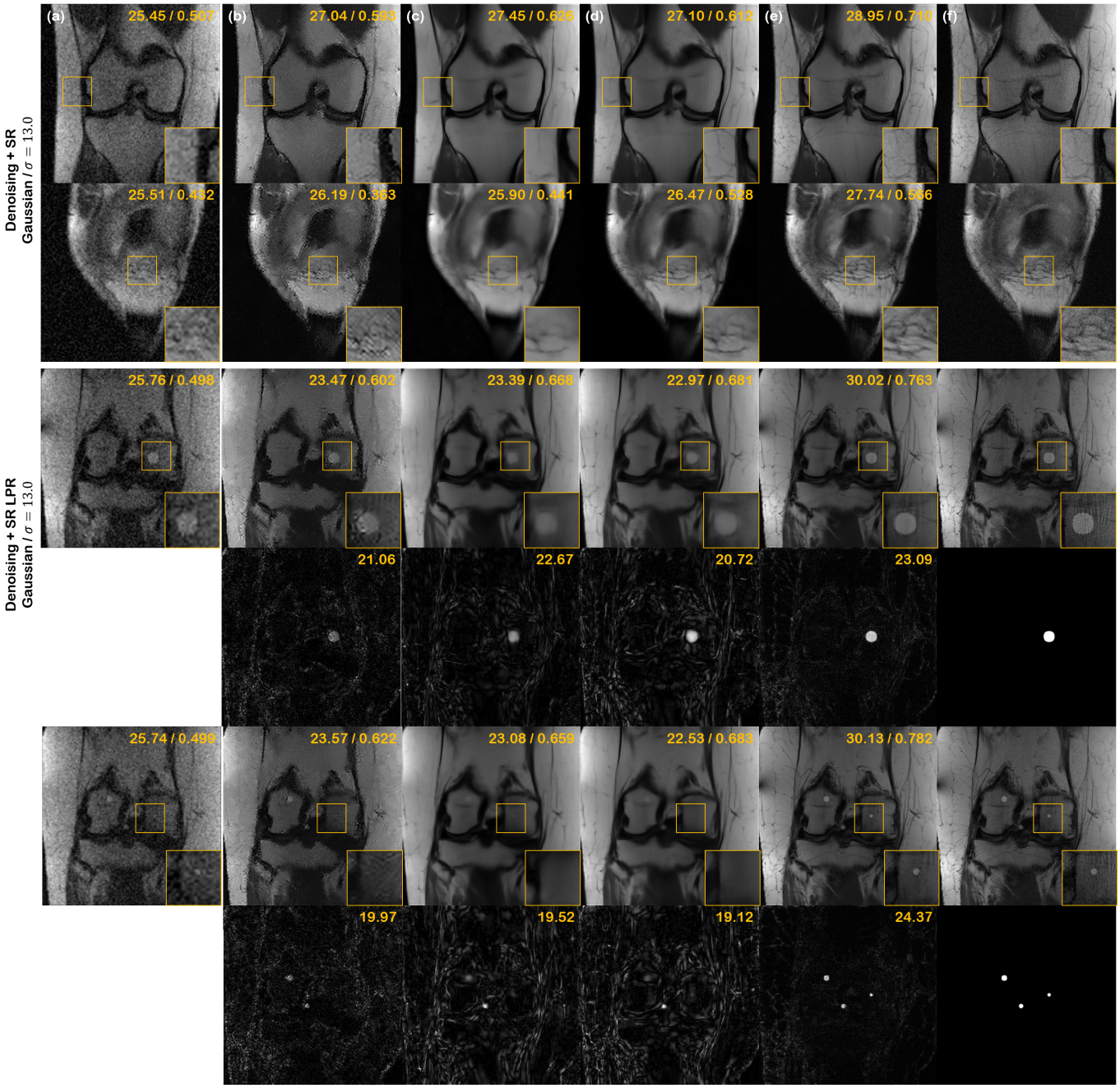

Fig. 4. Denoising+SR results of the simulation study on fastMRI knee dataset. (a) Input noisy image, (b) SNLM [49], (c) RicianNet [51], (d) N2N [13], (e) proposed method, and (f) ground-truth. Yellow boxes show results that are magnified. Third, fourth row: Results of quantifying the LPR metric (medium size); fifth, sixth row: LPR metric (small size). SR is applied via pre-trained SwinIR [52] model.

