arXiv:2305.01166v1 [cs.LG] 2 May 2023

利用从噪声数据中学习的基于分数的生成先验求解逆问题

Asad Aali[1], Marius Arvinte[1,2],Sidharth Kumar[1],Jonathan I. Tamir[1]德克萨斯大学奥斯汀分校钱德拉家族电气与计算机工程系，美国德克萨斯州奥斯汀[2]英特尔公司，美国俄勒冈州希尔斯伯勒 电子邮件： {asad.aali, arvinte, sidharth.kumar, jtamir}@utexas.edu

摘要—我们提出 SURE-Score，这是一种利用加性高斯噪声污染的训练样本来学习基于分数的生成模型的方法。近期研究表明，当有大量干净样本可供训练时，利用基于分数（扩散）的生成模型求解逆问题，其模型在完整采样数据分布上训练，效果优于端到端监督深度学习。在实践中，获取如此大规模的训练数据集，前期成本可能高得令人难以承受。本文提出一种方法，利用噪声训练数据近似学习干净数据分布的基于分数的生成模型。我们提出并论证了一种新颖的损失函数，该函数利用斯坦无偏风险估计，仅使用噪声样本，通过去噪分数匹配联合去噪数据并学习得分函数。我们通过学习先验并应用后验采样，求解来自不同领域的两项实际应用中的病态逆问题，以展示 SURE-Score 的通用性：压缩无线多输入多输出信道估计，以及加速二维多线圈磁共振成像重建。结果表明，在信噪比值分别为0 和10 dB 时进行学习，均可获得具有竞争力的重建性能。

索引词—**得分、扩散、生成、多输入多输出、磁共振成像、SURE、逆问题**

I. 引言

近年来，基于得分的（扩散）生成建模 [1], [2] 取得了显著进展，极大提升了在医学成像 [3]–[6], 、数字通信 [7], [8], 、图像超分辨率 [9], 等多个领域中，利用更少测量值并以更高重建保真度求解病态成像逆问题的能力。

然而，目前要为真实世界信号的分布学习高质量的基于分数的生成模型，仍假设可以获取大量完全采样且无噪声的训练样本 [10]。在许多应用领域，实际上不可能获得这样的训练集，因为用于采集测量数据的传感器不可避免地会受到噪声影响。例如，单线圈或多线圈磁共振成像（MRI）中始终存在噪声，这是由射频（RF）硬件中的热噪声以及人体造成的信号损耗所致。类似地，通信收发器中也存在噪声，尤其是在无线通信场景下，移动用户在成本和能耗受限的条件下使用设备。因此，仅使用噪声样本学习基于分数的生成模型是一项重要的研究课题。在这项工作中，我们通过将基于斯坦无偏风险估计（SURE）的去噪 [11] 与扩散模型训练中使用的去噪得分匹配（DSM）目标[12], 相结合，提出了这一问题的解决方案，从而利用单个深度学习模型实现联合去噪与得分学习。

我们的框架利用在适当噪声水平下的 Tweedie 单步去噪公式 [13] ，将这两个原本相似的目标联系起来。我们证明，该目标与 DSM 表述天然兼容，只需对训练函数进行简单修改，加入两个额外项和一个标量权重，即可用于训练数据受到独立同分布（i.i.d.）高斯噪声污染的情形。我们将学习到的模型用作先验，在两种不同应用中评估所提方法的有效性：压缩多输入多输出（MIMO）信道估计和加速MRI重建。结果表明，即使训练数据集是在信噪比（SNR）仅为 0 dB 的较差条件下采集的，利用所提方法学习到的先验仍可可靠地用于求解线性逆问题。

A. 相关工作

关于结构化信号去噪的研究已相当丰富；近期综述见 [14] 。与我们的工作最相关的是自监督 SURE 目标 [11], ，它仅利用噪声数据即可对监督去噪损失进行无偏估计。此前已有研究在独立的去噪场景中将 SURE 损失用作深度神经网络去噪器的训练目标 [15]–[17], ； 而将其与多层级去噪得分匹配联系起来，则是近期才开始探索的方向。[18], 研究了如何在受限的噪声水平子集上学习得分函数。

近来还有一些方法使用自监督学习，在没有真实值训练数据的情况下进行盲去噪。一个例子是 Noise2Noise [19],, 它使用同一样本的多个含噪测量值来训练深度神经去噪器。文献 [20] 进一步提出了 Noise2Self 框架；与我们提出的方法类似，该框架在训练集中只需要每个样本的一次含噪测量值。近期的文献 [17] 提出了

本研究得到 NSF IFML 2019844、NSF CCF-2239687（CAREER）以及 InterDigital 的资助。InterDigital 隶属于德克萨斯大学奥斯汀分校无线网络与通信组内的 6G@UT 中心。

Noise2Score 框架，用于在中高信噪比条件下自监督学习得分函数，但在低信噪比条件下如何学习得分函数仍未解决。更近期的文献 [21] 研究了如何从含噪的扩散磁共振成像样本中学习基于分数的生成模型，但仍需要获取同一受试者的多次含噪扫描。

预训练生成先验也已用于求解线性逆问题，且性能超越了经典的压缩感知方法 [22], [23]。近期一系列研究表明，在监督学习设定下，基于分数的生成模型用于求解加速MRI时能够取得有竞争力的重建性能 [3]–[6] ，并且可以通过 MMSE 去噪器隐式学习得分 [24]。文献 [7] 将相同思路用于从有限数量的导频测量中学习 MIMO 信道估计先验。近期研究 [25] 则探索了用于 MRI 重建的分数模型自监督学习方法：将测量数据划分为不同子集，其思路与 Noise2Self 类似。

