用于磁共振成像重建与合成的基于优化的深度学习方法

arXiv:2303.01515v1 [math.OC] 2 Mar 2023

作者:
Wanyu Bian

提交给佛罗里达大学研究生院的学位论文，作为满足哲学博士学位要求的一部分

佛罗里达大学

2022

© 2022年 Wanyu Bian

谨以此学位论文献给我的家人和朋友。

致谢

感谢我的导师 Yunmei Chen，感谢她在过去五年中的所有帮助与激励。感谢我的朋友和家人。

目录

page

致谢.4表格目录.8图目录.9摘要.11章节

1 引言.13

1.1 磁共振成像中压缩感知的优化模型.141.2 用于磁共振成像重建的可学习优化算法 . . . . . 151.3 用于参数训练的双层优化算法 .191.4 学位论文概述 .21

2 一种无需线圈灵敏度的基于优化的深度并行磁共振成像重建网络. . . 24

2.1 引言.242.2 相关工作.262.3 所提出的方法.272.3.1 无线圈灵敏度的联合通道图像重建并行磁共振成像 272.3.2 基于残差学习的近端梯度网络 292.3.3 网络架构与训练.302.4 实验结果.322.5 结论.36

3 一种无需线圈灵敏度的联合通道并行磁共振成像重建最优控制框架.37

3.1 引言.373.2 相关工作.403.3 所提出的方法.443.3.1 背景.443.3.2 问题设置.453.3.3 变分模型中的正则化设计 . . . . . 493.3.4 从拉格朗日乘子法（MLM）的角度进行网络训练.533.4 实验结果.55

3.4.1 数据集.553.4.2 实现.56

3.4.3 与现有方法的比较.583.4.4 消融研究.61

3.4.4.1 组合算子与平方和平方根的比较 . . . . . 623.4.4.2 初始化.
633.4.4.3 复数卷积.653.4.4.4 与所提出的图像域重建和域混合重建的比
较.65

3.5 结论 66

4 基于优化的元学习模型：使用 DiverseDataset 进行磁共振成像重建.67

4.1 引言.684.2 相关工作.704.3 所提方法.724.3.1 预备知识.724.3.2 LOA诱导重建网络.
744.3.3 用于网络训练的双层优化算法 . . . . . 784.4 实现.824.4.1 特征提取算子.
824.4.2 设置.834.5 数值实验.834.5.1 数据集.834.5.2 实验设置.844.5.3 径向掩模中不
同压缩感知比率下的实验结果 . . 874.5.4 不同采样模式下不同未见压缩感知比率的实
验结果.904.5.5 未来工作与开放挑战.994.6 结论.994.7 收敛性分析.100

5 基于优化的模型：多模态磁共振成像重建与合成
108

5.1 引言.1085.2 所提出的方法.1115.2.1 模型.1115.2.1.1 模态特定特征提取算子 . . . . .
. . 1135.2.1.2 多模态合成模块 .1135.2.2 高效可学习优化算法 .1135.2.3 收敛性分析 .
1165.2.4 用于网络训练的双层优化算法 . . . . . 1185.3 实验.1195.3.1 初始化.
1195.3.1.1 初始化网络.120

5.3.1.2 初始化网络训练.1205.3.2 实验设置.1205.3.3 超参数选择.121

5.3.4 性能评估.1225.4 结论.124参考文献.126个人简介.145

表格目录

表格第2页-1 冠状位 FSPD 数据重建的定量测量值.....32 2-2 冠状位质子密度数据重建的一定量测量值.....32 3-1 本文中通用的一些符号及其含义.....46 3-2 各阶段 FSPD 数据和 PD 数据的测试平均峰值信噪比.....58 3-3 冠状位FSPD与PD数据重建的定量评估，以及各参考方法的重建时间。时间（秒）指每种方法的测试时间。.....60 3-4 比较重建多线圈图像的方法时，PD 数据中 u 与 u^* 之间的均方误差。.....61 3-5 所提并行磁共振成像重建网络各变体的标签。.....62 3-6 使用所提方法的不同变体对冠状位FSPD与PD数据进行重建的定量评估。.....62 4-1 使用不同径向掩模采样率重建 T1脑图像的定量评估。.....92 4-2 使用不同径向掩模采样率重建 T2脑图像的定量评估。.....92 4-3 使用不同径向掩模采样率重建 T1脑图像的定量评估。.....93 4-4 使用不同径向掩模采样率重建 T2脑图像的定量评估。.....93 4-5 使用不同随机笛卡尔掩模采样率重建 T2脑图像的定量评估。.....93 5-1 最先进的多模态合成方法与所提方法的合成结果定性比较。.....123 5-2 所提模型采用联合重建，与仅合成和重建 T1 和 T2 的重建结果定量比较。.....124

图目录

图 page

2-1 第 $+1$阶段的网络结构范式。Re$\mathbf{b}^{(t+1)}$、Re$\mathbf{u}^{(t+1)} \in \mathbb{R}^{m\times n\times N_c}$和 Im$\mathbf{b}^{(t+1)}$、Im$\mathbf{u}^{(t+1)} \in \mathbb{R}^{m\times n\times N_c}$分别表示 $\mathbf{b}^{(t)}$ 和$\mathbf{u}^{(t)}$的实部和虚部。.....3

12-2 各卷积算子的结构。网络中实部和虚部的权重不共享，但实部和虚部卷积算子的结构相同.312-3 规则笛卡尔采样（采样率为 31.56%）下冠状位 FSPD 膝关节图像的结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、deepcomplexMRI、pMRI-Net 和真实值.33

2-4 规则笛卡尔采样（采样率为 31.56%）下冠状位 PD 膝关节图像的结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、pMRI-Net 和真实值.33

2-5 规则笛卡尔采样（采样率为 31.56%）下冠状位 FSPD 膝关节图像的更多实验结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、所提出的方法和真实值.3
4

2-6 规则笛卡尔采样（采样率为 31.56%）下冠状位 PD 膝关节图像的更多实验结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、所提出的方法和真实值.

