arXiv:2304.01963v1 [eess.IV] 4 Apr 2023

用于快速有限视角光声断层成像的模型校正学习型原始-对偶模型

Andreas Hauptmann*IEEE* 高级会员和 Jenni Poimala

摘要—**学习型迭代重建在加速断层成像方面大有可为，并且对模型扰动具有经验证的鲁棒性。然而，由于需要反复评估计算开销巨大的前向模型，这种方法在光声断层成像中的应用受到阻碍。采用快速近似模型可以提高计算可行性，但也因此需要补偿模型误差。在本研究中，我们将模型校正嵌入学习型原始-对偶框架，从而推进学习型图像重建中模型校正的方法与理论基础。在展开式端到端学习型迭代重建方法中，数据空间中的模型校正与图像空间中的学习型更新算子联合学习。所提出的公式还可扩展为原始-对偶深度平衡模型，实现不动点收敛，同时降低训练所需的内存。我们通过理论分析和实证研究深入探讨所提出的模型，并在真实的二维有限视角设置中进行数值验证。模型校正的学习型原始-对偶方法具有出色的重建质量和快速的推理速度，为光声断层成像中可实时运行且可扩展的迭代重建奠定了方法学基础。**

索引词——**学习型图像重建、逆问题、光声断层成像、模型校正、原始-对偶方法、有限视角。**

I. 引言

LEARNED 迭代重建因能够结合模型信息与学习组件，在对模型扰动具有鲁棒性的同时取得最先进的结果而广受欢迎 [1]–[4]。尤其是所谓的端到端方法，它们将迭代方案展开为固定次数的迭代，并联合训练所有迭代点，在许多应用中都能提供更优的重建质量。然而，这类端到端方法的适用范围在很大程度上局限于图像尺寸较小、训练和评估可行的二维场景。对于高分辨率成像，乃至三维数据，我们会遇到若干关键难题。首先，内存需求会随展开迭代的数量线性增长；其次，对模型方程进行计算开销高昂的评估，可能会严重限制合理的训练时间。在光声断层成像中尤其如此，因为其前向模型由声波方程给出。此外，许多光声测量配置采用有限视角几何，使逆问题严重不适定，因此重建过程需要先进的重建方法和强先验。

一类适用的先进重建方法是原始-对偶方法，这类方法在经典的基于模型的成像场景中已取得显著成效 [5], [6], 其学习型版本尤其能提供当前最先进的结果 [7]。在图像（原始变量）空间和数据（对偶变量）空间中结合学习型更新尤为有效，但也需要以端到端方式进行联合训练，因此又面临适应高维问题的主要局限。尤其是，由于引入了测量空间中的更新网络，内存需求进一步增加。

近期已有若干进展，在不同情境下克服这些计算限制，大致遵循两种范式：

（i）将网络训练与算子评估解耦。（ii）改进学习型迭代方法，以减轻高计算需求的影响。

第一类方法包括所谓的即插即用方法：离线训练去噪器，随后在交替方向乘子法或近端梯度下降等迭代方案中使用该去噪器，以取代近端算子 [4],[8]–[11]。另一种方法是贪心训练，它要求逐次迭代达到最优，而不是对展开的迭代过程进行端到端训练 [12], [13],，从而能够应用于三维有限视角光声断层成像 [12]。遗憾的是，学习型原始-对偶方法的联合特性使其无法采用贪心训练策略。

第二类方法更加多样，因为不同方法可能只针对某个特定子问题。例如， [14]–[16] 使用了一种近似模型，以加快模型的训练和部署。深度平衡模型（DEQ） [17], [18], 旨在将展开算法重新表述为不动点迭代，并且仅对不动点方程求导，从而降低内存消耗，使其不受迭代点数量的影响。另一项研究 [19]旨在通过将展开算法表述为多尺度方法，同时减少内存消耗和算子求值次数，从而降低前向算子的计算成本以及较低尺度上的内存消耗。最后，可逆神经网络 [20]–[22],可用于消除存储中间

本研究得到芬兰科学院项目 338408、346574 和 353093 的支持。作者感谢艾萨克·牛顿数学科学研究所（剑桥）在项目 丰富与非线性断层成像 期间提供的支持与接待；本文的相关研究工作即在该项目期间开展。该项目由英国工程与自然科学研究委员会资助，资助编号为 EP/R014604/1。JP 和 AH 任职于芬兰奥卢大学数学科学研究组。AH 同时任职于英国伦敦大学学院计算机科学系。

状态以进行反向传播的需要，但在展开算法中，仍需计算前向算子。

除了上述计算方面的进展，近期人们也越来越努力地为学习型图像重建提供理论结果 [3], [10], [23], [24], ，例如以展开方法的收敛结果为形式。其中一种收敛保证针对 DEQ 模型给出，要求更新算子为压缩映射，从而收敛到一个不动点 [17], [18]。虽然可以满足更强的收敛概念，但通常需要对学习组件施加更严格的约束，这往往会降低网络的表达能力和性能。关于可证明收敛的学习型图像重建方法综述，请参阅 [3] 。

受上述进展的启发，本文旨在构建一种模型校正的学习型原始-对偶模型，该模型可利用快速近似前向算子进行端到端训练，以实现高分辨率重建。此外，我们在深度平衡形式下提供关于不动点收敛的理论见解和条件。为此，我们从经典的基于模型的原始-对偶混合梯度法 [5]出发，将计算代价高昂的前向算子和伴随算子替换为快速但近似的版本 [14]。受 [15],启发，我们在对偶空间中通过算子校正来修正引入的误差，并在原始空间中采用学习型更新算子。由此得到的算法称为模型校正学习型原始-对偶（MC-PD）。随后，我们将模型校正的学习型原始-对偶模型构建为深度平衡模型（PD-DEQ），并分析参数和网络需满足的条件，以确保原始变量上的压缩映射。我们的主要结果表明，在对偶变量满足充分下降假设的条件下，可以实现不动点收敛。

本文对所提出的方法在二维有限视角场景下进行了全面评估。我们将其与采用 U-Net 的后处理方法，以及所提出的 MC-PD 的一个不使用权重共享的版本进行比较；后者因此接近原始的学习型原始-对偶方法。我们从内存消耗和评估时间两方面评估了方法扩展到更大图像尺寸时的可扩展性。结果表明，MC-PD 无论采用还是不采用权重共享，都能取得同样出色的性能。DEQ模型的表现不及 MC-PD，峰值信噪比和结构相似性指数有所下降，但仍能提供良好的定性结果。此外，我们讨论了在所研究的有限视角设置下训练 DEQ模型时面临的挑战，并通过计算验证了更新网络的压缩映射性质。最后，我们将提供模型的代码和实现。

本文结构如下：在第 II 节中，我们介绍光声断层成像的主要概念以及重建的变分形式。随后，我们回顾学习型重建方法，并介绍本研究中使用的快速近似算子。在第 III 节中，我们介绍模型校正学习型原始 - 对偶及其深度平衡形式。我们为所提出的模型提供一些理论见解，随后推导关于压缩性质的条件。第 IV 节讨论计算方面和数据模拟。我们在第 V 节中展示结果，并讨论所提出模型的性能与挑战。最后，第 VI 节给出最终结论。

II. 光声断层成像与先前研究

在光声断层成像（PAT）中，生物组织中的发色团会吸收一束短脉冲近红外光[25], [26]。对于足够短的脉冲，这将在组织内部引起空间变化的压力升高 x ，从而激发一个超声（US）脉冲（光声效应），该脉冲传播至组织表面。在本研究中，我们仅考虑光声断层成像中的后一个声学问题。也就是说，声传播的前向模型由声波方程的一个初值问题建模 [27],[28]

$$(\partial_{tt} - c^2\Delta)p(\mathbf{x},t) = 0,\ p(\mathbf{x},t_0) = x(\mathbf{x}),\ \partial_t p(\mathbf{x},t_0) = 0, \tag{1}$$

其中 $\mathbf{x} \in \mathbb{R}^2$ 和 $t_0 = 0$ 表示初始时刻。我们在有限时间窗内测量计算域 Ω 边界上的压力场。随后，可将测量数据建模为作用于压力场$p(\mathbf{x}\ t)$的线性算子 $\mathcal{M}$ ：

$$y = \mathcal{M}\, p_{|\partial\Omega\times(0,T)}. \tag{2}$$

声传播 (1) 和测量算子 (2) 定义了线性前向模型

$$Ax = y \tag{3}$$

从初始压力 $x \in X$ 到测量时间序列 $y \in Y$。这代表理想且准确的前向模型，可以通过例如 [29], [30] 中概述的伪谱时间步进模型进行模拟。

如果测量数据稀疏，或者检测几何仅覆盖边界的一部分，逆问题就会变得严重病态 [31], ，例如本文考虑的有限视角场景。测量装置示意图见图1。这就需要正则化和先进的先验来弥补信息损失，这可以在变分框架中表述 [32]–[34]。也就是说，给定测量数据 y ，我们将重建结果作为下式的极小化解：

$$J(x) = \|Ax - y\|_2^2 + \lambda R(x), \tag{4}$$

其中，第一项衡量数据拟合项，第二项对问题进行正则化，并纳入关于目标对象的先验知识； $\lambda > 0$ 则平衡这两项的影响。随后，可以通过合适的优化算法迭代求解式 (4)。例如，采用一种适用于不可微项$R(x)$的近端梯度法：给定初始重建 x_0 ，可以通过以下更新方程迭代最小化式 (4)：

$$x_{k+1} = \mathrm{prox}_{R,\lambda\gamma_k}\left(x_k - \gamma_k A^*(Ax_k - y)\right), \tag{5}$$

采用合适的步长 $\gamma_k > 0$。近端算子通过求解相应的去噪问题，将迭代点投影到满足正则化项 R要求的解上

$$\mathrm{prox}_{R,\alpha}(x) = \arg\min_u \left\{ R(u) + \frac{1}{2\alpha}\|u - x\|_2^2 \right\}. \tag{6}$$