B. 贡献

本文的贡献如下：

- 我们提出 SURE-Score：一种利用同一个深度神经网络联合学习去噪器和基于分数的生成模型的框架。该损失函数将 SURE、DSM 和 Tweedie 定则整合到同一个目标函数中。
- 我们通过从MIMO信道和磁共振成像的噪声数据中学习基于得分的生成模型，评估了SURE-Score。尽管这两个领域中的分布结构各不相同，我们的结果表明，即使信噪比为0 dB，SURE-Score也能学习到准确先验。
- 我们的仿真结果表明，使用在0 dB下采集的MIMO信道进行压缩信道估计，与完全监督的方法相比，归一化均方误差（NMSE）最多增加8 dB。我们还表明，使用在10 dB下采集的噪声样本学习先验后，可以实现自监督去噪和 $5\times$ 加速MRI重建，误差约增加15%。

我们用小写和大写字母分别表示向量（包括标量）和矩阵。我们用 p_X表示随机变量 X的概率分布函数（p.d.f.）。我们用x表示 X的展平版本，并用 X^{H} 表示 X 的厄米特共轭。除非另有说明，所有信号均假定为复数值。

II. 系统模型与背景

A. 线性逆问题

我们对受加性白高斯噪声污染的信号建模如下：

$$\tilde{x} = x + w, \tag{1}$$

其中 $x \in \mathbb{C}^N$ 从分布 p_X 中采样。噪声 $w \sim \mathcal{CN}(0,\ \sigma_w^2 I)$ 服从独立同分布的复值高斯分布，I 是单位矩阵，而 σ_w^2 是噪声功率。在线性逆问题的一般情况下，测量值如下：

$$y = Ax + n, \tag{2}$$

其中，$x \sim p_X$ 是真实信号，$y \in \mathbb{C}^M$ 是观测测量值，$n \sim \mathcal{CN}(0, \sigma_n^2 I)$ 是测量噪声，$A \in \mathbb{C}^{M\times N}$ 是前向算子。求解线性逆问题的目标是：给定测量值 y、前向算子 A 以及关于 x 和 n分布的统计知识，估计信号x。

B. 基于得分的生成建模

概率分布 p_X 的得分函数 ψ_X 定义为其对数相对于样本 [26]的梯度：

$$\psi_X(x) = \nabla_x \log p_X(x). \tag{3}$$

学习概率分布的得分函数可以避免归一化带来的挑战，并有助于下游任务，例如从学得的分布中采样或求解逆问题 [3]。基于得分的生成建模旨在给定从 p_X 中采样的训练数据集，学习得分函数 ψ_X。

去噪得分匹配 [12] 通过噪声条件得分网络 $s_\theta(x;\sigma)$ 学习一系列扰动信号分布的得分。当以噪声水平 σ为条件时，网络经过训练，用于预测实值信号 $\tilde{x} = x + z$经扰动后所得分布$p_{\tilde{X}}$ 的得分，其中 $z \sim \mathcal{N}(0, \sigma^2 I)$。DSM 损失定义如下：

$$L_{\mathrm{DSM},\theta}(x;\sigma) = \left\| s_\theta(\tilde{x};\sigma) - \nabla_{\tilde{x}} \log p_{\tilde{X}|X}(\tilde{x}|x;\sigma) \right\|_2^2. \tag{4}$$

当有无噪声数据可用于训练，并选择将添加的噪声 z 设为独立同分布的高斯分布时，条件分布的梯度可以表示为闭式形式 $-z/\sigma^2$。基于得分的生成模型通过在多个噪声水平上同时进行去噪得分匹配来训练 [1]。遵循 [27],, 我们使用 σ^2 对每个噪声水平下的损失加权，以归一化所有噪声水平下 DSM 损失的幅值。由此得到多级去噪得分匹配训练损失的表达式：

$$L_{\mathrm{DSM},\theta} = \mathbb{E}_{\substack{x\sim p_X \\ \sigma\sim p_\Sigma \\ z\sim\mathcal{N}(0,\sigma^2 I)}} \sigma^2 \left\| s_\theta(x+z;\sigma) + \frac{z}{\sigma^2} \right\|_2^2. \tag{5}$$

训练噪声水平 p_Σ 服从均匀分布，范围涵盖从 $\sigma_{\min}$ 到 $\sigma_{\max}$ 的几何级数分布噪声水平。我们遵循 [27] 中的指导原则，并将噪声条件得分网络实现为 $s_\theta(x;\sigma) = s_\theta/\sigma$，其中 $s_\theta(x)$ 是一个具有可学习权重 θ的深度神经网络。

C. 使用退火朗之万动力学求解逆问题

训练过程中学习到的基于分数的生成模型 s_θ 不使用式（2）设置中的任何线性算子 A 或接收到的测量值 y 信息，但仍然是解决线性逆问题的强大工具。我们用于生成点估计 x_{est} 的算法基于后验采样，并已在 [28] 中证明具有样本最优性

适用于线性逆问题。该随机算法从后验分布中抽取一个样本：

$$x_{\text{est}} \sim p_{X|Y}(x|y). \tag{6}$$

我们使用经过修改的退火朗之万动力学 [1]从后验分布中采样。这是一种迭代算法，从随机的 $x_0 \sim N(0,\ I)$ 开始，并且每一步都需要使用噪声条件得分函数 $\psi_{X_t|Y}(x_t|y;\sigma)$。为简洁起见，以下记法中省略 σ_t 。更新规则如下：

$$x_{t+1} = x_t + \alpha_t \cdot \psi_{X_t|Y}(x_t|y) + \sqrt{2\beta\alpha_t}\eta_t, \tag{7}$$