343-1 所提出的适用于所有阶段的框架范式.52

3-2 所提出的框架范式在第 $t+1$ 阶段。 $W = Re(W) + iIm(W)$表示复数卷积滤波器核.5

33-3 所提方法的重建通道合并图像 $\mathcal{J}(\bar{\mathbf{u}}(t))$ 对于 $t = 1$、2、3、4 阶段.583-4 冠状位 PD 膝关节图像上各重建方法的定性比较结果.603-5 冠状位 FSPD 膝关节图像上各重建方法的定性比较结果.60

3-6 采用三种不同初始重建方法（包括零填充初始值、SPIRiT 重建初始值和学习得到的初始值）的所有阶段框架范式.

633-7 第一行（从左到右）显示初始值 $\bar{\mathbf{u}}(0)$.643-8 不同方法重建的线圈图像中某一线圈的相位信息，并与所提方法和参考图像进行比较.65

4-1 从上到下依次为：T2脑图像重建结果、放大细节、带色标的逐点误差以及相应的 **径向**掩模。右上方图像为真实值全采样图像 .944-2 从上到下依次为：元学习、传统学习和ISTA-Net的重建结果、放大细节、带色标的逐点误差以及相应的 **径向**掩模+.95

4-3 图像（从上到下）展示了T1脑图像重建结果、放大细节、带有色标的逐点误差，以及相应的**径向**掩模.96

4-4 图像（从上到下）展示了T2脑图像重建结果、放大细节、带有色标的逐点误差，以及相应的**径向**掩模.9
7

4-5 图像（从上到下）展示了T2脑图像重建结果、放大细节、带有色标的逐点误差，以及相应的**笛卡尔**掩模.
98

4-6 使用径向掩模、四种不同压缩感知比率进行 T1加权磁共振成像图像重建时算法的收敛行为2 。**左图**：目标函数值ϕ 与相位数的关系。**右图**：峰值信噪比值与相位数的关系。。。。。。。

995-1 特征张量 h_{w_i}和特征向量 $h_{w_i,\ j}$位于空间位置 j。。。。。。1145-2 合成映射 g_θ 将拼接特征 $[h_{w_1},\ h_{w_2}]$ 映射为图像$\mathbf{x}_3$.114

5-3 所提出的网络架构。我们将 L^l 定义为方程（5–10）中定义的恒等损失函数。底部橙色框中的损失函数（记为 ℓ）定义于（5–6）中.115

5-4 最先进的多模态合成方法与所提方法之间的定性比较。从第一行到最后一行：T1 + T2 → 液体衰减反转恢复，T1 +液体衰减反转恢复 → T2加权成像，T2 + 液体衰减反转恢复 → T1加权成像和T1 + T2 → T1CE.124

5-5 合成图像与其对应真实值之间的逐点误差图。从上至下各行依次为：T1 + T2 → 液体衰减反转恢复、T1 + 液体衰减反转恢复 → T2加权成像、T2+ 液体衰减反转恢复 → T1加权成像和 T1 + T2 → T1CE.125

提交佛罗里达大学研究生院的学位论文摘要，旨在部分满
足哲学博士学位的要求

用于磁共振成像重建与合成的基于优化的深度学习方法

By

Wanyu Bian

2022年8月

主席：Yunmei
Chen 专业：数学

本学位论文致力于提出先进的磁共振成像（MRI）重建非凸非光滑变分模型、高效且可学习的图像重建算法以及参数训练算法，以提高基于优化的深度学习方法在压缩感知磁共振成像重建与合成中的准确率和鲁棒性。

第一部分介绍了一种新型的基于优化的深度神经网络，其架构受用于求解无须了解线圈灵敏度图的并行磁共振成像重建问题变分模型的近端梯度下降启发。正则化函数由非线性组合算子和稀疏特征编码器构成。

第二部分是在第一部分初步工作的基础上进行的重要扩展，通过在离散时间最优控制框架下解决无校准快速并行磁共振成像重建问题。网络架构由指定变分模型所诱导的离散时间动力系统确定。变分模型中的正则化包含两部分：一部分用于增强重建图像在图像域中的稀疏性；另一部分作用于 k 空间域，以进一步去除高频伪影。

第三部分旨在元学习框架下开发一种具有良好泛化能力的磁共振成像（MRI）重建方法。具体而言，我们开发了一种深度

重建网络，该网络由可学习优化算法（LOA）驱动，用于求解MRI图像重建的非凸非光滑变分模型。我们将这些网络参数划分为两部分：一部分是正则化中通用特征编码器组件的任务不变部分，另一部分是任务特定部分，用于应对异构训练和测试数据中的差异。