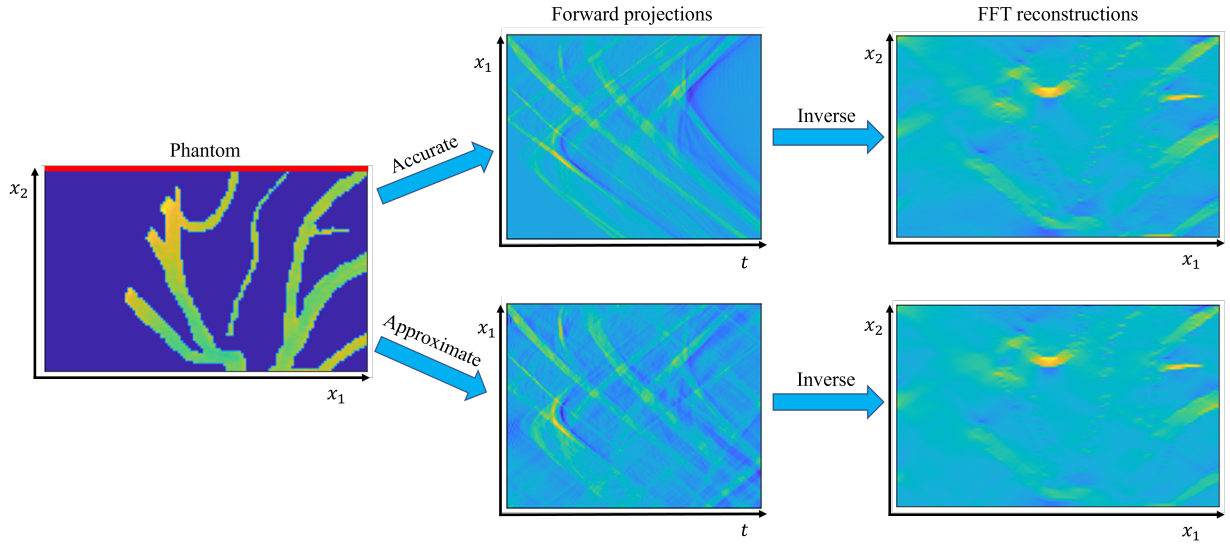


图1. 测量装置及快速模型示意图。仿体（左）由顶部的单侧传感器测量（红线）。理想测量结果显示在上方中间。快速近似前向映射显示在下方，其中包含明显的混叠伪影。应用逆映射后得到的重建结果存在严重的有限视角伪影。从右下方可以看出，基于近似模型得到的重建结果比上方的结果更平滑。

或者，利用对偶空间，我们可以在原始空间和对偶空间中构造一个极小-极大优化问题。这将导出广受认可的原始-对偶混合梯度（PDHG）方法， [5]。在此，求解（4）时，原始变量$x \in X$ 和对偶变量 $q \in Y$的更新方程为

$$q_{k+1} = \frac{q_k + \sigma(A\widetilde{x}_k - y)}{1+\sigma}, \tag{7}$$

$$x_{k+1} = \operatorname{prox}_{R,\lambda\tau}\left(x_k - \tau A^* q_{k+1}\right), \tag{8}$$

$$\widetilde{x}_{k+1} = x_{k+1} + \alpha(x_{k+1} - x_k), \tag{9}$$

其中 σ， $\tau > 0$且 $\theta \in [0,1]$。给定前向算子 L 的算子范数：$= \|A\| = \|A\|_{X\to Y}$，当 $\sigma\tau L^2 < 1$ 且 $\theta = 1$[5]时，可证明 $(q,\ x)$ 收敛到一个鞍点。需要指出的是，式（7）对应于对偶空间的近端算子，即式（4）中的最小二乘数据保真项。接下来，我们将基于式（7）和式（8）的表述构建方法。

A. 学习型模型驱动重建方法

学习型重建方法在医学成像中日益普及，因为它们能够将强大的数据驱动先验信息与人工设计的模型组件相结合 [1], ，同时还显著缩短重建时间。

在图像质量和稳健性方面表现最出色的是所谓的基于模型的技术，它们将数据驱动组件与模型组件结合起来。这意味着在网络架构中使用前向算子和伴随算子 [7], [12], [35], [36]。其中一个重要类别是遵循公式（5）的展开迭代方法，例如，将更新方程重新表述为

$$x_{k+1} = G_{\theta_k}\left(x_k - \lambda_k A^*(Ax_k - y)\right). \tag{10}$$

也就是说，用可学习的更新网络G_{θ_k} 替代近端算子，该网络具有参数 θ，其网络权重通过监督训练数据学习得到。请注意，在上述表述中，每次迭代都可以改变网络权重。我们也可以采用权重共享，为所有更新全局固定参数 θ 。固定的更新算子可以通过约束网络（权重）或更新算子的架构 [3]，使迭代重建更易于分析。

一种相关方法是所谓的即插即用方法[4] ，它使用单独训练的去噪器 G_θ，随后将其用于迭代过程。需要注意的是，此处的去噪器同样对所有迭代点保持固定。网络 G_θ 以去噪器身份运行这一特性，主要将该方法限制在去噪类逆问题中；而我们在此遇到的是有限视角问题。尽管如此，近期研究进展表明，如果 G_θ 由生成网络给出，就能在相关的采样框架 [37]中获得高质量的重建结果和不确定性图。

与式 (5)–(8) 中经典的基于模型的重建类似，我们可以将学习型迭代方法扩展为学习型原始-对偶方法 [7] ，方法是在原始变量和对偶变量空间中用学习型算子替换近端算子。由此得到的学习型更新可一般地表示为

$$q_{k+1} = F_{\phi_k}\left(q_k, Ax_k, y\right)$$
$$x_{k+1} = G_{\theta_k}\left(x_k, A^* q_{k+1}\right).$$

学习型原始-对偶方案在X射线计算机断层成像中展现出了优异的性能，但由于原始变量和对偶变量更新之间存在耦合，必须以端到端方式对所有迭代点进行联合训练，这可能带来重大的计算挑战。通过使用可逆网络 [38], [39], 在降低内存开销方面已经取得了一些进展。

尽管如此，学习型基于模型的技术在光声断层成像中的应用仍十分有限 [12],

[14], [40], 尽管研究表明，这些方法在定量和定性方面均表现更优 [41]。这主要是因为计算前向模型的成本高昂，限制了重建速度，也常常使训练难以开展。因此，PAT 中的大多数学习型方法采用一步或两步方法：前者学习从数据到重建的直接映射，后者则是在获得初始重建后，使用后处理网络。关于 PAT 学习型重建方法的全面综述，请参阅 [42]–[44]。

如前所述，文献 [12], 提出了一种克服PAT训练过程中昂贵前向模型局限的方法：采用贪心方法训练更新算子，只要求逐次迭代达到最优性。对于平面（或线性）传感器，使用更快的近似模型 [45], [46] 可进一步缩短训练和部署时间 [12]。另一种途径是完全学习前向模型，例如使用傅里叶神经算子 [47], [48] ，并将其嵌入高效的迭代重建算法中。

B. 快速近似前向算子与逆算子模型

近期研究探讨了在重建中使用近似或不完美算子 [49]–[51] 的影响，以及通过学习型校正 [15], [16]改善结果的可能性。最重要的是，研究表明，采用适当的学习型校正可以准确求解经典变分问题。我们将在此先讨论近似模型，随后在第III节介绍学习型重建。

要实现学习型迭代重建的可行训练，需要快速的正向和反向模型。这里我们将采用一种基于快速傅里叶变换（FFT）的解析近似方法，该方法适用于线传感器和平面传感器[45], [46]。值得注意的是，圆形域也存在基于 FFT 的算法 [52]，因此我们的方法可以自然扩展。我们简要回顾一下此前在 [14]中使用的近似模型。

该近似模型利用了测量点在二维情况下位于$\mathbf{x}_2 = 0$ 线上的事实（在三维情况下则位于平面上），并假设声速为常数 c_0。在二维情况下，传感器上的压力可通过以下关系与 x 联系起来 [45], [46]:

$$p(\mathbf{x}_1, t) = \frac{1}{c^2}\mathcal{F}_{k_1}\left\{\mathcal{C}_\omega\left\{B(k_1,\omega)\tilde{x}(k_1,\omega)\right\}\right\}, \qquad (11)$$

其中， $\tilde{x}(k_1,\omega)$由$\hat{x}(\mathbf{k})$通过色散关系 $(\omega/c)^2 = k_1^2 + k_2^2$ 得到，而$\hat{x}(\mathbf{k}) = \mathcal{F}_{\mathbf{x}}\{x(\mathbf{x})\}$ 是 $x(\mathbf{x})$的二维傅里叶变换。 $\mathcal{C}_\omega$ 是从 ω 到t的余弦变换， $\mathcal{F}_{k_1}$ 是探测器线上的一维逆傅里叶变换。最后，加权因子

$$B(k_1,\omega) = \omega / \left(\operatorname{sgn}(\omega)\sqrt{(\omega/c)^2 - k_1^2}\right), \qquad (12)$$

含有可积奇异性，这意味着若通过矩形网格上的离散化来计算（11），以便应用快速傅里叶变换进行高效计算，则测量数据 $p(\mathbf{x}_1, t)$中会产生混叠。因此，使用快速傅里叶变换计算（11）会得到一个快速但近似的前向模型。可以像 [14], 中那样对某些入射角进行阈值处理，以控制混叠伪影的程度，但这会导致信号强度损失。因此，本文不采用阈值处理，而是通过学习型模型校正来处理混叠。

因此，式 (11) 定义了本研究的近似前向模型 $\widetilde{A}: X \to Y$。类似地，我们可以对式 (11) 求逆，得到从测量时间序列到初始压力的映射，即逆映射或重建算子$A^\dagger: Y \to X$。由此得到图像 X空间与数据 Y 空间之间的两种映射：

$$\widetilde{A}x = \tilde{y}, \quad A^\dagger y = \tilde{x}. \qquad (13)$$

需要指出的是，快速逆映射不会引入任何混叠，但信息自然受限于有限视角几何所包含的信息；示意图见图1。下文将在学习型重建中使用上述快速映射。

此外，我们使用 PyTorch 实现了式 (13) 中的映射，以便在展开算法的训练中灵活部署。这样，我们的模型便能全面支持图形处理器和自动微分，从而实现高效训练与评估。作为本研究的一部分，我们将向研究人员公开模型实现。