其中， α_t是随时间变化的步长， $\eta_t \sim \mathcal{N}(0,\ I)$ 是在每个更新步骤中采样并添加的退火噪声，而 β 是控制退火噪声水平的超参数。当$\beta = 1$时，式 (7) 中的算法就变为退火朗之万动力学。遵循 [27], ，我们按照 $\alpha_t = \alpha_0 \cdot (\sigma_t/\sigma_{\min})^2$，指数衰减步长，并通过两个超参数定义这一调度。项 σ_t 源自训练噪声水平的几何分布，该分布从 $\sigma_{\max}$ 开始，到 $\sigma_{\min}$结束。

条件得分函数 $\psi_{X_t|Y}$ 可通过贝叶斯规则进一步展开为：

$$\begin{aligned}\psi_{X_t|Y}(x_t|y) &= \nabla_{x_t} \log p_{X_t|Y}(x_t|y) \\ &= \nabla_{x_t} \log \frac{p_{Y|X_t}(y|x_t)p_{X_t}(x_t)}{p_Y(y)} \\ &= \psi_{Y|X_t}(y|x_t) + \psi_{X_t}(x_t),\end{aligned} \tag{8}$$

其中，分母项因不依赖于 x而消失。第一项与式（2）中逆问题的线性表述有关。当 $\sigma_t \to 0$时，该项变为：

$$\psi_{Y|X_t}(y|x_t) = \frac{A^{\mathrm{H}}(y - Ax_t)}{\sigma_n^2}, \tag{9}$$

其中， A^{H} 是 A的厄米特转置（伴随）。我们通过用因子 γ_t 进一步退火第一项，来近似式（8）中对依赖迭代的噪声水平 σ_t 的依赖，具体如 [3]所示：

$$\psi_{Y|X_t}(y|x_t) \approx \frac{A^{\mathrm{H}}(y - Ax_t)}{\sigma_n^2 + \gamma_t^2}, \tag{10}$$

本文中我们选择 $\gamma_t^2 = \sigma_t^2$。式（8）中第二项对 σ_t 的依赖由预训练的噪声条件得分模型 $s_\theta(x;\sigma_t)$纳入，该模型可以直接近似此项。从后验分布中采样的最终更新规则为：

$$x_{t+1} = x_t + \alpha_t \left(\frac{A^{\mathrm{H}}(y - Ax_t)}{\sigma_n^2 + \gamma_t^2} + s_\theta(x_t;\sigma_t) \right) + \sqrt{2\beta\alpha_t}\eta_t. \tag{11}$$

我们强调，我们的算法仅近似实现后验采样。

D. 斯坦无偏风险估计

从第 II-B 节可以清楚看出，通过去噪得分匹配学习先验需要一个干净的训练数据集 $x \sim p_X$。然而，在我们的设定中，我们只能获得一个由噪声样本构成的训练集 $\tilde{x}$ ，这些样本按照 (1) 进行采样。理想情况下，我们希望学习一个去噪模型 $g_\phi(x)$，其参数为 ϕ，使其能够将受扰样本 $\tilde{x}$ 映射为干净样本 x。一种方法是使用斯坦无偏风险估计 [29]，为每个样本x无偏估计真实的监督去噪损失 $\|g_\phi(\tilde{x}) - x\|_2^2$。当样本受到噪声水平为 σ_w的独立同分布高斯噪声污染时，SURE 表达式为：

$$L_{\text{SURE},\phi}(\tilde{x};\sigma_w) = \|\tilde{x} - g_\phi(\tilde{x})\|_2^2 + 2\sigma_w^2 \cdot \mathrm{div}_{\tilde{x}}(g_\phi(\tilde{x})), \tag{12}$$

其中，div$_x$($g_\phi(x)$)表示向量场（将输入映射为相同维度输出的任意函数） g_ϕ 在点 x处的散度，其定义为：

$$\mathrm{div}_x(g_\phi(x)) = \sum_i \frac{\partial [g_\phi(x)]_i}{\partial x_i}. \tag{13}$$

计算式（13）中的所有偏导数，需要进行与输入维度成正比的函数求值，这无法扩展到现实世界中的高维信号。按照 [30], ，我们使用蒙特卡洛近似来近似散度项，其中 $n \sim \mathcal{N}(0,\ I)$ 是为每个小批量中的每个样本采样的独立同分布高斯噪声，近似结果如下：

$$\mathrm{div}_x(g_\phi(x)) \approx n^T \left(\frac{g_\phi(x + \epsilon n) - g_\phi(x)}{\epsilon} \right), \tag{14}$$

当 ϵ 较小时。SURE 损失不涉及真实样本 x ，并且具有一个显著性质：对 $\tilde{x} \sim \mathcal{N}(n,\ \sigma_w^2 I)$ 求期望后满足 [11]:

$$\mathbb{E}_{\tilde{x}} L_{\text{SURE},\phi}(\tilde{x};\sigma_w) = \mathbb{E}_{\tilde{x}} \|g_\phi(\tilde{x}) - x\|_2^2, \tag{15}$$

因此，当只有带噪的训练样本集可用 [30],，且使用一个噪声样本小批量时，它可用作模型 g_θ 的训练函数。请注意，当数据集中的每个样本只有一个带噪版本时，也可以应用 SURE 损失，这与我们的设定相符。

III. 提出的方法：SURE-SCORE

我们提出的方法学习最小均方误差（MMSE）去噪样本的扰动分布得分。尽管这不同于在任意噪声水平下学习干净样本真实扰动分布的得分，但我们通过针对高斯扰动的 Tweedie 定则，将 SURE 学习与去噪得分匹配结合起来，使这两个目标在特定噪声水平下彼此一致。我们将该方法称为 SURE-Score。训练流程的框图如图 1 所示。