最后一部分旨在利用源模态的部分扫描 k空间数据来合成 MRI 的目标模态，而不是使用最先进的多模态合成中所采用的完整扫描数据。我们提出为这三种模态分别学习一个模态特定特征提取算子。随后，我们结合这些已学习的算子和一个鲁棒稀疏特征选择算子，为这些图像设计正则化项。为了利用源模态合成目标模态图像，我们采用另一个特征融合算子，该算子学习从源模态生成的特征到目标模态之间的映射。

第1章 引言

MRI 是最重要的无创、非电离医学成像技术之一，可生成精确的活体组织图像，用于疾病诊断和医学分析。磁共振成像领域的一项主要关注点，是在磁共振扫描期间加快数据采集速度。根据扫描身体部位的不同，获取一个序列的磁共振图像通常至少需要 30 分钟，这一时间明显长于大多数其他成像模态。然而，婴儿、老年人以及患有严重疾病且无法控制身体运动的患者，可能无法在长时间扫描期间保持安静。扫描过程延长会使患者感到不适，或在磁共振图像中引入运动伪影，从而降低诊断可及性。因此，加快磁共振扫描对于改善磁共振图像质量至关重要且意义重大。此外，MRI 能够从同一解剖结构中提供不同的对比图像，从而丰富临床诊断和研究所需的解剖信息。不同对比度的磁共振图像具有相似的解剖结构，但会突出显示不同的软组织，并且这些图像是在不同序列中扫描获取的，因此，为获得额外的磁共振序列，采集时间可能会进一步延长。

MRI 数据是在频域（k空间）中采集的，采集时间大致与 k 空间相位编码步数成正比。加快 MRI 扫描的常见方法是跳过相位编码线，减少 k空间测量值的数量，只采集部分 k 空间数据，从而减少数据采集期间的采样信号数量。然而，这违反了奈奎斯特准则 [137]，并且由于欠采样而产生混叠伪影。

磁共振成像重建是从欠采样的部分 k 空间数据中恢复清晰磁共振图像的过程，可用于诊断和临床应用。压缩感知（CS） [47] 磁共振成像重建和磁共振并行成像 [164, 142, 97]

是解决逆问题、加快磁共振成像扫描并消除伪影的有效方法。

本章首先介绍快速磁共振成像欠采样与重建的背景；然后概述压缩感知磁共振成像和并行成像中的经典重建方法及基于深度学习的方法；最后介绍过去几十年提出的几种知名方法。

1.1 磁共振成像中的压缩感知优化模型

设磁共振图像为$\mathbf{x} \in \mathbb{C}^{\sqrt{N}\times\sqrt{N}}$，它是一个二维（2D）数组，其中图像中有$N$个像素$\mathbf{x}$。在数值算法中，通常且方便地将$\mathbf{x}$表示为维度为$N$的列向量，并且$\mathbf{x}_i$是图像$x$的第**个强度值，**其中$i \in \{1, \cdots, N\}$。欠采样k空间测量值与图像之间的关系如下：[67]:

$$\mathbf{y} = \mathbf{P}\mathcal{F}\mathbf{x} + \mathbf{n}, \tag{1–1}$$

其中$\mathbf{y} \in \mathbb{C}^p$是k空间中的测量值，总共有$p$个采样数据点，$\mathbf{x} \in \mathbb{C}^{N\times 1}$是待重建的磁共振图像，包含$N$个像素，$\mathcal{F} \in \mathbb{C}^{N\times N}$是二维离散傅里叶变换（DFT）矩阵，而$\mathbf{P} \in \mathbb{R}^{p\times N}$($p < N$)是表示k空间采样轨迹的二值矩阵。$\mathbf{n}$表示k空间中的数据采集噪声。磁共振成像重建的目标是根据方程（$\mathbf{x}$从方程（1–1），利用部分k空间数据$\mathbf{y}$求解

考虑到$\mathbf{y}$不含噪声，因此$\mathbf{n} = 0$，我们可以得到$\mathbf{P}\mathcal{F}\mathbf{x} = \mathbf{y}$，其中运算$\mathbf{P}\mathcal{F} \in \mathbb{C}^{p\times N}$满足$p << N$，这是一个欠定系统，会产生无穷多个复数解。最直接的解是零填充重建方法，其公式为$\mathbf{x} = \mathcal{F}^{-1}\mathbf{P}^{-1}\mathbf{y}$，本质上是将未采集的k空间数据填零，以获得伪全 k 空间数据$\mathbf{P}^{-1}\mathbf{y}$，然后进行逆傅里叶变换 $\mathcal{F}^{-1}$。然而，该方法会导致截断效应，表现为吉布斯振铃 伪影和模糊 [197]。

对于并行磁共振成像中的多线圈采集，SENSE [142]利用预先计算的对角矩阵（称为第 j 个线圈的线圈灵敏度图 $S_j \in \mathbb{R}^{N\times N}$）在图像域重建并行磁共振成像数据；该矩阵可以预先给定或估计。$P\mathcal{F}(S_j \cdot x) \in \mathbb{C}^p$是欠采样的傅里叶系数，第 j 个接收线圈的k空间数据采集可表示为 $y_j = P\mathcal{F}(S_j \cdot x) + n_j$，其中 $j = 1, \cdots, N_c$；$\cdot$表示逐元素相乘，n_j 表示第 j 个接收线圈在k空间中的测量噪声。