III. 模型校正原始-对偶与深度平衡模型

现在，我们可以将上述内容结合起来，从经典的原始-对偶混合梯度法（PDHG）表述式（7）和（8）出发，构建一个学习型迭代方案。首先，我们将在前向算子中引入模型校正项，并在原始空间中加入一个可学习的更新算子。此外，我们的表述还允许将所提出的算法构建为深度平衡模型。

A. 模型校正的学习型原始-对偶方法

即插即用框架建议用学习型网络替代式（5）中的近端算子。已知这种方法在原始空间中效果良好。通常不会采用同样的方法来学习对偶空间中的近端算子，因为我们无法直接在对偶空间中学习去噪网络。此外，对偶空间中的近端算子与数据保真项及噪声模型有明确的联系。因此，对于我们的方法，将保留式（7）给出的对偶空间近端算子的显式形式，仅用网络替代图像空间中用于原始变量更新的近端算子 G_θ。

我们用快速的近似版本 $\widetilde{A}$替代准确但计算开销大的模型 A ，这会导致数据出现误差，进而影响对偶更新。为此，我们提出利用网络 F_ϕ学习对前向算子的校正。我们不在原始空间中使用（计算开销大的）伴随算子，而是使用快速的逆映射 $A^\dagger$。两者结合后，更新

步骤以及模型校正和学习型近端算子可表示为

$$q_{k+1} = \frac{q_k + \sigma(F_\phi(\widetilde{A}x_k) - y)}{1+\sigma} \tag{14}$$

$$x_{k+1} = G_\theta\left(x_k - \tau A^\dagger q_{k+1}\right). \tag{15}$$

请注意，我们对所有迭代点使用共享网络参数θ，ϕ 并设置最大迭代次数 $K > 0$。我们将这一无约束版本称为模型校正学习型原始-对偶（MC-PD），如算法 1 所示。

算法 1 模型校正的原始-对偶（MC-PD）

1: 给定 $y, \widetilde{A}, A^\dagger$
2: 设置 σ, τ, K
3: $x_0 \leftarrow A^\dagger y$
4: $q_0 \leftarrow 0$
5: $k \leftarrow 0$
6: **对于** $0 < k \leq K-1$ **做**
7: $q_{k+1} \leftarrow \frac{q_k+\sigma\left(F_\phi\left(\widetilde{A}x_k\right)-y\right)}{1+\sigma}$
8: $x_{k+1} \leftarrow G_\theta\left(x_k - \tau A^\dagger q_{k+1}\right)$
9: **结束**

B. 原始-对偶深度平衡模型

近年来，收敛型学习图像重建方法受到广泛关注。我们不只是通过监督方法训练学习型重建，而是可以对权重施加约束，从而为展开迭代的行为提供理论保证。一种方法是要求学习得到的迭代点收敛到一个不动点，这正是深度平衡模型的基本范式。最简单的做法是确保式（10）中的学习型近端梯度满足不动点方程，即

$$x^* = G(x^*; y) = G_\theta(x^* - \lambda A^*(Ax^* - y)). \tag{16}$$

首先需要指出，迭代点之间共享权重，且不假定最大迭代次数，而是通过确保更新（例如）构成压缩映射来考虑极限下的收敛。在深度平衡模型中，前向求解器采用不动点迭代，可以使用普通版本，也可以选择更先进的方法，例如安德森加速；详情及其在逆问题中的扩展，见 [17] 和 [18]。随后，通过对不动点形式进行微分来训练，从而无需对所有迭代点进行反向传播，进而降低内存消耗。

在我们的情形中，给定两个变量 x，q，我们希望据此构建深度平衡模型。这意味着，我们现在要寻找一对不动点 $z^* = (x^*, q^*) \in X \times Y$，使不动点方程满足

$$(x^*, q^*) = \mathrm{PD}_{(\phi,\theta)}(x^*, q^*; y), \tag{17}$$

其中，映射 PD 对应的更新方程由式 (14) 和式 (15) 给出。接下来，我们将先讨论前向传播，再讨论反向传播的计算。

1) 深度平衡形式的前向传播：在这里，我们将对原始变量和对偶变量均应用 Anderson 加速不动点法。更具体地说，给定上述针对迭代 k的原始-对偶更新：

$$(x_{k+1}, q_{k+1}) = \mathrm{PD}_{(\phi,\theta)}(x_k, q_k; y).$$

我们用上标表示所计算的更新，即 PD^q 和 PD^x，由此得到如下形式

$$q_{k+1} = \sum_{i=0}^{m-1} \beta_i \mathrm{PD}^q_{(\phi,\theta)}(x_{k-i}, q_{k-i}; y), \tag{18}$$

$$x_{k+1} = \sum_{i=0}^{m-1} \alpha_i \mathrm{PD}^x_{(\phi,\theta)}(x_{k-i}, q_{k-i}; y), \tag{19}$$

其中包含记忆项 $m \geq 1$、$\sum_i \alpha_i = 1$ 和 $\sum_i \beta_i = 1$。参数向量 α 和 β 通过求解以下问题得到

$$\min_\alpha \|H\alpha\|_2^2, \text{ subject to } \sum_i \alpha_i = 1,$$

其中，矩阵 H 由向量化残差组成

$$H = \Big[\mathrm{PD}^x_{(\phi,\theta)}(x_k, q_k; y) - x_k, \cdots, \mathrm{PD}^x_{(\phi,\theta)}(x_{k-m+1}, q_{k-m+1}; y) - x_{k-m+1}\Big],$$

类似地，我们对 β 和对偶变量 q 进行相同操作。原始-对偶深度平衡模型（PD-DEQ）的前向传播与算法 1 类似，不同之处在于，新的迭代点由式（18）和（19）中的安德森加速版本计算。此外，除了考虑最大迭代次数外，还可以检查不动点迭代是否已收敛，例如，对原始变量的相对残差设定上限

$$r = \frac{\|PD^x_{(\phi,\theta)}(x_{k+1}, q_{k+1}; y) - x_k\|_2}{\|PD^x_{(\phi,\theta)}(x_{k+1}, q_{k+1}; y)\|_2}.$$

2) 反向传播：在训练过程中，尽管我们会更新两个变量，但由于原始变量x的损失易于通过训练数据获得，因此我们只针对该损失优化网络参数。这意味着，我们在式（17）中仅针对 x构造不动点方程。稍微滥用一下符号，我们将对偶变量 q 和测量值 y下的不动点 x^* 记为

$$x^* = \mathrm{PD}_{(\phi,\theta)}(x^*; q, y). \tag{20}$$

为简化符号，我们将网络变量统称为$\zeta = \{\phi, \theta\}$。随后采用监督式训练网络，通过最小化其与真实值 x 之间的均方误差（MSE）来进行训练：

$$\ell(\zeta) = \|\mathrm{PD}_{(\zeta)}(x^*; q, y) - x\|_2^2 = \|x^* - x\|_2^2.$$

反向传播通过对不动点方程 (20) 进行隐式微分来计算。我们将不动点处相对于参数 ζ 的雅可比矩阵记为 $J_{\mathrm{PD}^{x^*}}(\zeta)$，将相对于 x^* 的雅可比矩阵记为 $J_{\mathrm{PD}_{(\zeta)}}(x^*)$，由此可得梯度所满足的隐式方程

$$\nabla\ell(\zeta) = \left(J_{\mathrm{PD}^{x^*}}(\zeta)\right)^T \left(\mathrm{Id} - J_{\mathrm{PD}_{(\zeta)}}(x^*)\right)^{-T} \left(x^* - x\right). \tag{21}$$

请注意，我们无需通过所有迭代点计算反向传播，因此内存消耗与展开迭代的数量无关。不过，式 (21) 需要一个求解器来求解隐式方程，这可能会增加一些计算开销。这里我们将考虑完整形式，但通过对 $J_{\mathrm{PD}_{(\zeta)}}(x^*)$ 进行近似（尤其是采用无雅可比变体 [53]），还可以进一步减少计算量。

虽然我们的主要关注点是原始变量 x 的收敛，但现在合理的问题是，我们需要对 q 作出什么假设。下面对此进行讨论。

C. 收敛分析

在此，我们将简要考察所提出方法的一些理论性质。我们将探讨模型校正如何在 MC-PD 表述中得到解释，随后讨论深度平衡模型的压缩映射和不动点性质。

1) 前向校正下的泛函收敛：我们首先考虑式 ~(7)-(9) 中的解析 PDHG 表述，其中使用近似算子 A、其对应的~伴随算子 A^*，以及与某种~ ~正则化 R相对应的近端算子。给定算子范数 $L = \|A\|$ 后，我们可以~选择超参数，使其满足 $\sigma = \tau < 1/L$ 和 $\alpha \in [0, 1]$。然后，应用原始-对偶混合梯度法将最小化代价泛函

$$\|\widetilde{A}x - y\|_2^2 + \lambda R(x).$$

如 [15]所示，使用近似前向算子而非校正项来最小化上述泛函，确实会导致次优结果。

因此，现在考虑在对偶更新中加入前向算子校正，即在式（7）中用 $F_\phi(A)$替换 A，同时在原始~变量更新（8）中使用正确的伴随算子。我们将校正项 F_ϕ 的算子范数记为 $\varepsilon_\phi = \|F_\phi\|$。那么，在伴随算子训练完美这一假设下，我们可以恢复原始-对偶混合梯度法的收敛性，具体如下。

~**命题1**。 给定校正后的前向算子 $F_\phi(A)$，其~范数为 $\varepsilon_\phi L$；精确的伴随算子 A^* 的范数为 L。假设对于$k \geq 0$，我们有

$$\big(F_\phi(\widetilde{A}x_k), y_k\big)_Y = \big(x_k, A^* y_k\big)_X. \tag{22}$$