ROI where the background resides, respectively. Moreover, μ refers to the mean value, and ς indicates the signal standard deviation.

When calculating the SNR, we manually selected 8 distinct regions per volume, corresponding to the 8 different liver segments [55]. Circular disk with the radius of 10 mm was chosen to mask out the ROIs. For the calculation of CNR, we focused on 2 different regions per volume, in which we select two major vessels in the left liver, and the right liver, respectively. The selection of the regions was performed by a board-certified radiologist.

V. Results

A. Simuation Study

We show the representative results of the simulation study in Fig. 3, and the quantitative results in Table II. The proposed method outperforms all other methods in terms of the perceptual metrics: LPIPS, and mAP. As the penultimate goal of medical imaging is to provide an accurate diagnosis, high values of mAP indicate that our method is able to serve as a "good" denoiser, that boosts diagnosability. The metrics are consistent with the qualitative results shown in

TABLE II
QUANTITATIVE EVALUATION OF DENOISING ON FASTMRI KNEE VALIDATION DATA. **BOLD**: BEST, UNDER: SECOND BEST

Method	Gaussian ($\sigma = 13.0$)				Rician ($\sigma = 13.0$)			
	PSNR ↑	SSIM ↑	LPIPS ↓	mAP ↑	PSNR ↑	SSIM ↑	LPIPS ↓	mAP ↑
Proposed	31.74	0.872	**0.182**	**0.316**	31.13	**0.908**	**0.218**	**0.314**
SNLM [49]	30.82	0.852	0.304	0.292	29.98	0.871	0.306	0.254
BM3D [56]	30.35	0.860	0.318	0.280	31.01	0.896	0.322	0.263
N2Score [15]	30.08	0.852	0.333	0.245	28.41	0.801	0.333	0.252
SID [50]	30.01	0.856	0.327	0.250	29.06	0.792	0.319	0.249
RicianNet [51]	29.73	0.846	0.361	0.202	**31.31**	0.907	0.221	0.219
N2N [13]	**31.99**	**0.889**	0.193	0.309	25.78	0.675	0.401	0.186

TABLE III
QUANTITATIVE EVALUATION OF DENOISING+SR TASK ON FASTMRI KNEE VALIDATION DATA. **BOLD**: BEST, UNDER: SECOND BEST

Method	Gaussian ($\sigma = 13.0$)			
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR-LPR ↑
Proposed	**30.46**	**0.846**	**0.205**	**24.41**
SNLM [49]	27.63	0.622	0.359	19.03
BM3D [56]	27.38	0.639	0.331	19.97
N2Score [15]	29.61	0.686	0.301	21.56
SID [50]	27.51	0.610	0.329	19.95
RicianNet [51]	29.73	0.750	0.318	21.83
N2N [13]	29.87	0.778	0.297	21.75

Fig. 3, where we see that our method produces sharp and high-fidelity reconstructions while effectively removing the overall noise, regardless of the noise type or contrast. One thing to note is that the distortion metrics (i.e. PSNR, SSIM) of the proposed method often fall short against the SOTA methods in some cases (e.g. N2N has higher metrics in the Gaussian noise case). This is expected since MMSE denoisers are optimized specifically to minimize distortion, whereas the proposed method better matches the distribution. Moreover, one widely known issue of generative model-based methods is the hallucinatory pseudo-lesions, which is especially a big problem in medical imaging. Nonetheless, thanks to the well-controlled regularization of the proposed method, we do not observe such artifacts and see that our method provides faithful reconstructions that are close to ground truth. Namely, a board-certified radiologist with 15 years of experience was unable to spot pseudo-lesions across the whole dataset.

We perform an additional simulation study for denoising+SR, depicted in Fig. 4, with quantitative metrics shown in Table III. Here, we more clearly see the performance gap between the proposed method and the comparison methods. R2D2 performs accurate denoising + SR that closely mimics the ground truth. In contrast, the results of the comparison methods tend to be far blurrier, and often omitting important details. The further discrepancy is again mostly due to the inferior denoising performance. The denoised result is quite different from the training dataset of the SwinIR model. Subsequently, SwinIR underperforms on such out-of-distribution data, hence largening the gap. The superiority of the proposed method can also be seen in the LPR shown in the 3-6th row of Fig. 4, where we see that the computed residual with the proposed method closely matches the original perturbation. R2D2 is also the best-performing method in terms of PSNR-LPR, as can be seen in Table III.