由于逆问题（1–1）具有病态性，因此通过纳入待重建图像的先验信息来设计适当的正则化非常必要。压缩感知（CS）理论表明，当信号$\mathbf{x}$ 是稀疏的或可稀疏表示时，只需少得多的测量值即可重建出质量得以良好保留的图像。给定原始数据测量值$\mathbf{y}$，用于重建原始单线圈磁共振图像的通用压缩感知模型$\mathbf{x}$ 可写为：

$$\min_x \|\mathbf{P}\mathcal{F}\mathbf{x} - \mathbf{y}\|^2 + \mu R(\mathbf{x}), \tag{1–2}$$

用于并行磁共振成像重建的基于 SENSE 的模型可写为：

$$\min_x \sum_{j=1}^{N_c} \frac{1}{2}\|\mathbf{P}\mathbf{F}(\mathbf{s}_j \cdot \mathbf{x}) - \mathbf{f}_j\|^2 + \mu R(\mathbf{x}), \tag{1–3}$$

其中，正则化项R 提供重建图像的先验信息，而 $\mu > 0$ 是用于平衡数据保真项与正则化项的权重参数。过去几十年来，全变分（TV）一直被广泛认为是一种有效的正则化项，其形式为$TV(\mathbf{x}) = \sum^N{}_{i=1} \|D_i\mathbf{x}\|$ 其中$D_i \in \mathbb{R}^{2\times N}$是一个二值矩阵，其元素为 -1 和 1，分别对应于沿第一和第二坐标方向的偏导数的前向有限差分。由于 TV 是梯度$D_i\mathbf{x}\|$的l_1 范数，因此在压缩感知方法 [155]中常将其用作稀疏变换。

1.2 用于磁共振成像重建的可学习优化算法

成功的单线圈磁共振成像重建传统方法包括但不限于以下方法：1）基于全变分的模型 [117, 122, 198, 221, 181];

用于求解TV正则化图像重建问题的几种常用算法，包括交替方向乘子法（ADMM）[20, 62, 198],原始-对偶混合梯度（PDHG）算法 [23, 52, 229] 以及迭代收缩阈值算法（ISTA） [10]。2）基于小波的模型 [66, 26] ，其利用小波域中的稀疏性假设。3）基于低秩的模型 [121, 46]，常用于动态磁共振成像。4）基于图像块或非局部的模型 [206, 178, 50] ，通过捕捉图像块之间的几何自相似性来考虑稀疏表示。5）基于字典学习的模型 [151, 216, 84] ，利用字典学习进一步提升信号的稀疏性，等等。6）并行成像包含两类：使用逐线圈自动校准的k空间方法，如GRAPPA [65], SPIRiT [118]，以及诸如SENSE的图像域方法 [142] ，其求解一个需要完美线圈灵敏度图知识的优化问题。

传统方法采用手工设计的正则化项，其求解算法遵循理论论证。然而，这些正则化项过度简化。例如，TV正则化模型往往会重建出“分片常数”的图像，而由于其促使梯度稀疏，精细的结构细节可能会变得模糊。此外，相关参数难以调节，而且通常需要数百甚至数千次迭代才能收敛，以捕捉细微细节并满足临床诊断质量要求 [34]。

基于深度学习的模型利用大型数据集，进一步挖掘相较于传统方法提升重建性能的潜力，并已成功应用于临床领域。然而，当数据稀缺时，训练通用深度神经网络（DNN）可能会导致过拟合。此外，深度网络结构如同黑箱，缺乏数学层面的解释。为提升深度模型拓扑结构与重建结果之间关系的可解释性，一类新兴的基于深度学习的方法——可学习优化算法 （LOA）——受到了广泛关注

例如， [116, 111, 157, 124, 223, 162, 24, 6, 113, 111]。LOA旨在将现有优化算法映射为结构化网络，其中网络的每个阶段对应于优化算法的一次迭代。

例如，近端点网络 [14, 16] 不需要指定正则化的形式，并进行以下两个步骤的迭代：

$$r_k = \mathbf{x}_{k-1} - \alpha_k \nabla f(\mathbf{x}_{k-1}), \tag{1–4a}$$

$$\mathbf{x}_k = \mathrm{prox}_{\alpha_k R}(r_k), \tag{1–4b}$$

其中 $\alpha > 0$ 为步长，函数f是式（1–1）中的数据保真度。邻近算子 $\mathrm{prox}_{\alpha_k R}$ 可参数化为深度可学习去噪器，也可替换为去噪网络 [219]。第 2章和第 3章将详细介绍近端点网络及其一些变体。在 ADMM-Net [205], 中，深度重建网络架构的每个阶段都模拟 ADMM算法的一次迭代，其中 ADMM-Net 将离散梯度算子 D 替换为卷积算子，该算子是给定滤波器组的线性组合。