~那么，对于选择 $\tau < 1/L$ 和 $\sigma < 1/(\varepsilon_\phi L)$，我们可以恢复式(7)-(9) 中的原始-对偶混合梯度法迭代，以求解精确泛函(4)。

证明。 条件（22）要求前向校正确实满足伴随方程。根据伴随算子的唯一性，这等价于完美训练条件

$$\big(F_\phi(\widetilde{A}x_k) - Ax_k, y_k\big)_Y = 0 \quad \forall k \geq 0.$$

因此，式 (7)-(9) 中的所有迭代点都与精确算子相同。通过选择超参数 σ 和 τ，我们有~，因此 $\tau\sigma LL\varepsilon_\phi < 1$，经典 PDHG 的收敛条件得以满足。 □

条件 (22) 可用于训练，而无需访问精确的前向算子。也就是说，可以在训练中增加一个额外的伴随损失

$$L(x; \phi) = (F_\phi(\tilde{A}x), h)_Y - (x, A^* h)_X \text{ for } h \in Y. \tag{23}$$

显然，实践中无法实现完美训练。在这种情况下，可以将其视为伴随算子中的扰动，或者实际上视为前向算子中的扰动。关于原始-对偶算法在伴随失配情况下的收敛讨论，请参阅 [54], [55]。

请注意，尽管上述伴随损失（23）仍需要精确的伴随算子。因此，我们现在考虑将前向算子和伴随算子都分别替换为对应快速算子的情况。此时，相应的伴随损失变为

$$L(x; \phi) = (F_\phi(\tilde{A}x), h)_Y - (x, A^\dagger h)_X \text{ for } h \in Y, \tag{24}$$

其中，强制要求前向校正与快速逆算子互为伴随。显然，这并不能最小化精确泛函，而是最小化相应的泛函

$$\|F_\phi(\widetilde{A})x - y\|_2^2 + \lambda R(x),$$

与真实模型的接近程度取决于快速逆算子 $A^\dagger$ 与伴随算子 A^*的接近程度。若希望相对于精确泛函实现收敛，则需要进行伴随校正 [15]。

2）深度平衡模型的不动点收敛：现在，我们将近端算子替换为学习型网络 G_θ 及相应的范数 ε_θ。首先需要指出，如果不对 G_θ施加任何额外条件，就不能再期待泛函收敛，但可以得到不动点收敛 [3]。首先提醒一下，PD-DEQ模型的训练仅针对原始变量 x进行。因此，我们希望验证以下两个条件

(i) 在什么参数选择以及对网络 F_ϕ 和 G_θ 施加什么条件下，原始变量上的映射是压缩映射?

(ii) 我们也能确保不动点收敛吗?

正如我们将看到的，这两个条件密切相关，因为条件 (ii) 要求映射是压缩映射；但我们也将看到，如果不作进一步假设，依赖关系 q 并不能保证不动点收敛。首先，让我们通过以下方式合并更新方程 (14) 和 (15)：

$$x_{k+1} = G_\theta\Big(x_k - \tau A^\dagger\big(q_k + \sigma\big(F_\phi(\widetilde{A}x_k) - y\big)\big)/(1+\sigma)\Big).$$

我们可以看到更新中对 q_k 的依赖关系。让我们通过引入算子 $T : X \to X$ 来单独考察 x 的更新，使得

$$T_{\mathrm{Id}}(x) := (\mathrm{Id} - T)(x) = x - \tau A^\dagger \sigma F_\phi(\widetilde{A}x)/(1+\sigma). \tag{25}$$

因此， T 是正规算子 A^*A的缩放版本。我们注意到，当 T 是坚定非扩张映射时，上述关于 x (25)的更新规则会收敛到一个不动点，这意味着 Id $-T$是坚定非扩张的 [56]。为了证明坚定非扩张性，我们需要一个伴随损失条件。我们不假设完美拟合，而是考虑一个近似伴随损失条件，使得

$$(F_\phi(\tilde{A}x), h)_Y \leq (1+\varepsilon)(x, A^\dagger h)_X \text{ for } h \in Y. \tag{26}$$

这意味着我们仅在一定的 $\varepsilon > 0$误差下满足伴随条件，该误差取决于右侧内积的值。这是一个更现实的情况，因为我们无法期望完美的伴随拟合。我们注意到，假设(26)要求$(x, A^\dagger h)_X$ 为正。然而，由于所有样本 x 和相关的 h 均为正，我们可以假设这一条件成立。现在我们可以证明 T的坚定非扩张性。

定理 1. 在条件 (26) 下，算子 $T\colon X \to X$ （定义见(25)）是坚定非扩张的，其中 $\varepsilon > 0$、 $L^\dagger = \|A^\dagger\|$ 以及 σ、$\tau > 0$ 的选取满足

$$(1+\varepsilon)\frac{\tau\sigma}{1+\sigma}(L^\dagger)^2 \leq 1. \tag{27}$$

证明. 我们需要证明

$$\|T(x) - T(v)\|_2^2 \leq (T(x) - T(v), x - v)_X, \tag{28}$$

其中 $T\colon X \to X$ 如（25）中所定义。我们先利用$A^\dagger$ 和 $\|A^\dagger\| = L^\dagger$的线性性，对上述不等式左侧进行估计，得到

$$\begin{aligned}
&\Big\|\frac{\tau\sigma}{1+\sigma}A^\dagger F_\phi(\widetilde{A}x) - \frac{\tau\sigma}{1+\sigma}A^\dagger F_\phi(\widetilde{A}v)\Big\|_2^2 \\
&\quad\leq \Big|\frac{\tau\sigma}{1+\sigma}\Big|^2\Big|L^\dagger\Big|^2\Big\|F_\phi(\widetilde{A}x) - A^\dagger F_\phi(\widetilde{A}v)\Big\|_2^2 \\
&\quad= \Big|\frac{\tau\sigma L^\dagger}{1+\sigma}\Big|^2\Big(F_\phi(\widetilde{A}x) - F_\phi(\widetilde{A}v), F_\phi(\widetilde{A}x) - F_\phi(\widetilde{A}v)\Big)_X.
\end{aligned}$$

现在利用半线性和对称性，将内积分开。然后使用（26）中的近似伴随条件分别估计各项并重新组合，得到

$$\begin{aligned}
&\Big(F_\phi(\widetilde{A}x) - F_\phi(\widetilde{A}v), F_\phi(\widetilde{A}x) - F_\phi(\widetilde{A}v)\Big)_X \\
&\quad\leq (1+\varepsilon)\Big(A^\dagger F_\phi(\widetilde{A}x) - A^\dagger F_\phi(\widetilde{A}v), x - v\Big)_X.
\end{aligned}$$

现在，在条件（27）下，我们得到所需的估计。

□

接下来，我们考察学习型原始-对偶更新的迭代，并在对偶更新有界的条件下，得到关于固定 k 的压缩映射。

定理 2. 假设 $\|q_k - q_{k-1}\|_2 < C_q$在 $k > 0$ 固定且满足 $\|G_\theta\| = \varepsilon_\theta < 1$的条件下成立。在定理 *1* 的假设下，存在一种 $\tau > 0$的取法，使得 x_{k+1} 的迭代是压缩的，即

$$\|x_{k+1} - x_k\|_2 < \|x_k - x_{k-1}\|_2.$$

证明。 我们首先利用T_{Id} 将 x_{k+1} 的更新重新表述为

$$x_{k+1} = G_\theta\Big(T_{\mathrm{Id}}(x_k) - \frac{\tau}{(1+\sigma)}A^\dagger(q_k - \sigma y)\Big).$$

利用上述更新规则，通过网络$\|G_\theta\| = \varepsilon_\theta$ 和 $\|A^\dagger\| = L^\dagger$ 的范数，运用三角不等式可得到迭代点的第一个上界：

$$\begin{aligned}
&\|x_{k+1} - x_k\|_2 \leq \\
&\quad\varepsilon_\theta\Big(\big\|T_{\mathrm{Id}}(x_k) - T_{\mathrm{Id}}(x_{k-1})\big\|_2 + \frac{\tau L^\dagger}{1+\sigma}\big\|q_k - q_{k-1}\big\|_2\Big).
\end{aligned} \tag{29}$$

令 $0 < \delta = 1 - \varepsilon_\theta < 1$，由于 $\|q_k - q_{k-1}\|_2 < C_q$ 且 $\|x_k - x_{k-1}\|_2$ 有限，存在足够小的 $\tau > 0$，使得

$$\frac{\tau L^\dagger}{1+\sigma}C_q < \delta/2\|x_k - x_{k-1}\|_2. \tag{30}$$

接着，利用第一项中 T_{Id} 的坚定非扩张性以及第二项中的式 (30)，得到式 (29) 的最终估计：

$$\begin{aligned}
\|x_{k+1} - x_k\|_2 &\leq \varepsilon_\theta(1+\delta/2)\|x_k - x_{k-1}\|_2 \\
&< \|x_k - x_{k-1}\|_2,
\end{aligned}$$

这是因为根据假设

$$\varepsilon_\theta(1+\delta/2) < \varepsilon_\theta + \delta/2 = (1+\varepsilon_\theta)/2 < 1.$$

□

上述定理仅在对偶变量 q一致有界的条件下，给出了逐迭代点收敛性。这意味着，如果 q_k不收敛， τ 可能变得任意小。需要指出的是，在实践中我们只考虑有限次迭代，因此存在一致界。对于极限情形，我们可以假设 q_k 至少以与 x_k一样快的速度收敛，由此可以得到 $\tau > 0$的一致取值。为此，假设两个变量都具有 $1/k$ 收敛性。

推论 1。 在定理 *2* 的假设下，进一步假设 $\|q_k - q_{k-1}\| = \mathcal{O}(1/k)$ 和 $\|x_k - x_{k-1}\| = \mathcal{O}(1/k)$ 均满足相应条件。则存在一个对所有 $k > 0$ 均成立的统一 $\tau > 0$，使得 x_k 是压缩的。