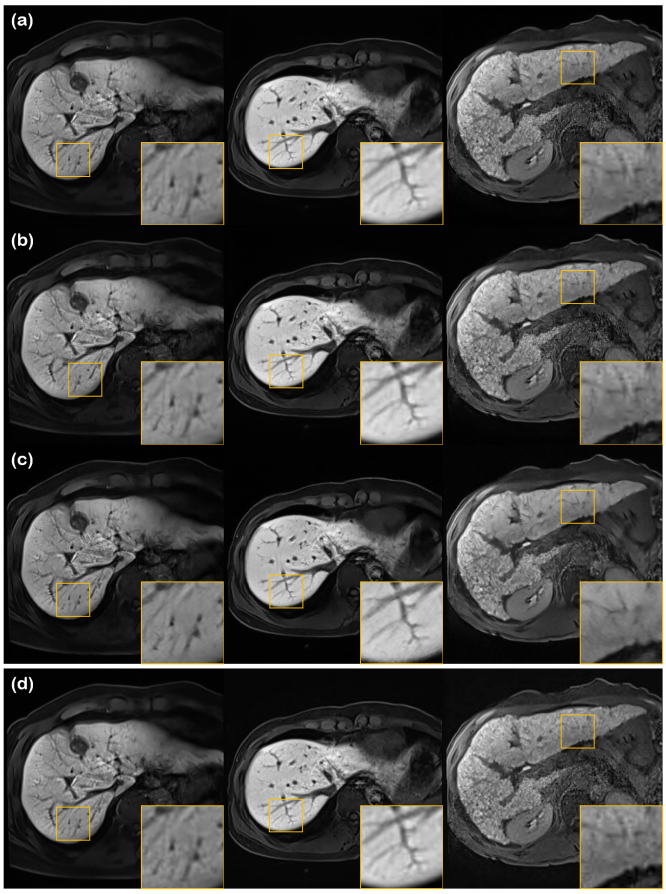

Fig. 5. Denoising results using various methods on different liver conditions. (a) N2N [13], (b) N2Score [15], (c) proposed method, and (d) input noisy image. Yellow boxes show results that are magnified. First row: early liver cirrhosis with hepatocellular carcinoma (HCC), second row: normal liver, third row: advanced liver cirrhosis.

B. In Vivo Study

Denoising results are summarized in Fig. 5, with the quantitative metrics in Table IV. In Fig. 5, we see that our method is the only method to effectively remove the noise while sharpening the image, regardless of the given condition of the liver. This is also consistent with the metrics given in Table IV, where our method is the highest scoring method both in terms of SNR and CNR. This can be deciphered in two aspects - 1) R2D2+ is the only method that explicitly enhances the resolution with the same network that was used to denoise the image. 2) Posterior sampling does not produce blurry output, as opposed to MMSE estimates.

From Table IV, we do observe that self-supervised learning-based methods (i.e., N2N, N2Score) have much better performance than BM3D/SNLM. Nevertheless, the metrics lack far behind R2D2+, especially in terms of CNR. This can also be visually inspected in Fig. 5 (b-d). While we can see relatively clearly that denoising has taken place as opposed to the noisy input, the denoised output tend to be washed out, with unclear boundaries and vessel structures. We note that Nei2Nei performs even worse than BM3D, which we conjecture to

TABLE IV
QUANTITATIVE EVALUATION OF SNR AND CNR WITH DIFFERENT METHODS (MEAN ± STD.) COMPUTED OVER 1312 TEST IMAGE SLICES. FOR ALL METHODS EXCEPT R2D2+, SR IS APPLIED VIA PRE-TRAINED SWINIR [52] MODEL

	SNR	CNR
Input	15.28 ± 5.47	<u>14.46</u> ± 4.95
SNLM [49]	16.49 ± 5.38	13.79 ± 6.21
BM3D [56]	16.38 ± 6.18	12.03 ± 7.95
N2N [13]	<u>19.38</u> ± 8.69	15.09 ± 1.31
SID [50]	19.01 ± 7.98	13.06 ± 8.21
N2Score [15]	19.28 ± 8.60	14.99 ± 8.72
Nei2Nei [38]	16.11 ± 5.57	12.08 ± 7.97
R2D2+	**21.28** ± 8.35	**16.08** ± 7.21

be stemming from its poor generalization capability. While SID [50] also uses an iterative denoising procedure with the same score function as ours, R2D2+ shows superior performance due to the advanced regularization techniques.