朴素地说，学习经 MMSE 去噪器处理后的扰动样本的得分，可以通过两个不同的函数来完成，每个函数都有自己的一组可学习参数，并依次应用：

$$L_{\text{SURE-Score},\theta}(\tilde{x};\sigma,\sigma_w) = \sigma^2 \left\| s_\theta(g_\theta(\tilde{x};\sigma_w)+z;\sigma) + \frac{z}{\sigma^2} \right\|_2^2 + \lambda \left\| \tilde{x} - g_\theta(\tilde{x};\sigma_w) \right\|_2^2 + 2\lambda\sigma_w^2 \cdot \text{div}_{\tilde{x}}(g_\theta(\tilde{x};\sigma_w)) \tag{16}$$

$$= \sigma^2 \left\| s_\theta(\tilde{x}+\sigma_w^2 s_\theta(\tilde{x};\sigma_w)+z;\sigma) + \frac{z}{\sigma^2} \right\|_2^2 + \lambda \left\| \sigma_w^2 s_\theta(\tilde{x};\sigma_w) \right\|_2^2 + 2\lambda\sigma_w^2 \cdot \text{div}_{\tilde{x}}(\tilde{x}+\sigma_w^2 s_\theta(\tilde{x};\sigma_w)). \tag{17}$$

1) 一个学习 MMSE 去噪器的函数 g_ϕ 。该函数可以仅使用含噪数据，通过第 II-D 节所述的 SURE 训练目标进行学习。

2) 一个基于分数的生成模型 s_θ ，在 g_ϕ 的输出上进行训练。如果有足够的训练数据，可以将其分配给两个阶段；否则，可以在第二阶段重复使用用于学习 g_ϕ 的同一训练集。

受 [17]，近期工作的启发，我们利用 Tweedie 定则 [13] 所揭示的两个步骤之间的内在联系，从而能够高效地将同一组网络参数用于这两种目的，并减少存储需求。假设已有一个基于分数的生成模型 $s_\theta(\tilde{x};\sigma)$，且该模型已完成训练，那么在加性高斯白噪声情况下，MMSE 去噪器可通过其 *Tweedie* 重新参数化表示为 [17]:

$$g_\theta(\tilde{x};\sigma) = \tilde{x} + \sigma^2 \cdot s_\theta(\tilde{x};\sigma). \tag{18}$$

反过来，若已获得训练完成的 MMSE 去噪器 $g_\theta(\tilde{x};\sigma)$，则其得分重新参数化形式为 [17]:

$$s_\theta(\tilde{x};\sigma) = \frac{g_\theta(\tilde{x};\sigma) - \tilde{x}}{\sigma^2}, \tag{19}$$

上述结论源于扰动后的 $\tilde{x}$ 分布具有有限均值和方差时 MMSE 估计器的存在性与唯一性，因此可以直接对 $\sigma > 0$ 应用 Tweedie 定则的逆变换。

前述恒等式表明，对于特定噪声水平，基于分数的生成模型与相同噪声水平下的 MMSE 去噪器可以互换。因此，我们提出在单一噪声水平 σ 下，使用以下损失函数对它们进行联合学习，训练数据则是在噪声水平 σ_w 下加入噪声的:

$$L_{\text{SURE-Score},\theta}(\tilde{x};\sigma,\sigma_w) = L_{\text{DSM},\theta}(g_\theta(\tilde{x};\sigma_w);\sigma) + \lambda \cdot L_{\text{SURE},\theta}(\tilde{x};\sigma_w), \tag{20}$$

其中， $\lambda > 0$ 是用于控制两个损失相对权重的超参数。请注意，在上述公式中，DSM 项使用去噪器 g_θ 的输出——这会导致对同一个底层网络进行循环调用。上述两项分别对应两个目标：第一项学习 MMSE 去噪样本扰动分布的基于分数的生成模型，第二项则学习 MMSE 去噪器本身。

Tweedie 重参数化在噪声水平 σ_w 下使两个目标保持一致，也就是说，当$\sigma = \sigma_w$ 在式（20）中时：在噪声水平 σ_w 下，对真实得分函数的近似越准确，去噪效果就越好，反之亦然。这可以视为近期提出的一致性目标的一种粗略

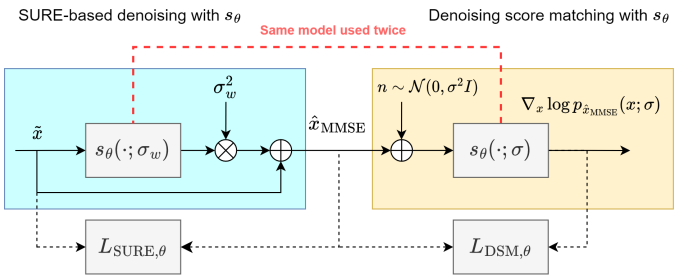


图 1. SURE-Score 在训练过程中的流程。同一个深度神经网络 s_θ 首先用于去噪，随后用于去噪得分匹配。(17) 式中的各项由输入、$\hat{x}_{\text{MMSE}}$ 和最终输出计算得到。

版本，见[31]，；该目标确保在噪声水平 σ_w 下去除噪声并重新添加噪声，能够使扰动得分函数的梯度得到一致的学习。在实践中，由于 MMSE 去噪存在非零残差， $\sigma = \sigma_w$ 处的两项均有界且非零。此外，上述一致性仅针对 $\sigma = \sigma_w$ 引入，但在实践中我们发现，这不会影响其他噪声水平下的性能，因为外推足够准确。

展开式 (20) 中两个损失的期望形式，可得到式 (16)。将 g_θ 替换为其 Tweedie 重参数化形式，可得到式 (17) 中的 SURE-Score损失函数，该函数仅依赖于评分网络 s_θ 和正则化λ。我们在实践中使用的训练损失是式 (17) 的小批量版本，其期望与式 (5) 中的期望相同。