图像重建问题的压缩感知理论通常要求在正则化中使用稀疏变换$\Psi \in \mathbb{R}^{n\times n}$ （例如：小波变换），因为自然图像通常并不稀疏，单个 ℓ_1 并不足够。因此，我们可以得到一个不同的重建模型：

$$\min_{\mathbf{x}} f(\mathbf{x}) + \mu R(\Psi \mathbf{x}), \tag{1–5}$$

其中，我们将f写作数据保真项，而 $\Psi\mathbf{x}$ 是一个稀疏向量。ISTA-NET$^+$ [218]考虑求解 $\mathbf{u} = \Psi\mathbf{x}$ ，并重建$\mathbf{x} = \Psi^\top \mathbf{u}$。变换器 Ψ 和 $\Psi^\top$ 分别由多层卷积神经网络$H^{(k)}$ 和 $\tilde{H}^{(k)}$ 替代，然后迭代执行以下方案：

$$r_k = \mathbf{x}_{k-1} - \alpha_k \nabla f(\mathbf{x}_{k-1}), \tag{1–6a}$$

$$\mathbf{x}_k = \tilde{H}^{(k)}\, \mathrm{prox}_{\alpha_k R}(H^{(k)} r_k) = \tilde{H}^{(k)} S_{\theta_k}(H^{(k)} r_k), \tag{1–6b}$$

其中，S_{θ_k} 表示收缩算子，因为$R(\mathbf{u}) = \mu\|\mathbf{u}\|_1$，且 $\theta_k = \alpha_k\mu$。网络$H^{(k)}$ 和 $\tilde{H}^{(k)}$ 采用对称结构进行参数化，但其参数分别独立学习。在第2和3章提出的迭代算法网络架构+源自ISTA-NET[3, 35, 75, 126] 。原始-对偶网络（PD-Nets） [23]是深度网络，其中每个连续阶段都模拟原始-对偶混合梯度算法的相应迭代 [35] 。PD-Net解决基于压缩感知的磁共振成像重建问题，其中原始近端算子和对偶近端算子由多层可学习卷积神经网络去噪器替代。

上述可学习优化算法（LOA）只是对迭代算法的似是而非的模仿，因此缺乏变分模型的理论基础，也没有任何收敛保证。例如，为了避免求解非光滑非凸问题的困难，使用卷积神经网络替代近端算子 [35, 14, 218], 矩阵变换 [203, 69, 218], 非线性算子 [203, 69], 以及去噪器/正则化器 [4, 160] 等。这些方法只是将网络嵌入优化迭代的展开方案中，因此其网络结构缺乏收敛性分析和可解释性，也容易出现过拟合或欠拟合问题。

第4 章和第5 章提出的用于图像重建的网络，旨在通过执行具有收敛保证的可学习优化算法（LOA）来解决这一问题，其中网络结构严格遵循该算法。由于（1–2）中的最小化问题是非光滑且非凸的，我们首先对正则化项中的$\ell_{2,1}$ 范数进行平滑处理，然后设计具有可证明收敛性的算法展开方法。在本学位论文中，我们将 LOA 视为前向图像重建优化过程，同时分析训练过程中用于更新可学习参数的反向参数优化问题。通过最小化损失函数获得迭代训练算法是机器学习中的一个主要问题，而用于参数训练的双层优化算法则是本学位论文的重点考虑内容。

1.3 双层优化参数训练算法

双层优化可表述为以下形式：

$$\min_{u} \phi(v^*(u), u) \tag{1–7a}$$

$$\text{subject to } v^*(u) = \arg\min_{v} \varphi(v, u), \tag{1–7b}$$

其中（1–7a）是上层/外层优化问题，而受约束问题（1–7b）称为下层/内层优化问题。下层的解是上层的一个自变量。

在传统的监督机器学习中，训练数据以输入和输出对的集合形式给出 $\mathcal{D} = \{(\mathbf{y}_m, \mathbf{x}_m^*)\}_{m=1}^{\mathcal{M}}$，我们希望训练一个预测模型$\phi_\theta$，该模型由参数 $\theta \in \Theta$ 参数化，并将输入$\mathbf{y}_m \in \mathcal{Y}$映射到相应的输出 $\mathbf{x}_m^* \in \mathcal{X}$。参数 θ 通过最小化预定义的损失函数$L: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$进行学习，该损失函数用于衡量输出 $F_\theta(\mathbf{y}_m)$与参考值$\mathbf{x}_m^*$之间的差异，这里我们记作

$$\ell(\theta; \mathcal{D}) := \sum_{m=1}^{\mathcal{M}} L(F_\theta(\mathbf{y}_m), \mathbf{x}_m^*). \tag{1–8}$$

传统监督学习的目标是 ^找到最优参数 θ，使得

$$\hat{\theta} = \arg\min_{\theta} \ell(\theta; \mathcal{D}). \tag{1–9}$$

近年来的机器学习研究中，提升机器学习模型的泛化能力受到广泛关注。一些监督学习模型，包括深度双层学习 [89]和基于优化的元学习[54, 149, 110, 135, 156]，是专门用于提升模型泛化能力并防止估计模型过拟合的有效策略。