C. Flexibility and Uncertainty

As was discussed earlier, we can flexibly adjust the parameter α so that we can get just the right amount of denoising that we want. An illustration of such control is shown in Fig. 6, where we vary the value of α from 0.2 to 1.0.

In the first sample, it is clearly seen that we are able to achieve better results with higher values of α. With a lower factor, the noise is insufficiently removed. This is the case where noise is distributed in a holistic manner, as can be seen in the input image.

On the other hand, in the second sample, we see that with a high value of α, the texture and the important details of the liver are washed out, and hence it would be better to keep the denoising factor α at a low level. Notice that pulsation artifact [57] is apparent in the central region of the input image. In such cases, the periphery of the image is rather clean, whereas the central region of the image is contaminated with very high scale of noise. Hence, it can be seen that the distribution of noise is far from Gaussian, leading to an over-estimation of the noise variance. Therefore, lower values of α may be needed. In any case, being able to control the amount of denoising is a prominent feature of our proposed method, and would be of great importance in clinical settings.

Another important feature of the proposed method is the ability to quantify uncertainty. Denoising is an inverse problem, and hence there can be many feasible solutions. In order to embrace this fact, prior works have proposed several ways to estimate the uncertainty. For instance, [58] leveraged MC dropout [59]. Other works tried to estimate the variance directly [60]. In contrast, with the proposed method, we can simply sample multiple posterior samples, and directly compute the statistics from these samples. In order to show such property of R2D2+, we took 5 different posterior samples given the noisy image, and computed the mean and the standard deviation of the samples. The results are shown in the uncertainty maps of Fig. 6. The first sample does not depict any excessive variance in specific regions, and hence we can be fairly confident with the resulting denoised image.

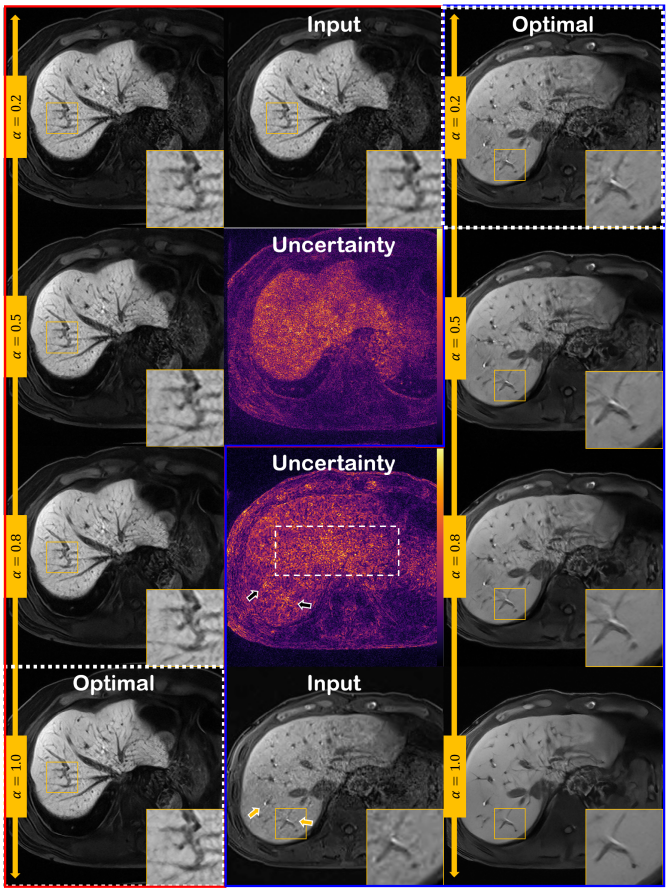

Fig. 6. Illustration of flexibility and uncertainty estimating property of the proposed method for the two patient data. (Red box, Blue box) In the second, input noisy image and uncertainty map (standard deviation of the posterior mean) are shown. The first and the third row illustrates denoised results with varying values of α. The arrows indicate regions of high uncertainty, and the white box with dotted line points out region with concentrated noise. White boxes with the dotted line indicate the values of α used to compute the posterior mean. Uncertainty maps were computed using the optimal α values indicated with the white dotted box.

Contrarily, we do observe some regions of higher variance in the second sample (marked with arrows), which may indicate that a clinician should be conservative on decision making based on the denoised samples.