请注意，尽管迄今为止的所有推导都基于这样的假设：$\tilde{x}$ 受到独立同分布高斯噪声的污染，但 SURE 损失和 Tweedie 重参数化都可以轻松扩展到其他指数族分布 [17], [32], ，因此我们提出的方法也很容易应用于其他常见的噪声模型。

最后，我们的训练和重建算法需要选择若干超参数，但在调参时无法获得真实值。具体而言，我们继承了与去噪得分匹配和 SURE 相关的所有超参数；其中值得注意的是 DSM 中由 p_Σ给出的步长调度，它已知取决于数据集 [27], ，以及蒙特卡洛近似 ϵ，它已知取决于噪声水平 σ_w [18], [30]。我们还必须选择 λ ，以平衡式 (17) 中的两个损失。

IV. 实验结果与讨论

我们在两个线性逆问题上测试了我们的方法：从压缩的

导频测量值中估计 MIMO 无线信道矩阵，以及根据有限的 k 空间测量值进行多线圈磁共振成像重建。我们将该方法与以下基线进行比较：(i) 监督得分模型训练，(ii) 在噪声数据上进行朴素 DSM 训练，(iii) BM3D，(iv) Noise2Self，以及 (v) Noise2Score。在 (iii-iv) 中，我们在去噪后的数据集上进行 DSM 训练。

我们对在不同信噪比水平（由 SNR^w 给出）的训练数据上学习得到的分数模型进行定性评估。我们还对后验重建性能进行定量评估。按照惯例，我们报告信道估计的归一化均方误差（NMSE，单位为 dB），以及磁共振成像的归一化均方根误差（NRMSE）；两者均针对特定的欠采样率以及不同的 σ_w^2 和 σ_n^2值进行评估。

A. 模型架构

我们采用基于 RefineNet 模型[33], 的深度神经网络，并遵循 [27][1]中的 NCSNv2 架构。为了处理这两种应用中的复值信号，我们将同一样本的实部和虚部分别视为两个实值通道。根据其他研究的建议 [16], ，我们在（14）中设置 $\epsilon = 10^{-3}$。所有实验均使用 Adam 优化器，学习率固定为 10^{-4}。（20）中的参数 λ 设为$L_{\text{DSM},\,\theta}/L_{\text{SURE},\,\theta}$，取值依据为训练开始时的第一个批次。

B. 多输入多输出信道建模与估计

我们考虑一种点对点多输入多输出基带通信场景，其中发射端和接收端分别配备 N_{t} 和 N_{r} 根天线。信道估计需要根据形式如下的测量值 Y 来估计信道矩阵 $H \in \mathbb{C}^{N_{\text{r}} \times N_{\text{t}}}$:

$$Y = HP + N, \tag{21}$$

其中，矩阵 $P \in \mathbb{C}^{N_{\text{t}} \times N_{\text{p}}}$ 是预先已知的测量矩阵（通常称为导频矩阵），而 N 是方差为 σ_{n}^2 的加性白高斯噪声。式（21）可以向量化为:

$$\underline{\text{y}} = \left(P^{\text{H}} \otimes I_{N_{\text{r}}}\right) \underline{\text{h}} + \underline{\text{n}}, \tag{22}$$

其中，$I_{N_{\text{r}}}$ 表示大小为 N_{r} 的方形单位矩阵。因此，该问题属于式（2）所描述的一般线性含噪逆问题。我们为 P 的各元素使用随机选取的正交相移键控（QPSK）信号。当导频密度比 $\alpha = N_{\text{p}}/N_{\text{t}} < 1$ 时，信道估计是一个欠定逆问题。在训练和推理过程中，我们通过除以 $\mathbb{E}_{H_{\text{train}}}[\|H\|_2^2]$ 对信道进行归一化。信道分布的信噪比定义为：$\mathbb{E}_H[\|H\|_2^2]/\sigma_w^2 = 1/\sigma_w^2$。

我们使用所提出的 SURE-Score 框架，在˜扰动后的簇状时延线（CDL）信道上训练了一个基于得分的模型 s_θ，其第一个模块包含 32 个隐藏层；具体而言，训练数据来自 CDL-C 非视距分布，信噪比为$w = 0$dB。我们为每种方法模拟了 10000 个（可能含噪的）样本用于训练，其中 $N_{\text{r}} = 16$ 和 $N_{\text{t}} = 64$。

[1]论文发表后，我们将在源代码中提供完整实现。

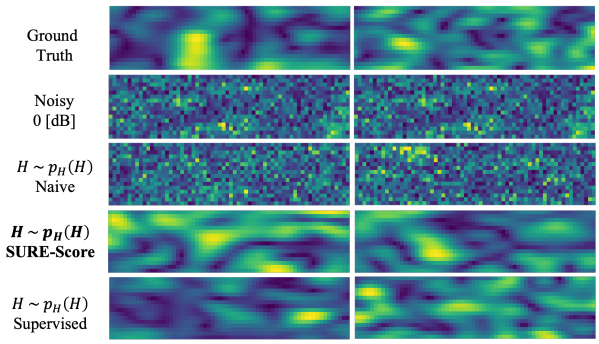


图 2. 三种方法的先验采样性能：朴素方法（ii）、信噪比为$w = 0$ dB 时的 SURE-Score，以及监督方法（i）。每列展示 CDL-C 信道的一种不同实现。