在本论文中，我们更加关注元学习。在有监督元学习中，元知识 ω 被引入，用于指导“如何学习”，并通过为一般任务学习算法来提高学习新任务的泛化能力。预先指定的

假设（例如学习率、初始参数和优化器的选择）可以在元学习中有所不同 [85]。目标是在任务分布$p(\mathcal{T})$上寻找最优元知识 ω，这也被称为跨任务知识 [78]。一般设置可分为元训练、元验证和元测试，这些阶段分别对应互不相交的任务集。我们将 $\mathcal{T}_i = (\mathcal{D}_{\tau_i}^{tr}, \mathcal{D}_{\tau_i}^{val}, \mathcal{D}_{\tau_i}^{test})$、$\mathcal{T}_i \sim p(\mathcal{T})$记为，元训练阶段使用数据集 $\{(\mathcal{D}_{\tau_i}^{tr}, \mathcal{D}_{\tau_i}^{val})\}_{i=1}^{P}$，其中包含$P$个任务，每个任务都有训练集和验证集。元测试阶段或评估阶段使用包含Q个任务的数据集 $\{(\mathcal{D}_{\tau_i}^{tr}, \mathcal{D}_{\tau_i}^{test})\}_{i=1}^{Q}$，此时允许模型在训练集上针对单个任务更新参数，随后可在任务特定参数更新后使用测试集评估模型性能。元训练过程可以表述为一个双层优化问题：

$$\hat{\omega} = \arg\min_{\omega} \sum_{i=1}^{P} \ell_{\tau_i}(\theta_i(\omega), \omega; \mathcal{D}_{\tau_i}^{val}) \tag{1–10a}$$

$$s.t. \quad \theta_i(\omega) = \arg\min_{\theta} \ell_{\tau_i}(\theta, \omega; \mathcal{D}_{\tau_i}^{tr}), \tag{1–10b}$$

其中 ℓ_{τ_i} 是与任务 τ_i相关联的任务感知损失函数。（1–10b）是内层/下层/基础层优化，它从相应的训练集中为每个单独的任务求解特定参数，并固定任务特定参数 ω。（1–10a）是外层/上层/元层优化，通过最小化验证损失来学习超参数 ω。

近年来，元学习方法结合不同技术，在各个领域展现出了良好的效果 [78]。元学习技术可分为三类 [209, 101, 85]：基于度量的方法 [95, 184, 163]、基于模型的方法 [131, 150, 143, 64], 和基于优化的方法 [54, 149, 110]。基于优化的方法通常被表述为双层优化问题，并且对于更广泛的任务分布表现出相对更强的泛化能力。本文主要关注基于优化的元学习。

对于元学习应用，下层问题会遇到新任务，并尝试根据训练观测快速学习相关特征；外层则在以往任务中积累任务特定元知识，而元学习器为内层提供支持，使其能够快速适应新任务。例如，在许多现有的基于优化的元学习应用中，下层问题通过一个或几个梯度下降步进行近似，例如模型无关元学习（MAML） [54], 以及大量后续提出的 MAML 相关工作，这些工作采用类似策略来改善泛化能力 [101, 156, 56, 63, 134, 186, 208, 210]。一些变体模型 [156, 186, 208] 专注于多个初始条件下的快速学习，这通常依赖于固定的优化器，例如带动量的随机梯度下降或其变体。深度双层学习 [89] 旨在获得比在单个任务上训练时更好的泛化能力，并能够良好地泛化到另一个任务。该模型用于优化正则化损失函数，从训练集中寻找网络参数并确定超参数，使网络在验证数据集上表现良好。

或者，面向优化器的方法 [7, 150, 104, 196] 侧重于学习内层优化器，元知识 ω 可用于为每次基础学习迭代定义基于梯度的优化步骤。超参数优化（HO）与元学习具有相同的优点，主要区别在于，HO 通常考虑单个任务，并将其划分为训练数据和验证数据；内层目标是在训练数据上定义的正则化经验损失函数，旨在调整模型参数，而外层目标则旨在调整超参数 [58, 139, 57, 129]。

1.4 学位论文概述

以下章节安排如下：

第2章介绍了一种新颖的深度神经网络架构：将用于并行磁共振成像（pMRI）快速图像重建的鲁棒近端梯度方案映射为正则化函数，并利用数据进行训练。所提出的网络无需了解灵敏度图，而是学习自适应地组合来自

不完整并行磁共振成像数据的多线圈图像，生成一幅具有均匀对比度的单一图像，然后将其输入非线性编码器，以高效提取图像的稀疏特征。

第3章旨在开发一种新颖的无校准快速并行磁共振成像（pMRI）重建方法，并将其与离散时间最优控制框架相结合。该重建模型旨在学习一种正则化方式，通过利用多线圈图像各通道之间的信息共享来融合通道并提取特征。重建网络被构建为一个结构化离散时间最优控制系统，从而将参数训练表述为一种最优控制形式，其中目标函数的参数充当控制变量。我们证明，用于求解控制问题的拉格朗日方法等价于反向传播，从而确保训练算法的局部收敛。

第4章旨在开发一种基于元学习框架的通用型磁共振成像重建模型。元学习中的标准基准测试面临着在多样化任务分布上进行学习的挑战。所提出的网络通过利用异构数据集，学习变分模型中的正则化函数，并重建具有各种欠采样率或采样模式的磁共振图像；这些采样率或模式可能在训练数据中出现过，也可能未曾出现。

生成同一解剖结构的多对比度/多模态 MRI 可丰富诊断信息，但在实践中受限于过长的数据采集时间。第 5章提出了一种新颖的深度学习模型，利用多个源模态的不完整 k 空间数据作为输入，实现多模态 MRI 的联合重建与合成。模型输出包括源模态的重建图像，以及目标模态中的高质量合成图像。所提出的模型被表述为一个变分问题，并结合了多个可学习的模态特定特征提取器和一个多模态合成模块。我们提出了一种可学习优化算法来求解该模型，由此构建出一个多阶段网络，其参数可利用多模态 MRI 数据进行训练。此外，我们采用双层优化框架，以实现稳健的