Recall that our method involves numerically solving for the reverse SDE, and each discretization step involves addition of randomly sampled gaussian noise (z). Consequently, the more iterations we use (higher values of α), the more variation that it will lead to. In Fig. 7, we illustrate such effect, where we see that the uncertainty indeed increases for larger values of α. Thus, when the objective is to minimize the variation in the solutions of the inverse problem, one should conservatively choose the value of α.

VI. DISCUSSION AND RELATED WORKS

A. Impact of Regularization

First, we perform an ablation study on the two regularization schemes that we propose - MCG, and LF regularizer in Table V. We test the performance of denoising gaussian ($\sigma = 13.0$) noise on the simulation dataset, by removing

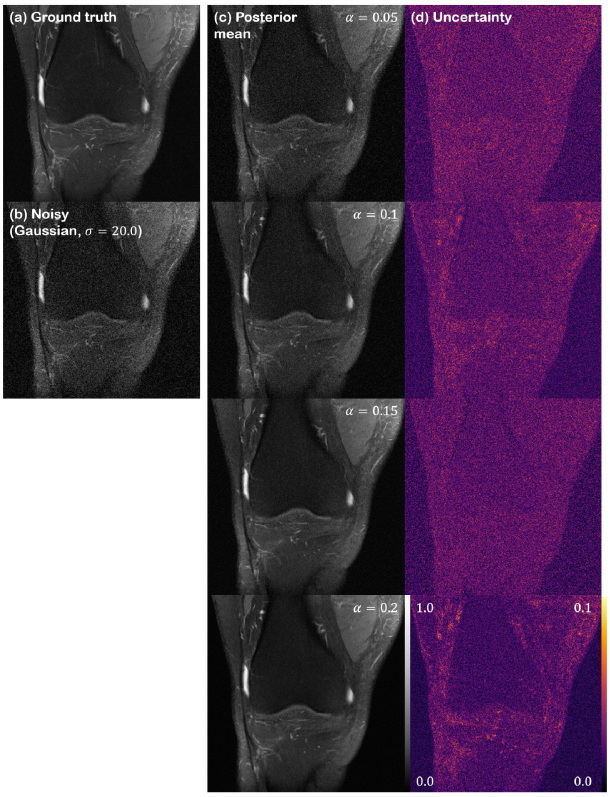

Fig. 7. Change of uncertainty with varying α values (5 posterior samples each). (a) Ground truth, (b) noisy image, (c) posterior mean, (d) standard deviation (uncertainty).

TABLE V
ABLATION STUDY ON THE TWO REGULARIZATION SCHEMES - MCG, LF

Configuration	PSNR	SSIM
MCG o, LF o	**31.74**	**0.872**
MCG o, LF x	31.26	0.864
MCG x, LF o	31.33	0.867
MCG x, LF x	29.61	0.848

each component. We achieve the highest metrics when we use both methods, and observe slightly degraded performance when using only one of the methods. When removing both components, we can see that the performance drops by a large margin, showing the efficacy of the proposed regularizations.

We show an example of the effect of low frequency regularization in Fig. 8. In the first column (see yellow arrow), we see that the vessel structure is altered when the regularization is not performed. Contrarily, the structure is conserved with the regularization present. Furthermore, in the blue arrow of the first column of Fig. 8, we observe some shaggy-looking structured artifact when we do not perform any regularization. Again, this artifact is eliminated as we impose our regularization. Consistent with the observation that was made in the first column, we again see altered structure in the second column, when regularization is not imposed. Thus, we can safely conclude that the proposed regularization is able to lead our reconstructions toward a conservative outcome, thereby eliminating false positives/negatives.

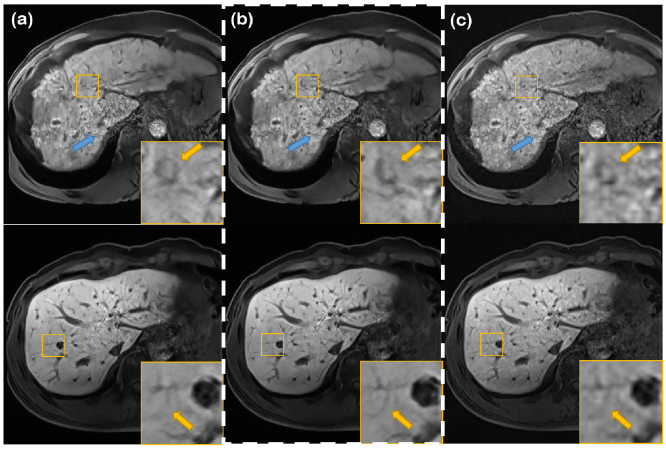

Fig. 8. Ablation study on low-frequency regularization method. (a) Without regularization, (b) with regularization, and (c) noisy image. First row: The alteration of the tortuous vascular structure might be misinterpreted as a round hepatic focal lesion by radiologists in clinical practice. Blue arrows indicate the region where structured shaggy artifact is present in the reconstruction without regularization. Second row: clearer vessel structure can be seen in the yellow arrows. White dotted lines indicate the results by the proposed method.