1) 先验采样：为了从学习到的先验 $p_{\hat{x}_{\text{MMSE}}}$ 中进行无条件采样，我们使用式（11）中的更新，并令 $A = 0$，从而消除依赖于测量的项。图 2 展示了来自训练集的 CDL-C 样本、信噪比为 w 的噪声样本，以及使用三种学习到的先验生成的样本：(i) 朴素 DSM；(ii) SURE-Score；(iii) 使用干净数据训练的 DSM。从定性结果来看，在噪声信道上以朴素方式训练的分数模型生成的先验样本受到了破坏。然而，监督样本和 SURE-Score 样本在定性上都与参考样本中的信号结构相匹配。我们的后验采样实验进一步验证了这一点。

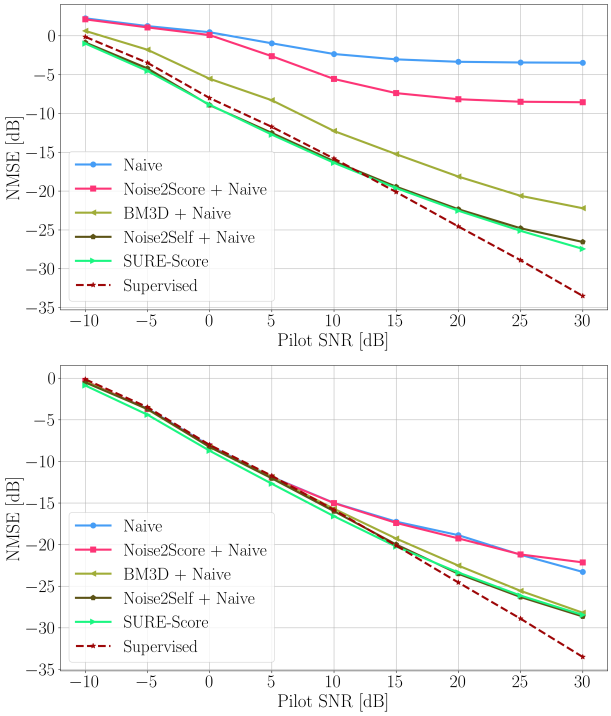


图 3. 使用在 CDL-C 信道上训练的分数模型进行信道估计的性能，信噪比为w: 0 dB（上）和 10 dB（下）； $\alpha = 0.6$ （38 个导频）。

2) 后验重建：我们使用式（11）中的退火朗之万动力学更新，并采用基于式（22）的数据一致性项。我们模拟的试验密度为 $\alpha = 0.6$，即 38 个导频。

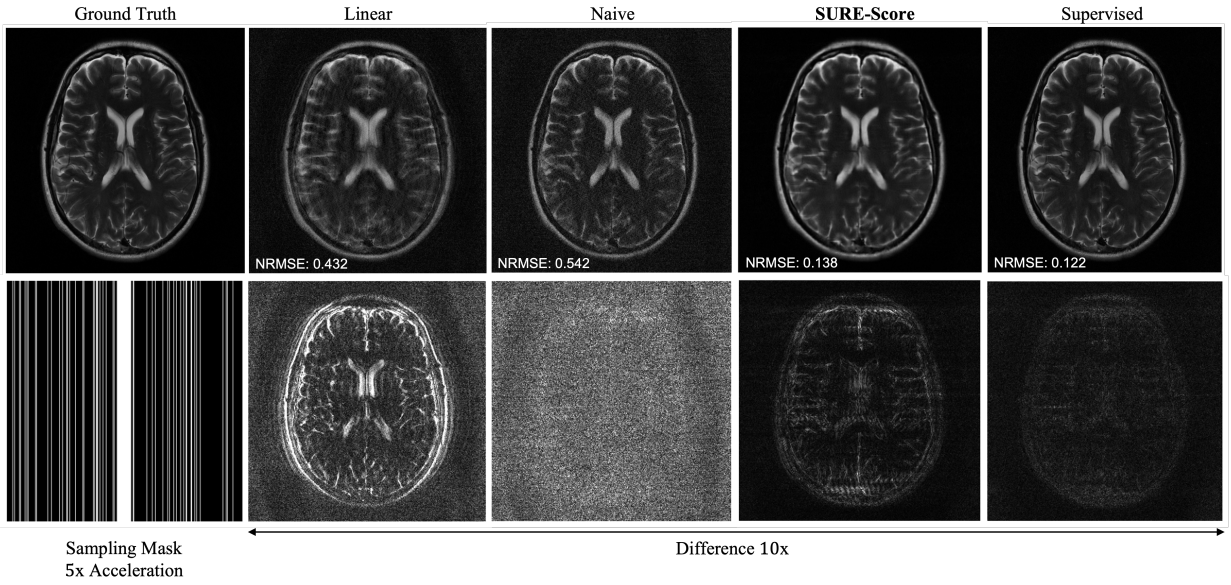


图 4. 加速因子为 $5\times$ 时的多线圈磁共振成像重建。从左到右依次为：全采样真实值、线性重建、在信噪比为^{w}的噪声数据上进行朴素训练后的后验采样、使用 SURE-Score 训练后的后验采样，以及使用无噪声数据训练后的后验采样。底行依次显示各方法的采样模式和差异图像。

图 3 总结了使用不同训练方法、在不同导频信噪比下进行后验采样的结果，˜其中我们假设两个图中的训练样本 H 分别含有信噪比为 0 dB 和 10 dB 的噪声。我们观察到，朴素训练的估计性能会趋于停滞，这是由于对含噪训练集过拟合所致。Noise2Score 和 BM3D 在较低信噪比下表现较差，信噪比提高时性能有所改善，这与文献中的报告一致 [17]。除了导频信噪比较高时以外，SURE-Score 的表现接近监督 DSM。作为去噪预处理步骤的 Noise2Self 也表现良好。我们观察到，在导频信噪比较高时存在性能差距，这可能是由于针对特定架构学习得到的 MMSE 去噪器性能有限，以及训练数据量有限。