证明。 根据假设，估计式（30）的两边均可除以 k （两边可能还需乘以额外的常数）。因此，存在一个足够小的 $\tau > 0$，使其对所有 k 均成立。

□

首先，上述分析表明，应将 τ 和 σ 保守地选择为较小的值，以确保迭代具有压缩性。正如我们将看到的，这对于 DEQ 形式的稳定性实际上十分重要。其次，我们希望对偶 q 中的残差具有较小的范数，并且最好能够快速收敛。第三，值得注意的是，实际上只要求对偶网络 G_θ 是压缩的，而模型校正网络 F_ϕ 通过近似伴随误差与逆算子 $A^\dagger$ 相连接。在实验中，我们将看到， q 确实能够快速收敛，并且其残差的范数显著小于 x 的残差范数。因此，实验确认，即使理论假设依赖于 q 的压缩性，我们仍然能够在原始变量和对偶变量上恢复不动点迭代。

IV. 光声数据模拟与计算方面

在本研究中，我们将通过模拟的二维场景评估所提出算法的性能。为此，我们首先讨论训练数据生成。随后，我们将继续讨论正向模型和逆向模型实现的相关细节，最后介绍所采用的训练流程。

A. 二维光声模拟与训练数据

为模拟训练数据，我们使用基于伪谱方法的 k-wave，在二维中准确模拟声波方程（1）。我们考虑一个大小为 80×128 、各向同性空间像素尺寸为 $dx = 106\mu$m 的矩形计算域。光声传感器位于计算域的上边缘，示意图见图1。声速设为恒定的 1500 m/s，时间采样率设为 $dt = 50$ ns；测量数据的维度为 160×128, 间距为 $dt \times dx$。我们在模拟测量值中加入了 1% 的高斯随机噪声，这相当于信噪比约为 18 分贝。

在生成训练数据时，我们使用了*DRIVE*：用于血管提取的数字视网膜图像[1]，，其中包含用于训练的视网膜照片和分割掩膜。我们使用了 20 张图像及其对应的掩膜来创建血管结构仿体。首先，将图像转换为灰度图并进行归一化，然后将掩膜与图像相乘。每张生成的图像尺寸为 584×565，随后我们从图像中提取尺寸为80×128 的小图像块，并通过检查所有像素值之和，仅保留特征足够丰富的图像块，即像素值之和大于 150 的图像块。我们对每张图像转置后重复相同的流程，以进行数据增强，共得到 893 张图像。随后，我们通过垂直翻转图像进一步扩充了数据集，因为这样可以产生不同的有限视场伪影。最终共获得 1786 张图像，其中 $N = 1600$ 用于训练，93 张用于验证，另有 93 张用于测试。

B. 快速模型的实现

在本研究中，我们完全使用 PyTorch 实现了两个模型：快速近似前向算子 $\widetilde{A}$ 和快速逆算子 $A^\dagger$，以便进行高效评估。这也使得能够在基于模型的学习型重建中使用自动微分。该实现主要遵循第 II-B 节中的描述，并使用 PyTorch 内置的快速傅里叶变换。对于图像 k-空间 $\mathbf{k} = (k_1, k_2)$ 与测量 k-空间 (k_1, ω) 之间的映射，我们在 PyTorch 中自行实现了网格插值器，因为没有可用的内置函数。

为提高训练期间的计算效率，所有必要的静态变量和 k- 网格均已预先计算，并传递给快速模型。这遵循了与 k-wave 工具箱 [29]类似的范式。作为本研究的一部分，我们发布了快速模型和网络代码。 [2]

C. 模型校正的原始-对偶模型的实现与训练

所提出模型的网络 F_ϕ 和 G_θ 均选用 U-Net，包含 3 个尺度，即 2 个下采样层；第一个尺度有 64 个通道，后续尺度分别有 128 和 256 个通道。我们将两种算法的最大迭代次数设为 10。PD-DEQ 的 Anderson 加速使用 $m = 5$ 个记忆单元。对于超参数 τ 和 σ，我们~计算得到了 $L \approx 0.7$ 的近似算子范数，并将 ~ 设为 $\tau = \sigma = 1/(10L)$。由于所提出的理论，我们将其高估了 10 倍，这在实践中提升了稳定性。

如前所述，我们只针对原始变量训练网络。已知真实图像x^i 及其对应的测量值 y^i，以及相应的重建算子 $\mathcal{R}_\zeta : Y \to X$（MC-PD 或 PD-DEQ）及其参数 $\zeta = \{\phi, \theta\}$。随后，我们最小化ℓ^2-损失（均方误差），以求得最优参数集 ζ^*：

$$\zeta^* = \arg\min_\zeta \sum_{i=1}^{N} \|\mathcal{R}_\theta(y^i) - x^i\|_2^2. \tag{31}$$

受内存限制，每次训练迭代仅使用 1 个样本，并且每次迭代时，训练样本均从所有训练数据对中按均匀分布抽取。我们使用 PyTorch 中的 Adam 优化器，并采用余弦衰减调整学习率，初始学习率为 $2{\cdot}10^{-4}$，总共进行 25000 次训练迭代。在配备 24 GB 内存的 Nvidia Quadro RTX 6000 上，MC-PD模型的训练耗时 100 分钟，PD-DEQ模型的训练耗时 180 分钟。

训练 PD-DEQ模型时，我们在不动点形式下使用自动微分，计算公式 (21) 给出的梯度。需要注意的是，要稳定地求解公式 (21)，网络必须是压缩的。因此，我们测试了使用谱归一化来显式约束 G_θ 的利普希茨常数，但不使用谱归一化时效果更好；相关讨论见第 V-C2 节。这表明，压缩映射通过不动点形式得到了隐式保证。尽管如此，我们提醒读者，只有原始空间中的更新网络G_θ 需要满足利普希茨约束，而对偶网络 F_ϕ 可以不受约束。我们还指出，针对对偶变量进行监督训练并不直接，因为理想的参考值应仅为残差噪声。

1）一种组合混合方法：我们将看到，尽管 PD-DEQ 模型能够提供良好的图像质量，但在所展示的有限视角设置下，它难以获得较高的峰值信噪比。因此，我们还将考虑 MC-PD 与 PD-DEQ 两种方法的混合组合。具体而言，我们先执行固定次数的 MC-PD 迭代，然后再使用 PD-DEQ。之所以采用这一顺序，是因为先通过 MC-PD 减轻有限视角设置的影响，再由最后的 PD-DEQ 迭代提供用于最终重建结果的定点形式。正如后文将要讨论的，我们需要在 MC-PD 的性能与 PD-DEQ 带来的内存减少之间进行权衡。下文将这种方法称为 PD-Hybrid

对于训练目标，我们以交替方式训练网络，即在每次训练迭代中，我们首先评估 MC-PD，记为 $\mathcal{R}^{MC}_{\zeta_1}$，并对样本 i 最小化式（31）。然后，给定输出（x^i_k, q^i_k）$= \mathcal{R}^{MC}_{\zeta_1}(y^i)$，我们评估 PD-DEQ，记为 $\mathcal{R}^{DEQ}_{\zeta_2}$，并仅针对 PD-DEQ模型的参数最小化损失函数，

[1]https://drive.grand-challenge.org/[2]链接将在修订后添加

即，

$$\|\mathcal{R}^{DEQ}_{\zeta_2}(x^i_k, q^i_k, y^i) - x^i\|^2_2.$$

我们将使用 5 次 MC-PD 迭代，随后进行 5 次 PD-DEQ 迭代。训练 PD-Hybrid模型约需 160 分钟。

D. 对比方法

我们还将把这两种模型校正的原始-对偶模型与两种成熟的方法进行性能比较。首先，我们采用一种后处理方法：先通过逆算子 $x_0 = A^\dagger(y)$ 获得初始重建，然后训练一个后处理网络以去除噪声和有限视角伪影，即，

$$x_{\mathrm{rec}} = G_\theta(x_0) = G_\theta(A^\dagger(y)).$$

这里，我们将对网络G_θ 采用与 MC-PD 和 PD-DEQ模型相同的 U-Net架构。

其次，我们将把我们的方法与不采用权重共享的 MC-PD 版本进行比较，因此它与原始的学习型原始-对偶（LPD） [7], 相近；下文中我们也将这种模型称为 LPD。该网络显然具有最强的表达能力，因为每次迭代在原始空间和对偶空间中都有各自的网络权重，因此我们将其视为与我们的方法进行比较的最先进方法。需要指出的是，这里的网络同样采用相同的 U-Net架构。训练 U-Net 仅需 12 分钟，而训练 LPD 需 120 分钟。

V. 结果与讨论

我们将评估所提出方法的定量和定性表现，以及其提供可扩展且稳健的重建结果的能力。随后讨论深度平衡模型。

A. 二维模拟数据的结果

我们首先评估第 IV-A 节所述测试数据上的性能；这些数据与训练数据一致，即分辨率为 80×128，测量值中含有 1% 的噪声。对于每种算法，我们训练了 3 个实例，并选择验证误差最高的实例，在包含$N = 93$ 个样本的测试集上计算峰值信噪比和结构相似性指数定量指标；所得数值见表 I。

表 I 所有纳入考察的方法在测试数据（$N = 93$）上的定量指标（峰值信噪比和结构相似性指数）。

	U-Net	LPD	MC-PD	PD-DEQ	混合方法
PSNR	21.95	29.06	**29.34**	22.1	24.27
SSIM	0.943	**0.991**	**0.991**	0.959	0.974

我们可以得出两个有趣的观察。首先，尽管 LPD 并不像 MC-PD 那样采用权重共享，但二者的表现非常相近。这与已有文献的结论基本一致：采用权重共享的版本与其对应版本表现同样出色。有趣的是，在本例中，MC-PD 的峰值信噪比甚至略高一些，但结构相似性指数相同。其次，我们注意到，受约束版本 PD-DEQ 的峰值信噪比显著下降。这一结果令人遗憾，尤其是与 U-Net 相比，我们只能看到微弱的提升。尽管如此，与采用 U-Net 的后处理相比，其结构相似性指数仍有明显提升。结合 MC-PD 和 PD-DEQ 的混合方法能够改善这两项定量指标，但性能仍不及无约束版本。这与以往文献的结论相反；已有研究大多报告称，DEQ模型能够达到几乎与无约束版本同样好（甚至更好）的结果 [18]。我们在此指出，本研究面对的是一个病态程度严重的有限视角设置。因此，我们认为性能下降是由 PAT 有限视角问题带来的额外复杂性所致，后文将对此作进一步讨论。