TABLE VI
PSNR VALUES BY VARYING THE LOW FREQUENCY REGULARIZATION PARAMETERS - λ (COLUMN), AND BANDWIDTH (ROW). **BOLD** INDICATES OUR CHOICE

	0.1	**0.05**	0.01	0.001	0.005
High	31.31	31.56	30.09	29.66	30.00
Mid	31.23	**31.58**	30.18	29.54	29.89
Low	31.03	31.02	30.02	29.69	29.70

TABLE VII
QUANTITATIVE EVALUATION OF SNR AND CNR WITH/ WITHOUT RESOLUTION ENHANCEMENT SCHEME

	SNR	CNR
SR ✓	21.28 $\pm$ 8.35	**16.08** $\pm$ 10.00
SR ✗	**22.53** $\pm$ 9.06	11.24 $\pm$ 7.32

In order to further investigate the effect of low frequency regularization, we check the performance of the regularized denoisers as a function of λ, and bandwidth of the mask Ω with the simulation fastMRI dataset with Gaussian noise ($\sigma = 13.0$). We search over three different bandwidths: High (5/8 cutoff), Mid (1/2 cutoff), and Low (3/8 cutoff), and also search over different values of λ. The results are reported in Table VI, where we see that the performance peaks at around $\lambda = 0.05$, when using Mid bandwidth. Since we also observe an increasing trend of PSNR values as we increase the λ value from 0 to 0.05, we again confirm that low-frequency regularization plays a crucial role.

B. Ablation Study on Super-Resolution

In Fig. 9, we compare the reconstruction results with and without using the post-hoc super-resolution method, with the corresponding metrics reported in Table VII. From the figure and the table, we see that the additional SR step greatly enhances the resolution, and the CNR, at the expense of a

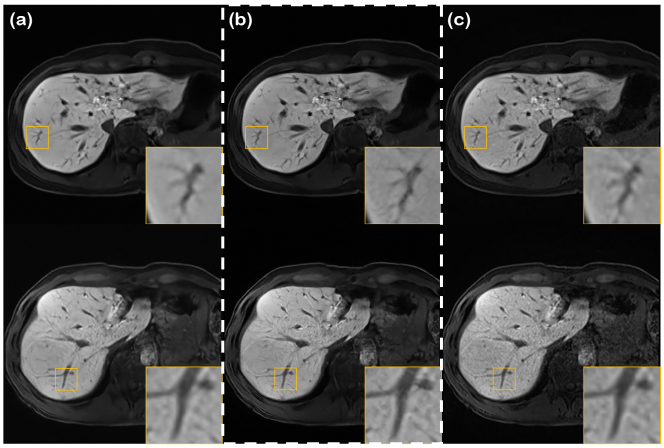

Fig. 9. Ablation study on the post-hoc super-resolution method. (a) Without SR, (b) with SR, and (c) noisy image. White dotted lines indicate the results by the proposed method.

slight decrease in SNR. Hence, we can safely conclude that the proposed method has an overall positive effect on the reconstruction.

C. Related Works

Denoising based on posterior sampling was discussed in two of the recent workshop papers [50], [61]. Ohayon et al. [61] trains a conditional GAN (cGAN), which is trained in a retrospective supervised manner. At the inference stage, one can sample multiple different latent vectors z, so that one would be able to acquire multiple posterior samples given the same noisy image. Kawar et al. [50] is the closest to our work, in that the authors propose to use a score-based generative model.[4] However, there are several contributions that our paper makes over [50]. First, the prior work [50] uses un-regularized denoising steps, whereas ours use regularized steps based on the theory of stochastic contraction [26]. We have already shown in Section VI-A the superiority in preserving the content. Second, our method is fully unsupervised in that we use heavily out-of-distribution open source data to train our network, and validate our method on real-world *in vivo* data. In opposition, [50] only tests their denoising method on well-controlled in-distribution datasets. Finally, we utilize a more advanced form of score matching and sampler, which was verified to have a large gap against the previous work [19].