参数训练。我们通过大量数值实验展示了所提方法的有效性。

第2章 一种无需线圈灵敏度的基于优化的深度并行磁共振成像重建网络

2.1 引言

本章提出一种新颖的深度神经网络架构，将用于并行磁共振成像（pMRI）快速图像重建的鲁棒近端梯度方案映射为网络，并采用从数据中训练得到的正则化函数。所提出的网络能够自适应地将不完整并行磁共振成像数据中的多线圈图像组合成一幅具有均匀对比度的图像，再将其输入非线性编码器，以高效提取图像的稀疏特征。与大多数现有的深度图像重建网络不同，我们的网络无需灵敏度图信息；灵敏度图难以准确估计，且一直是现实并行磁共振成像应用中图像重建的主要瓶颈。实验结果表明，我们的方法在多种并行磁共振成像数据集上表现出良好的性能。

并行磁共振成像（pMRI）是一种先进的医学磁共振成像技术，它通过在扫描对象周围布置多个接收线圈并行采集k空间（傅里叶）数据。为加快扫描过程，pMRI采用部分数据采集，增加k空间中读出线之间的间距。然而，k空间数据采样减少会导致图像中出现混叠伪影，必须通过图像重建过程予以去除。pMRI图像重建主要有两种方法：第一种是k空间方法，利用多个接收线圈采集的数据对未采样的k空间数据进行插值[45], ，例如广义自校准部分并行采集（GRAPPA）[65]。另一种是图像空间方法，通过求解一个方程组来去除图像域中的混叠伪影；该方程组通过线圈灵敏度将待重建图像与部分k空间数据关联起来，例如灵敏度编码（SENSE） [142]。

本文提出一种基于深度学习的新型重建方法，以解决图像空间中并行磁共振成像重建的若干关键问题。考虑一个具有N_c 个接收线圈的并行磁共振成像系统，以分辨率$m \times n$ 采集二维磁共振图像（下文中，我们将二维图像$\mathbf{v} \in \mathbb{C}^{m\times n}$ 及其列向量形式$\mathbf{v} \in \mathbb{C}^{mn}$ 视为可互换的表示）。令$\mathbf{P} \in \mathbb{R}^{p\times mn}$ 为表示欠采样掩膜的二值矩阵，其在k空间中有p 个采样位置；令$\mathbf{s}_i \in \mathbb{C}^{mn}$ 表示线圈灵敏度，令$\mathbf{f}_i \in \mathbb{C}^p$ 表示第i个接收线圈在$i = 1$、...、N_c处的部分 k空间数据。因此，$\mathbf{f}_i$与图像$\mathbf{v}$之间的关系为$\mathbf{f}_i = \mathbf{PF}(\mathbf{s}_i \cdot \mathbf{v}) + \mathbf{n}_i$，其中 $\cdot$ 表示两个矩阵的逐点乘法，$\mathbf{n}_i$ 是每个接收线圈在k空间中的未知采集噪声。于是，基于SENSE的图像空间重建方法通常可表述为一个优化问题：

$$\min_{\mathbf{v}} \sum_{i=1}^{N_c} \frac{1}{2}\|\mathbf{PF}(\mathbf{s}_i \cdot \mathbf{v}) - \mathbf{f}_i\|^2 + R(\mathbf{v}), \tag{2–1}$$

其中，$\mathbf{v} \in \mathbb{C}^{mn}$ 表示待重建的磁共振图像，$\mathbf{F} \in \mathbb{C}^{mn\times mn}$ 表示离散傅里叶变换，而 R(v) 是对图像 v 的正则化项。$\|\mathbf{x}\|^2 := \|\mathbf{x}\|_2^2 = \sum_{j=1}^{n} |\mathbf{x}_j|^2$ 表示任意复向量 $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top \in \mathbb{C}^n$。使用（2–1）进行并行磁共振成像图像重建时，有两个关键问题：能否获得准确的线圈灵敏度 $\{\mathbf{s}_i\}$，以及能否采用恰当的图像正则化。大多数现有的基于 SENSE 的重建方法都假设已知线圈灵敏度图，但在实际应用中很难准确估计这些图。另一方面，对于逆问题（2–1）而言，正则化项 R 至关重要，它能从大幅欠采样的数据中生成所需图像；然而，许多现有方法采用手工设计的正则化，无法有效提取图像中的复杂特征。

在本文中，我们在一个称为 pMRI-Net 的统一深度学习框架中解决上述两个问题。具体而言，我们考虑重建所有接收线圈的多线圈图像 $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_{N_c}) \in \mathbb{C}^{mnN_c}$，从而避免使用线圈灵敏度图（但可以将其作为副产品恢复），并设计了一种深度残差网络，该网络能够