现在，让我们考察图2所示测试集中一个样本的定性表现。我们可以立即注意到，最显著的差异在于各方法恢复出的定量值。没有任何一种方法能很好地恢复左上方血管的对比度；LPD 和 MC-PD 在这方面表现最佳。两种受约束的 DEQ 方法主要在中心区域出现对比度损失，而 U-net 则严重高估了数值，且对比度变化突兀。就有限视角伪影而言，这些伪影主要出现在底部两角，所有方法的表现都非常好，只有 U-net 和 PD-DEQ 在下方血管中呈现轻微模糊。总体而言，考虑到有限视角设置，视觉效果相当出色。尤其是在血管形状方面，另外三种方法 LPD、MC-PD 和 PD-Hybrid 的表现非常相似，整体视觉效果良好。

B. 可扩展性

使用近似模型和深度平衡模型表述的主要动机，是实现可扩展的学习型迭代方法，即使其适用于大尺寸图像，并缩短计算时间，从而使训练时间保持在合理范围内。因此，接下来我们将考察不同算法随图像尺寸和计算时间变化的扩展性。

为此，我们逐步增大所考虑的图像尺寸，从初始的 $80 \times 128 \approx 10^4$ 像素增至 $640 \times 1024 \approx 6 \cdot 10^5$ 像素，同时保持网络架构不变。需要注意的是，最大规模情形下的测量数据尺寸因此为 1280×1024。图 3 展示了不同方法在数据规模增大时的可扩展性。首先可以看到，PD-DEQ模型和 U-Net 的内存消耗随图像尺寸增大而表现良好，因此可应用于最终图像尺寸。相比之下，MC-PD 在图像尺寸达到初始尺寸的 3 倍时就已耗尽内存。PD-Hybrid 可利用较低的内存消耗，扩展性略好。此处省略了 LPD，因为其表现与 MC-PD 类似。需要注意的是，通过采用较小的网络（尤其是减少通道数），并为 MC-PD 和 PD-Hybrid 减少迭代次数，可以轻松进一步改善内存消耗方面的可扩展性；而 DEQ 的内存消耗仅取决于网络大小，与迭代点无关。在运行时间方面，U-Net 明显优于

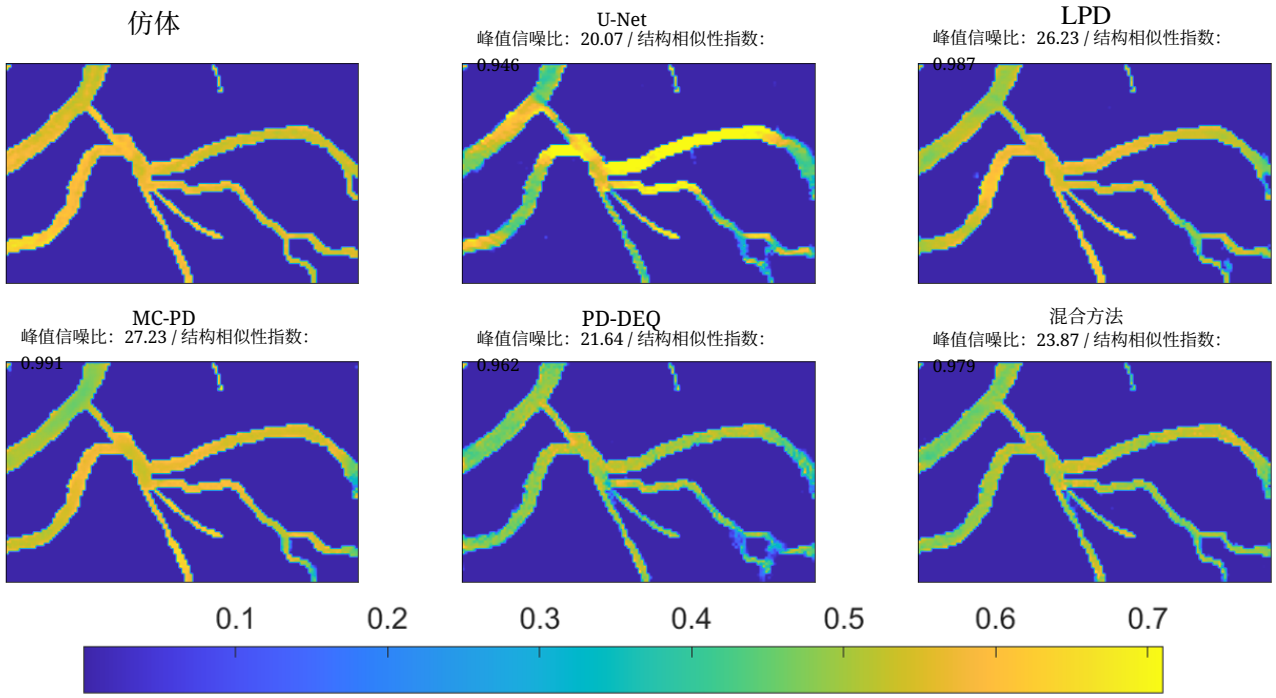


图 2. 所考虑方法在测试集某个样本上的重建结果及其定量指标。所有图像使用相同的色标。测量值在上边缘采集，噪声为 1% 的高斯噪声。重建结果的分辨率为 80×128。

所有其他方法，因为它只需使用快速逆算子进行一次重建，并进行一次网络评估。所有迭代方法的表现相近；与 PD-Hybrid 和 MC-PD 相比，PD-DEQ模型的运行时间略有增加，但差异并不显著。对于不超过 4 倍的图像尺寸（即图像尺寸为 320×512），PD-DEQ模型的运行时间可低于 1 秒。通过减少展开迭代次数，还可以轻松进一步加速。需要注意的是，前向传播耗时 1 秒，意味着在 25000 个训练迭代点下，训练时间约为 7 小时。因此，即使是 PD-DEQ 的最大图像尺寸，其训练时间也仍可接受，为 $\sim$35 小时。

最后，我们考察了模型应用于比训练时更大的图像尺寸时的表现。在这项测试中，我们采用数据生成时使用的完整图像尺寸，并生成了尺寸为 160×512 的样本进行测试。结果如图4所示。可以看到，当应用于比训练时更大的图像尺寸时，这些网络仍能良好运行。尽管所有迭代方法都能提供良好的图像质量，与测试集上的结果一致，但我们可以看到，U-Net后处理会在空白区域产生一些残余伪影。与前一种情况类似，受约束方法无法完全恢复高吸收区域的对比度，并且底部角落中略微模糊的特征显示出一些有限视角伪影。

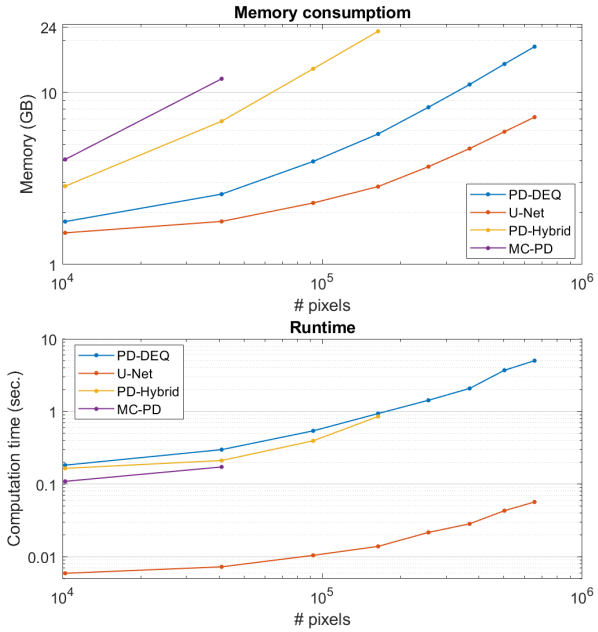


图 3. 所考察方法的可扩展性测试。上图显示内存消耗，下图显示前向传播的运行时间。

C. 有限视角设置中的挑战

在实验中，我们发现，在有限视角设置下训练DEQ模型会面临一些挑战。首先，正如结果所示，DEQ网络的表现不如其无约束相比之下。这在一定程度上可归因于受约束网络的表达能力有限；在逆问题中，其他受约束网络也能观察到类似现象，例如受到凸性约束的网络 [57]。此外，有限视角设置会导致初始重建的对比度不正确，这是因为单侧测量造成能量损失。为解决这一问题，我们将