D. Limitations and Future Directions

The iterative solver for the proposed method is based on first-order Euler discretization. Some of the recent works that leverage higher-order solvers for the integration of reverse SDE [34], [63] have shown promising improvements over the first order baselines. Extending our method to higher-order solvers may thus be a promising direction of research.

The mathematical foundation of diffusion models make them suitable for Gaussian noise removal, but not to other noise distributions. While our empirical observations show that the proposed method well extends to other distributions, there still exists model mismatch that has not been theoretically justified. The same holds for our low frequency regularization scheme. Albeit being effective, the motivation remains heuristic and further theoretical analysis on this method would be beneficial.

While this work contains extensive performance analysis both with the simulation study and in vivo study, the work does not contain in-depth human-involved perceptual analysis. It will be an interesting direction of future study to perform perceptual evaluation study conducted by radiologists, to establish the validity of the proposed method in actual clinical settings.

[4]Based on NCSNv2 [62].

VII. Conclusion

In this work, we proposed a denoising method for MRI, based on reverse diffusion and non-expansive mapping regularization. Along with it, we propose a method to super-resolve the denoised image with the *same* neural network. Our method achieves state-of-the-art both in terms of SNR and CNR, outperforming comparison methods by a large margin. With statistical analysis, we show that our method is the only method that can boost SNR such that it is statistically relevant, and at the same time boost the CNR. Furthermore, our method is able to quantify uncertainty in the solutions to the given inverse problem, while being flexible so that users can tweak how such noise should be eliminated, according to their needs. We believe that our research opens up an exciting new direction of denoising for medical imaging.

References

[1] J. Yang, J. Fan, D. Ai, S. Zhou, S. Tang, and Y. Wang, "Brain MR image denoising for Rician noise using pre-smooth non-local means filter," *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–20, Dec. 2015.
[2] J. V. Manjón et al., "MRI denoising using non-local means," *Med. Image Anal.*, vol. 12, no. 4, pp. 514–523, Aug. 2008.
[3] S. P. Awate and R. T. Whitaker, "Feature-preserving MRI denoising: A nonparametric empirical Bayes approach," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1242–1255, Sep. 2007.
[4] V. Hanchate and K. Joshi, "MRI denoising using BM3D equipped with noise invalidation denoising technique and VST for improved contrast," *Social Netw. Appl. Sci.*, vol. 2, no. 2, pp. 1–8, Feb. 2020.
[5] J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-Aron, J. Sijbers, and E. Fieremans, "Denoising of diffusion MRI using random matrix theory," *NeuroImage*, vol. 142, pp. 394–406, Nov. 2016.
[6] M. M. López, J. M. Frederick, and J. Ventura, "Evaluation of MRI denoising methods using unsupervised learning," *Frontiers Artif. Intell.*, vol. 4, p. 75, Jun. 2021.
[7] S. Fadnavis, J. Batson, and E. Garyfallidis, "Patch2Self: Denoising diffusion MRI with self-supervised learning," 2020, *arXiv:2011.01355*.
[8] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[10] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.
[11] 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.
[12] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.
[13] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data," in *Proc. ICML*, 2018, pp. 2971–2980.
[14] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2129–2137.

[15] K. Kim and J. C. Ye, "Noise2Score: Tweedie's approach to self-supervised image denoising without clean images," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 864–874.

[16] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[17] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.

[18] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[19] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *Proc. 9th Int. Conf. Learn. Represent. (ICLR)*, 2021, pp. 1–36.

[20] P. Dhariwal and A. Q. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds. Curran Associates, 2021, pp. 8780–8794.

[21] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Med. Image Anal.*, vol. 80, Aug. 2022, Art. no. 102479.

[22] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 14938–14954.

[23] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–18.

[24] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning method for denoising diffusion probabilistic models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 1–14.

[25] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," 2021, *arXiv:2104.07636*.

[26] H. Chung, B. Sim, and J. C. Ye, "Come-Closer-Diffuse-Faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12413–12422.

[27] J. Zbontar et al., "FastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.

[28] G. Chen, F. Zhu, and P. A. Heng, "An efficient statistical method for image noise level estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 477–485.

[29] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022pp. 1–28.

[30] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.