C. 多线圈磁共振成像重建

多线圈磁共振成像数据通过在成像解剖结构周围放置多个射频线圈，在频域中采集。我们假设一个向量化图像 $x \in \mathbb{C}^N$，其中有 N_{c} 个射频接收灵敏度分布，每个分布均由一个对角矩阵 $S_i \in \mathbb{C}^{N\times N}$ 表示；傅里叶采样为 $F_\alpha \in \mathbb{C}^{\alpha N\times N}$。来自第 i^{th} 个线圈的采集 k 空间数据为：

$$y_i = F_\alpha S_i x + n_i, \tag{23}$$

其中，n_i 是方差为 σ_w^2 的加性白高斯噪声。按照惯例，我们定义相对于单个全采样线圈的加速因子 $1/\alpha$，因此采集到的测量值总数为 $\alpha \cdot N_{\mathrm{c}} \cdot N$。我们可以根据式 (2) 表示磁共振成像多线圈前向模型。

我们使用 fastMRI 数据集中的 10000 次多线圈 T2加权脑部磁共振成像扫描 [34]，训练第一个模块包含 128 个隐藏层的基于分数的模型。SURE-Score 模型由于训练时间有限，该模型仅使用部分数据（2000 个样本）进行训练。我们首先使用来自大小为 24×24 的中心窗口的k空间数据进行平方和开根号（RSS）重建，从而对训练集中的所有层面进行归一化。然后，将每个层面的 k空间数据除以该低分辨率重建的 95^{th}幅值百分位数。我们将信噪比定义为 $1/\sigma_w^2$。

1) 去噪性能： 我们使用 (18) 中的 Tweedie 重参数化，评估经过 (i) 朴素训练、(ii) 使用 SURE-Score 训练以及 (iii) 使用无噪声数据（监督）训练的分数模型的去噪性能。表 I 列出了 SNR^w 为 0 dB 和 10 dB 时，100 个验证切片上的归一化均方根误差（均值和标准差）。SURE-Score 的去噪性能几乎与监督学习相当，表明在 σ_w 处 SURE 目标与分数匹配目标之间具有一致性。因此，在多个层面进行去噪得分匹配不会降低 σ_w处的去噪性能。

去噪性能（归一化均方根误差）		
SNR^w	0 dB	10 分贝
朴素	2.48 ± 0.24	0.70 ± 0.07
监督	0.21 ± 0.01	0.14 ± 0.01
SURE-Score	0.23 ± 0.01	0.16 ± 0.01

表 I：使用受噪声污染的数据和不同目标函数训练的分数模型的磁共振成像去噪性能。

2）后验重建：我们使用文献（7）中所述的退火朗之万动力学。我们使用垂直采样掩模模拟五倍加速因子，其中中心 k 空间完全采样，其余部分进行均匀随机欠采样。我们假设信噪比较高，并在推理过程中设置 $\sigma_n = 0$。

表 II 展示了在包含 50 个层面的验证集上的重建归一化均方根误差（均值和标准差）。我们比较

线性重建、朴素训练、监督训练和 SURE-Score。我们观察到的趋势与多输入多输出实验中的相似。图 4 比较了在信噪比为 ^{w} 的噪声数据上训练后得到的重建示例。朴素训练会导致噪声放大，而 SURE-Score 在定性和定量方面均有所改善。

重建性能（归一化均方根误差）		
SNR^{w}	0 dB	10 dB
线性	0.53 ± 0.19	0.53 ± 0.19
朴素	1.42 ± 0.35	0.71 ± 0.15
监督	0.16 ± 0.05	0.16 ± 0.05
SURE-Score	0.22 ± 0.06	0.19 ± 0.05

表 II：fastMRI 验证集上的重建归一化均方根误差（均值和标准差）比较。

V. 结论

通过实验结果，我们观察到基于 SURE- 的去噪对先验采样以及求解无线通信和磁共振领域逆问题的影响；在此过程中，模型从噪声信号比高达 0 dB 的受污染数据集中学习。我们还发现，SURE-Score 等自监督技术能够达到与监督去噪相当的性能。尽管我们的算法在训练轮次方面与监督分数模型的收敛速度相当，但由于每一步都需要额外进行一次网络前向计算，我们发现每次迭代的运行时间有所增加。与其他自监督方法一样，另一个挑战是在无法获取真实数据的情况下选择超参数。最后，我们目前的工作假设噪声污染为白高斯噪声，但可借助广义 SURE 将其扩展至任意指数族 [32]。

参考文献

[1] Y. Song 和 S. Ermon，“通过估计数据分布的梯度进行生成建模”，神经信息处理系统进展，第 32 卷，2019 年。[2] J. Ho、A. Jain 和 P. Abbeel，“去噪扩散概率模型”，神经信息处理系统进展，第 33 卷，第 6840–6851 页，2020 年。[3] A. Jalal、M. Arvinte、G. Daras、E. Price、A. G. Dimakis 和 J. Tamir，“利用深度生成先验进行鲁棒压缩感知MRI”，神经信息处理系统进展，第 34 卷，2021 年，第 14 938–14 954 页。[4] Y. Song、L. Shen、L. Xing 和 S. Ermon，“利用基于分数的生成模型解决医学成像中的逆问题”，*arXiv*预印本 *arXiv:2111.08005*，2021 年。[5] H. Chung 和 J. C. Ye，“用于加速MRI的基于分数的扩散模型”，医学图像分析，第 80 卷，第 102479 页，2022 年。[6] G. Luo、M. Blumenthal、M. Heide 和 M. Uecker，“使用扩散模型进行联合不确定性估计的贝叶斯MRI重建”，磁共振医学，2023 年。[7] M. Arvinte 和 J. I. Tamir，“使用基于分数的生成模型进行MIMO信道估计”，*IEEE*无线通信汇刊，2022 年。[8] N. Zilberstein、C. Dick、R. Doost-Mohammady、A. Sabharwal 和 S. Segarra，“大规模MIMO检测的退火朗之万动力学”，*IEEE*无线通信汇刊，2022 年。[9] B. Kawar、M. Elad、S. Ermon 和 J. Song，“去噪扩散复原模型”，*arXiv*预印本 *arXiv:2201.11793*，2022 年。[10] F. Koehler、A. Heckett 和 A. Risteski，“分数匹配的统计效率：等周不等式视角”，*arXiv*预印本 *arXiv:2210.00726*，2022 年。[11] C. M. Stein，“多元正态分布均值的估计”，统计年刊，第 1135–1151 页，1981 年。[12] P. Vincent，“分数匹配与去噪自编码器之间的联系”，神经计算，第 23 卷，第 7 期，第 1661–1674 页，2011 年。