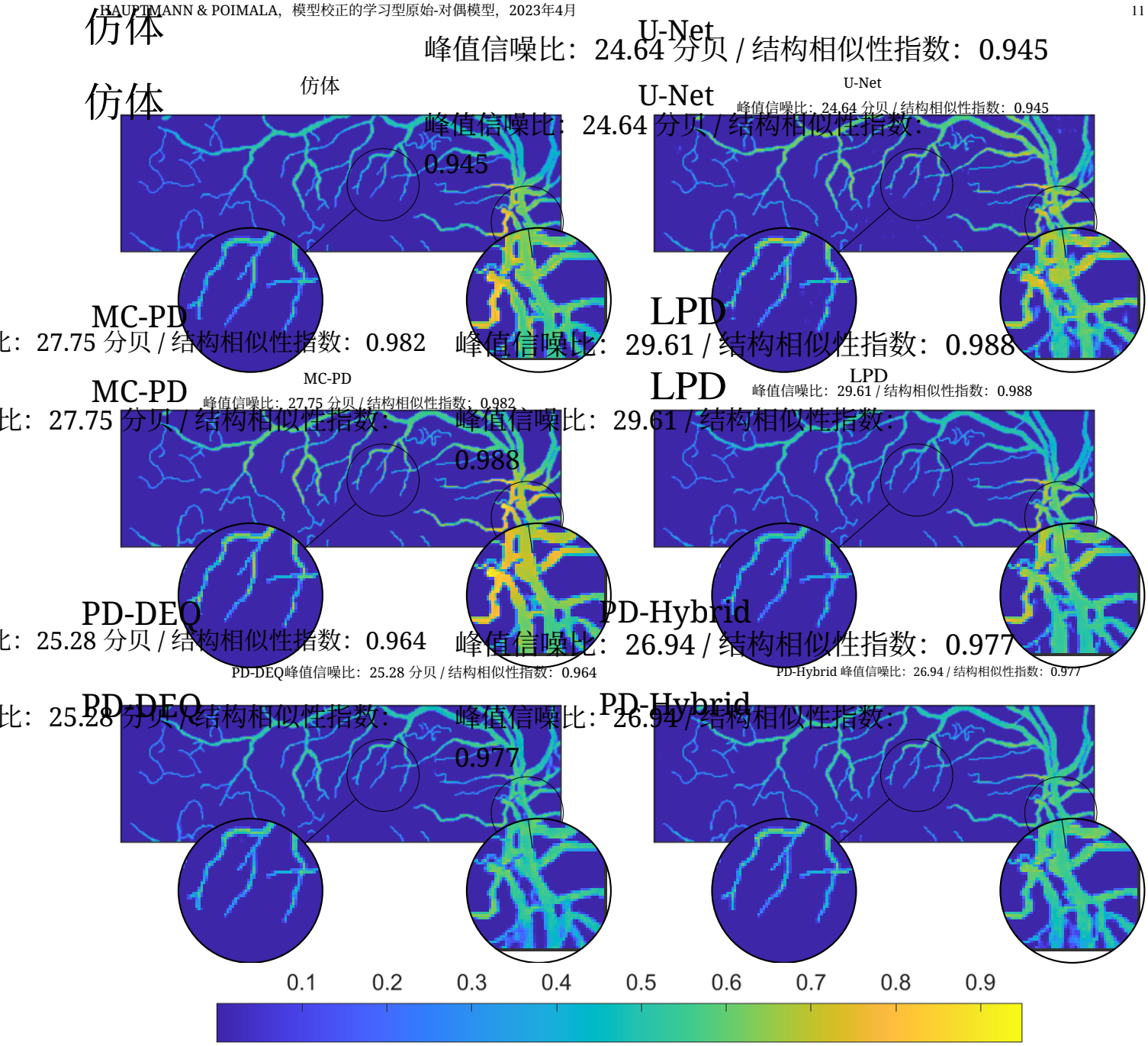


图 4. 所考虑方法在尺寸为 160×512 的较大测试图像上获得的重建结果及其定量指标。所有图像均采用相同的色标。测量值取自顶部边缘，并加入了 1% 的高斯噪声。

输入值放大了（乘以 4），使 $[0,1]$中的对比度大致达到正确水平。尽管如此，探测器附近的区域重建得更清晰，对比度也高于底部区域。我们推测，在这种情况下，DEQ模型的压缩特性并不适合弥补这一缺陷，因此与MC-PD和U-Net相比，PD-DEQ和PD-Hybrid的重建结果普遍存在对比度损失。我们认为，这是峰值信噪比下降的主要原因。

1）算法参数的影响：第三节 C 的理论预测，需要较小的 τ, σ 才能保证收敛。事实上，我们观察到，当 $\tau, \sigma > 1/(5L)$ 过大时，迭代过程会不稳定，训练可能无法成功。而 $< 1/(20L)$ 过小则会削弱更新的影响。我们的实验表明，在这些取值之间，训练具有良好的稳定性，并且可以观察到收敛。

对于深度平衡模型，迭代次数同样发挥着重要作用。这里也观察到了类似的现象：迭代点过少 < 5 无法带来足够的改进，而迭代点过多 > 20 往往无法收敛，并导致训练不稳定。因此，本研究选择了 10 个迭代点。

2）谱归一化：我们分别训练了采用和未采用谱归一化的原始变量更新网络 G_θ。尽管谱归一化能够稳定训练，并且可以选择更大的

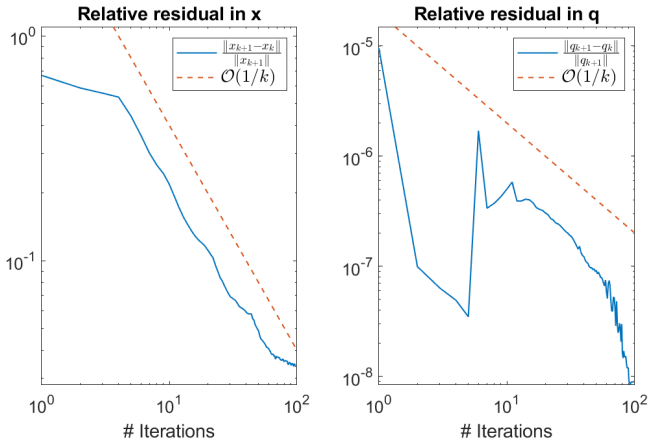


图 5. PD-DEQ表述中的残差收敛情况，与速率 $1/k$ 相比。网络训练 10 个迭代点，并测试 100 个。

τ，σ，但我们也观察到，训练过程中更容易陷入较差的局部极小值。未采用谱归一化时，早期迭代点可能不稳定；但当 τ，σ 足够小时，我们能够稳定地训练网络，并且大多数训练运行都取得了更好的性能。如第 IV-C 节所述，即使不采用谱归一化，也有观察表明训练会隐式地保证压缩性，从而能够实现稳定训练。此外，谱归一化会增加计算开销。因此，我们最终的训练和测试均未采用谱归一化。

3) 不使用模型校正进行训练：针对深度平衡模型，进行了一项消融研究：训练时不使用模型校正网络 F_ϕ，这种做法更接近原始深度平衡模型。遗憾的是，我们未能取得足够理想的结果。不稳定性问题加剧的同时，网络也未能找到良好的局部极小值。较小的 τ，$\sigma <$ 有助于稳定训练，但无法获得良好的结果。在我们的实验中，验证集上的峰值信噪比始终未能超过 15 分贝。这表明，在这种设置下，对偶空间中的模型校正网络确实是必不可少的。

*D. PD-DEQ*模型的收敛

我们通过数值实验测试了在达到最多10次训练迭代后继续运行迭代点时PD-DEQ模型的表现。理论上，迭代点在原始变量上应具有压缩性，而对偶变量应保持一致有界。对于图2所示的测试示例，这一点可以在图5中观察到。x 中的残差确实以 $1/k$的速率收缩，而对偶变量在开始时并不收缩，但其相对范数显著更小；在最初的10次迭代之后，对偶变量也以高于 $1/k$ 的速率收缩，从而满足不动点收敛的要求。然而，最佳重建质量出现在最初的10次迭代时，之后对比度进一步降低。

VI. 结论

本研究旨在构建一种模型校正与学习型迭代重建相结合的模型，使其能够扩展到更大的图像尺寸。我们构建了模型校正的学习型原始-对偶模型（MC-PD）及其相应的深度平衡模型表述（PD-DEQ）。MC-PD能够提供出色的重建质量，但由于计算图随迭代次数线性增长，因此需要更高的内存需求。PD-DEQ具有出色的可扩展性，但定量结果较差。此外，我们还提出了PD-Hybrid方法，在降低内存需求的同时改善重建质量，从而实现二者之间的平衡。预计进一步修改混合方法可以改善定量结果。

其次，本研究旨在进一步探讨学习式迭代重建与模型校正的方法论和理论问题，以及在有限视角设置下使用深度平衡模型。尽管我们实现的 PD-DEQ 性能不及无约束版本，但我们也发现，不包含模型校正组件的深度平衡模型根本无法成功训练。这表明纳入模型校正十分重要，也为未来研究提供了有前景的方向。

参考文献

[1] Simon Arridge、Peter Maass、Ozan ¨Oktem 和 Carola-Bibiane Sch¨onlieb。使用数据驱动模型求解逆问题。《数值学报》，28:1–174，2019。[2] Vishal Monga、Yuelong Li 和 Yonina C Eldar。算法展开：用于信号与图像处理的可解释、高效深度学习。《*IEEE*信号处理杂志》，38(2):18–44，2021。.[3] Subhadip Mukherjee、Andreas Hauptmann、Ozan Oktem、Marcelo Pereyra 和 Carola-Bibiane Sch¨onlieb。具有收敛保证的学习型重建方法：概念与应用综述。《*IEEE*信号处理杂志》，40(1):164–182，2023。[4] Ulugbek S Kamilov、Charles A Bouman、Gregery T Buzzard 和 Brendt Wohlberg。用于在计算成像中整合物理模型与学习模型的即插即用方法：理论、算法与应用。《*IEEE*信号处理杂志》，40(1):85–97，2023。[5] Antonin Chambolle、Stacey E. Levine 和 Bradley J. Lucier。一种用于基于全变差的图像平滑的迎风有限差分法。*SIAM*成像科学杂志，4(1):277–299，2011。[6] Emil Y Sidky、Jakob H Jørgensen 和 Xiaochuan Pan。使用Chambolle–Pock算法进行计算机断层成像图像重建的凸优化问题原型设计。《医学与生物学中的物理学》，57(10):3065，2012。[7] Jonas Adler 和 Ozan ¨Oktem。学习型原始-对偶重建。*IEEE*医学影像学汇刊，2018。[8] Singanallur V Venkatakrishnan、Charles A Bouman 和 Brendt Wohlberg。用于基于模型的重建的即插即用先验。载于*2013 IEEE*全球信号与信息处理大会，第945–948页。IEEE，2013。[9] Kai Zhang、Yawei Li、Wangmeng Zuo、Lei Zhang、Luc Van Gool 和 Radu Timofte。结合深度去噪器先验的即插即用图像复原。《*IEEE*模式分析与机器智能汇刊》，44(10):6360–6376，2021。[10] Ernest Ryu、Jialin Liu、Sicheng Wang、Xiaohan Chen、Zhangyang Wang 和 Wotao Yin。经过适当训练的去噪器可使即插即用方法得到可证明的收敛。载于国际机器学习会议，第5546–5557页。PMLR，2019。[11] R´emi Laumont、Valentin De Bortoli、Andr´es Almansa、Julie Delon、Alain Durmus 和 Marcelo Pereyra。使用即插即用先验的贝叶斯成像：当朗之万遇上特威迪。*SIAM*成像科学杂志，15(2):701–737，2022。[12] A. Hauptmann、F. Lucka、M. Betcke、N. Huynh、J. Adler、B. Cox、P. Beard、S. Ourselin 和 S. Arridge。用于加速有限视角三维光声断层成像的基于模型的学习。*IEEE*医学影像学汇刊，2018。