[31] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–13.

[32] B. D. O. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes Appl.*, vol. 12, no. 3, pp. 313–326, May 1982.

[33] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Comput.*, vol. 23, no. 7, pp. 1661–1674, 2011.

[34] A. Jolicoeur-Martineau, K. Li, R. Piché-Taillefer, T. Kachman, and I. Mitliagkas, "Gotta go fast when generating data with score-based models," 2021, *arXiv:2105.14080*.

[35] G. Parisi, "Correlation functions and computer simulations," *Nucl. Phys. B*, vol. 180, no. 3, pp. 378–384, May 1981.

[36] H. H. Bauschke et al., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408. New York, NY, USA: Springer, 2011.

[37] D. Kim, S. Shin, K. Song, W. Kang, and I.-C. Moon, "Score matching model for unbounded data score," 2021, *arXiv:2106.05527*.

[38] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2Neighbor: Self-supervised denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14781–14790.

[39] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, vol. 10. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[40] S. Pyatykh, J. Hesser, and L. Zheng, "Image noise level estimation by principal component analysis," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 687–699, Feb. 2012.

[41] W. Liu and W. Lin, "Additive white Gaussian noise level estimation in SVD domain for images," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 872–883, Mar. 2012.

[42] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[43] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, Dec. 1995.

[44] J. V. Manjón and P. Coupe, "MRI denoising using deep learning and non-local averaging," 2019, *arXiv:1911.04798*.

[45] M. H. Yu, J. M. Lee, J.-H. Yoon, B. Kiefer, J. K. Han, and B.-I. Choi, "Clinical application of controlled aliasing in parallel imaging results in a higher acceleration (CAIPIRINHA)-volumetric interpolated breath-hold (VIBE) sequence for gadoxetic acid-enhanced liver MR imaging," *J. Magn. Reson. Imag.*, vol. 38, no. 5, pp. 1020–1026, Nov. 2013.

[46] Y. Song, C. Durkan, I. Murray, and S. Ermon, "Maximum likelihood training of score-based diffusion models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 1415–1428.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.

[48] M. Tancik et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7537–7547.

[49] J. Hu, J. Zhou, and X. Wu, "Non-local MRI denoising using random sampling," *Magn. Reson. Imag.*, vol. 34, no. 7, pp. 990–999, Sep. 2016.

[50] B. Kawar, G. Vaksman, and M. Elad, "Stochastic image denoising by sampling from the posterior distribution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1866–1875.

[51] S. Li, J. Zhou, D. Liang, and Q. Liu, "MRI denoising using progressively distribution-based neural network," *Magn. Reson. Imag.*, vol. 71, pp. 55–68, Sep. 2020.

[52] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[53] R. Zhao et al., "FastMRI+: Clinical pathology annotations for knee and brain fully sampled multi-coil MRI data," 2021, *arXiv:2109.03812*.

[54] C. Chan and J. P. Haldar, "Local perturbation responses and checkerboard tests: Characterization tools for nonlinear MRI methods," *Magn. Reson. Med.*, vol. 86, no. 4, pp. 1873–1887, Oct. 2021.

[55] L. Sibulesky, "Normal liver anatomy," *Clin. Liver Disease*, vol. 2, no. S1, pp. S1–S3, Mar. 2013.

[56] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," in *Proc. SPIE*, Feb. 2006, Art. no. 606414.

[57] T. Budrys, V. Veikutis, S. Lukosevicius, R. Gleizniene, E. Monastyreckiene, and I. Kulakiene, "Artifacts in magnetic resonance imaging: How it can really affect diagnostic image quality and confuse clinical diagnosis?" *J. Vibroengineering*, vol. 20, no. 2, pp. 1202–1213, Mar. 2018.

[58] M.-H. Laves, M. Tölle, and T. Ortmaier, "Uncertainty estimation in medical image denoising with Bayesian deep image prior," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis*. Cham, Switzerland: Springer, 2020, pp. 81–96.

[59] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[60] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[61] G. Ohayon, T. Adrai, G. Vaksman, M. Elad, and P. Milanfar, "High perceptual quality image denoising with a posterior sampling CGAN," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1805–1813.

[62] Y. Song and S. Ermon, "Improved techniques for training score-based generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12438–12448.

[63] L. Liu, Y. Ren, Z. Lin, and Z. Zhao, "Pseudo numerical methods for diffusion models on manifolds," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–24.