[13] B. Efron，“Tweedie公式与选择偏差”，《美国统计学会杂志》，第 106 卷，第 496 期，第 1602–1614 页，2011 年。[14] M. Elad 、 B. Kawar 和 G. Vaksman，“图像去噪：深度学习革命及其后续发展——综述论文——”，arXiv预印本 arXiv:2301.03362，2023 年。[15] S. Soltanayev 和 S. Y. Chun，“无需真实数据训练基于深度学习的去噪器”，神经信息处理系统进展，第 31 卷，2018 年。[16] C. A. Metzler、A. Mousavi、R. Heckel 和 R. G. Baraniuk，“使用Stein无偏风险估计量的无监督学习”，arXiv预印本 arXiv:1805.10531，2018 年。[17] K. Kim 和 J. C. Ye，“Noise2Score：Tweedie方法用于无需干净图像的自监督图像去噪”，神经信息处理系统进展，第 34 卷，第 864–874 页，2021 年。[18] M. Arvinte 和 J. Tamir，“用于无线信道建模和估计的基于分数的生成模型”，载于《ICLR高度结构化数据深度生成模型研讨会》，2022 年。[19] J. Lehtinen、J. Munkberg、J. Hasselgren、S. Laine、T. Karras、M. Aittala 和 T. Aila，“Noise2noise：无需干净数据学习图像复原”，载于《国际机器学习会议》。PMLR，2018 年，第 2965–2974 页。[20] J. Batson 和 L. Royer，“Noise2self：通过自监督进行盲去噪”，载于《国际机器学习会议》。PMLR，2019 年，第 524–533 页。[21] T. Xiang、M. Yurt、A. B. Syed、K. Setsompop 和 A. Chaudhari，“Ddm2：利用生成式扩散模型进行自监督扩散MRI去噪”，arXiv预印本 arXiv:2302.03018，2023 年。[22] M. Lustig、D. L. Donoho、J. M. Santos 和 J. M. Pauly，“压缩感知MRI”，IEEE信号处理杂志，第 25 卷，第 2 期，第 72–82 页，2008 年。[23] A. Alkhateeb、G. Leus 和 R. W. Heath，“基于压缩感知的多用户毫米波系统：需要多少测量值？”，载于《2015年IEEE国际声学、语音与信号处理会议（ICASSP）》。IEEE，2015 年，第 2909–2913 页。[24] Z. Kadkhodaie 和 E. Simoncelli，“利用去噪器中隐含的先验，为线性逆问题求取随机解”，载于《神经信息处理系统进展》，M. Ranzato、A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，第 34 卷。Curran Associates, Inc.，2021 年，第 13 242–13 254 页。[25] Z.-X. Cui、C. Cao、S. Liu、Q. Zhu、J. Cheng、H. Wang、Y. Zhu 和 D. Liang，“Self-score：基于分数模型的MRI重建自监督学习”，arXiv预印本 arXiv:2209.00835，2022 年。[26] A. Hyv¨arinen 和 P. Dayan，“通过分数匹配估计非归一化统计模型”。机器学习研究杂志，第 6 卷，第 4 期，2005 年。[27] Y. Song 和 S. Ermon，“训练基于分数的生成模型的改进技术”，神经信息处理系统进展，第 33 卷，第 12 438–12 448 页，2020 年。[28] A. Jalal、S. Karmalkar、A. Dimakis 和 E. Price，“通过后验采样实现实例最优压缩感知”，载于《国际机器学习会议》。PMLR，2021 年，第 4709–4720 页。[29] K.-C. Li，“从Stein无偏风险估计到广义交叉验证方法”，统计年刊，第 1352–1377 页，1985 年。[30] S. Ramani、T. Blu 和 M. Unser，“蒙特卡洛SURE：针对一般去噪算法的正则化参数黑盒优化”，IEEE图像处理汇刊，第 17 卷，第 1540–1554 页，2008 年。[31] G. Daras、Y. Dagan、A. G. Dimakis 和 C. Daskalakis，“一致性扩散模型：通过学习保持一致来缓解采样漂移”，arXiv预印本 arXiv:2302.09057，2023 年。[32] Y. C. Eldar，“指数族的广义SURE：正则化应用”，IEEE信号处理汇刊，第 57 卷，第 2 期，第 471–481 页，2008 年。[33] G. Lin、A. Milan、C. Shen 和 I. Reid，“RefineNet：用于高分辨率语义分割的多路径细化网络”，载于《IEEE计算机视觉与模式识别会议论文集》，2017 年，第 1925–1934 页。[34] J. Zbontar、F. Knoll、A. Sriram、T. Murrell、Z. Huang、M. J. Muckley、A. Defazio、R. Stern、P. Johnson、M. Bruno 等，“fastMRI：用于加速MRI的开放数据集和基准测试”，arXiv预印本 arXiv:1811.08839，2018 年。