[13] Amir Aghabiglou、Matthieu Terris、Adrian Jackson 和 Yves Wiaux。面向大规模高动态范围成像的深度网络级数。arXiv预印本 arXiv:2210.16060，2022。[14] Andreas Hauptmann、Ben Cox、Felix Lucka、Nam Huynh、Marta Betcke、Paul Beard 和 Simon Arridge。用于快速光声重建的近似k空间模型与深度学习。载于《国际医学图像重建机器学习研讨会》，第103–111页。施普林格，2018。[15] Sebastian Lunz、Andreas Hauptmann、Tanja Tarvainen、Carola-Bibiane Schonlieb 和 Simon Arridge。关于逆问题中的学习型算子校正。SIAM成像科学杂志，14(1):92–127，2021。[16] Danny Smyl、Tyler N Tallman、Jonathan A Black、Andreas Hauptmann 和 Dong Liu。非高斯模型误差的学习与校正。计算物理学杂志，432:110152，2021。[17] Shaojie Bai、J Zico Kolter 和 Vladlen Koltun。深度平衡模型。神经信息处理系统进展，32，2019。[18] Davis Gilton、Gregory Ongie 和 Rebecca Willett。用于成像逆问题的深度平衡架构。IEEE计算成像汇刊，7:1123–1133，2021。[19] Andreas Hauptmann、Jonas Adler、Simon Arridge 和 Ozan ¨Oktem。多尺度学习型迭代重建。IEEE计算成像汇刊，6:843–856，2020。[20] Patrick Putzky、Dimitrios Karkalousos、Jonas Teuwen、Nikita Miriakov、Bart Bakker、Matthan Caan 和 Max Welling。i-rim在fastMRI挑战赛中的应用。arXiv预印本 arXiv:1910.08952，2019。[21] Jevgenija Rudzusika、Buda Bajic、Ozan ¨Oktem、Carola-Bibiane Sch¨onlieb 和 Christian Etmann。可逆学习型原始-对偶。载于NeurIPS 2021深度学习与逆问题研讨会，2021。[22] Rafael Orozco、Mathias Louboutin 和 Felix J Herrmann。用于三维光声成像的内存高效可逆神经网络。arXiv预印本 arXiv:2204.11850，2022。[23] Jean-Christophe Pesquet、Audrey Repetti、Matthieu Terris 和 Yves Wiaux。学习用于图像复原的极大单调算子。SIAM成像科学杂志，14(3):1206–1237，2021。[24] Housen Li、Johannes Schwab、Stephan Antholzer 和 Markus Halt-meier。Nett：使用深度神经网络求解逆问题。逆问题，36(6):065005，2020。[25] P. Beard。生物医学光声成像。《界面焦点》，1(4):602–631，2011。[26] Ben Cox、Jan G Laufer、Simon R Arridge 和 Paul C Beard。定量光谱光声成像：综述。《生物医学光学杂志》，17(6):061202–061202，2012。[27] Peter Kuchment 和 Leonid Kunyansky。光声与热声断层成像数学。载于《成像数学方法手册》，第817–865页。施普林格，2011。[28] Kun Wang 和 Mark A Anastasio。光声与热声断层成像：成像原理。载于《成像数学方法手册》，2015。[29] Bradley E Treeby 和 Benjamin T Cox。k-wave：用于模拟与重建光声波场的Matlab工具箱。《生物医学光学杂志》，15(2):021314，2010。[30] Bradley E Treeby、Jiri Jaros、Alistair P Rendell 和 B T Cox。使用k空间伪谱方法模拟具有幂律吸收的非均匀介质中的非线性超声传播。《美国声学学会杂志》，131(6):4324–36，2012。[31] 线探测器有限视角光声断层成像重建算法的实验评估。逆问题，23(6):S81，2007。[32] Simon Arridge、Paul Beard、Marta Betcke、Ben Cox、Nam Huynh、Felix Lucka、Olumide Ogunlade 和 Edward Zhang。通过压缩感知实现加速高分辨率光声断层成像。《医学与生物学中的物理学》，61(24):8908，2016。[33] 一种用于光声断层成像的新型压缩感知方案。SIAM应用数学杂志，75(6):2475–2494，2015。[34] Yoeri E Boink、Marinus J Lagerwerf、Wiendelt Steenbergen、Stephan A van Gils、Srirang Manohar 和 Christoph Brune。一种用于光声断层成像中方向性与高阶重建的框架。《医学与生物学中的物理学》，63(4):045018，2018。[35] Jonas Adler 和 Ozan ¨Oktem。使用迭代深度神经网络求解病态逆问题。逆问题，33(12):124007，2017。[36] Kerstin Hammernik、Teresa Klatzer、Erich Kobler、Michael P Recht、Daniel K Sodickson、Thomas Pock 和 Florian Knoll。学习用于加速MRI数据重建的变分网络。《医学磁共振》，79(6):3055–3071，2018。[37] Jiaming Liu、Rushil Anirudh、Jayaraman J Thiagarajan、Stewart He、K Aditya Mohan、Ulugbek S Kamilov 和 Hyojin Kim。Dolce：一种用于有限角度CT重建的基于模型的概率扩散框架。arXiv预印本 arXiv:2211.12340，2022 年。[38] Buda Baji´c 、 Ozan ¨Oktem 和 Jevgenija Rudzusika。采用内存高效可逆学习型原始-对偶方法的三维螺旋CT重建。arXiv电子预印本，第 arXiv–2205 页，2022年。[39] Nikita Moriakov、Jan-Jakob Sonke 和 Jonas Teuwen。Lire：用于锥束CT的学习型可逆重建。arXiv预印本 arXiv:2205.07358，2022年。[40] Yoeri E Boink、Srirang Manohar 和 Christoph Brune。一种用于光声联合重建与分割的部分学习型算法。IEEE医学影像学汇刊，39(1):129–139，2019年。[41] Ko-Tsung Hsu、Steven Guan 和 Parag V Chitnis。比较用于光声断层成像图像重建的深度学习框架。《光声学》，23:100271，2021年。[42] Andreas Hauptmann 和 Ben T Cox。光声断层成像中的深度学习：当前方法与未来方向。《生物医学光学杂志》，25(11):112903，2020年。[43] Janek Gr¨ohl、Melanie Schellenberg、Kris Dreher 和 Lena Maier-Hein。生物医学光声成像中的深度学习：综述。《光声学》，22:100241，2021年。[44] Changchun Yang、Hengrong Lan、Feng Gao 和 Fei Gao。光声成像深度学习综述。《光声学》，21 100215，2021： 。[45] K P Koestli、M Frenz、H Bebie 和 H P Weber。采用傅里叶变换方法对光声压力瞬变进行时间反向投影。Phys Med Biol，46(7):1863–1872，2001年。[46] B. Cox 和 P. Beard。采用k空间方法快速计算流体中的脉冲光声场。J. Acoust. Soc. Am.，117 3627，2005 6 3616( ): – 。[47] Ko-Tsung Hsu、Steven Guan 和 Parag V Chitnis。采用学习型物理模型进行光声断层成像快速迭代重建：理论验证。《光声学》，第100452页，2023年。[48] Steven Guan、Ko-Tsung Hsu 和 Parag V Chitnis。用于快速光声波仿真的傅里叶神经算子网络。算法，16(2):124，2023年。[49] Leon Bungert、Martin Burger、Yury Korolev 和 Carola-Bibiane Sch¨onlieb。针对前向算子不完善且噪声模型一般化的逆问题的变分正则化。逆问题，36(12):125014，2020年。[50] Arttu Arjas、Mikko J Sillanp¨a¨a 和 Andreas Hauptmann。非线性逆问题的序贯模型校正。arXiv预印本 arXiv:2301.10094，2023年。[51] Ruanui Nicholson、Noemi Petra、Umberto Villa 和 Jari P Kaipio。关于非线性贝叶斯逆问题的全局正态线性近似。逆问题，2023年。[52] Leonid Kunyansky。在具有柱对称或球对称性的特定区域中进行热声断层成像的快速重建算法。逆问题与成像，6(1):111–131，2012年。[53] Samy Wu Fung、Howard Heaton、Qiuwei Li、Daniel McKenzie、Stanley Osher 和 Wotao Yin。JFB：隐式网络的无雅可比反向传播。载于《AAAI人工智能大会论文集》，第36卷，第6648–6656页，2022年。[54] Dirk A Lorenz 和 Felix Schneppe。采用伴随失配的Chambolle–Pock原始-对偶方法。《应用数学与优化》，87(2):22，2023年。[55] Emilie Chouzenoux、Andr´es Contreras、Jean-Christophe Pesquet 和 Marion Savanier。伴随失配条件下原始-对偶算法的收敛结果。SIAM成像科学杂志，16(1):1–34，2023年。[56] Heinz H Bauschke、Sarah M Moffat 和 Xianfu Wang。牢固非扩张映射与极大单调算子：对应关系与对偶性。《集值与变分分析》，20:131–153，2012年。[57] Subhadip Mukherjee、S¨oren Dittmer、Zakhar Shumaylov、Sebastian Lunz、Ozan ¨Oktem 和 Carola-Bibiane Sch¨onlieb。用于逆问题的学习型凸正则化项。arXiv电子预印本，第 arXiv–2008 页，2020年。