联合学习多线圈图像的自适应组合方式以及从训练数据中获得的有效正则化。

本文的贡献可概括如下：我们的方法是首个用于基于深度学习的并行磁共振成像重建的“先组合、后正则化”方法。组合算子将多通道图像整合为单通道图像，其效果优于线性组合和平方和开根号（SOS）方法 [142]。该方法有三大优势：(i) 组合图像在整个视野内具有均匀对比度，适用于基于特征的图像正则化，且较少受到线圈图像中强度偏差的影响；(ii) 每次迭代都将正则化算子应用于该单体图像，所需的网络参数大幅减少，从而降低过拟合并提高鲁棒性；(iii) 我们的方法无需使用灵敏度图，自然规避了基于图像的并行磁共振成像重建中一直棘手的问题。

2.2 相关工作

大多数现有的基于深度学习（DL）的方法采用端到端神经网络，将部分 k 空间数据映射为重建图像 [191, 189, 96,145, 123]。这类方法的普遍问题是，深度神经网络需要大量数据进行训练，而得到的网络表现得如同难以解释和修改的“黑箱”。

近年来，一类基于深度学习的方法通过选用迭代优化算法方案，并设定阶段数T，将该方案的每次迭代映射为网络的一个阶段，从而改进了端到端训练。这些方法通常称为学习型优化算法（LOAs），[5, 35, 72, 202, 205, 218, 159]。例如，ADMM-Net [205], 、ISTA-Net$^+$ [218]和级联网络 [159] 用于常规磁共振成像重建。对于并行磁共振成像：变分网络（VN）[72] 通过应用给定的灵敏度 $\{\mathbf{s}_i\}$引入了梯度下降法。MoDL [5] 提出了一种递归网络，通过采用权重共享策略展开共轭梯度算法。

Blind-PMRI-Net [127] 设计了三个网络模块，基于半二次分裂的迭代算法，交替更新多通道图像、灵敏度图和重建的磁共振图像。该网络在 [171] 中提出了用于联合 MRI-PET 重建的贝叶斯框架。VS-Net [48]推导出了一种变量分裂优化方法。然而，现有方法在并行磁共振成像问题中仍面临缺乏准确线圈灵敏度图和适当正则化的问题。

近期，一种名为 DeepcomplexMRI [191] 的方法采用端到端学习，在不显式使用线圈灵敏度图的情况下恢复逐通道图像，并在测试时将其组合成单通道图像。

本文提出了一种新颖的深度神经网络架构，将鲁棒近端梯度方案融入其中，用于在未知线圈灵敏度图的情况下进行并行磁共振成像重建。我们的网络学习根据不完整数据自适应地组合逐通道图像，以辅助重建；同时利用欠采样逐通道 $\mathbf{k}$ 空间数据及其对应图像构成的训练数据对，学习一种非线性映射，以高效提取图像的稀疏特征。每次迭代中，多线圈图像组合算子和稀疏特征编码器的作用均有明确定义，并通过联合学习获得。因此，我们的网络在训练时具有更高的数据效率，重建结果也更加准确。

2.3 所提方法

2.3.1 无 coil 灵敏度的联合通道图像重建并行磁共振成像

我们提出一种针对（2–1）的替代性并行磁共振成像重建方法，通过联合恢复各个接收线圈的图像来实现。记 $\mathbf{u}_i$ 为第 $\mathbf{i}$ 个接收线圈的磁共振图像，即 $\mathbf{u}_i = \mathbf{s}_i \cdot \mathbf{v}$，其中灵敏度 $\mathbf{s}_i$ 和全视野图像 $\mathbf{v}$ 在实际应用中均为未知。因此，图像 $\mathbf{u}_i$ 与部分 $\mathbf{k}$ 空间数据 $\mathbf{f}_i$ 之间的关系为 $\mathbf{f}_i = \mathbf{P}\mathbf{F}\mathbf{u}_i + \mathbf{n}_i$，由此数据保真项可表示为最小二乘形式 $(1/2) \cdot \|\mathbf{P}\mathbf{F}\mathbf{u}_i - \mathbf{f}_i\|^2$。我们还需要对图像 $\{\mathbf{u}_i\}$进行适当的正则化 $\mathbf{R}$。然而，这些图像的

对比度差异很大，这是由于不同接收线圈的灵敏度图存在显著变化。因此，对（未知）图像$\mathbf{v}$进行正则化更为合适。

为了解决正则化问题，我们提出首先学习一个非线性算子$\mathcal{J}$，将 $\{\mathbf{u}_i\}$组合成具有均匀对比度的图像 $\mathbf{v} = \mathcal{J}(\mathbf{u}_1, \dots, \mathbf{u}_{N_c}) \in \mathbb{C}^{m\times n}$，并利用 $\ell_{2,1}$-范数和收缩阈值算子的稳健稀疏选择特性，对 $\mathbf{v}$ 施加参数化形式 $\|\mathcal{G}(\mathbf{v})\|_{2,1}$ 的正则化。这里，$\mathcal{G}(\mathbf{v})$表示一个通过数据训练得到的非线性稀疏编码器，可有效提取图像$\mathbf{v}$中的复杂特征。结合上述数据保真项，我们提出以下并行磁共振成像重建模型：

$$\mathbf{u}(\mathbf{f};\Theta) = \arg\min_{\mathbf{u}} \frac{1}{2}\sum_{i=1}^{N_c} \|\mathbf{P}\mathbf{F}\mathbf{u}_i - \mathbf{f}_i\|_2^2 + \|\mathcal{G}\circ\mathcal{J}(\mathbf{u})\|_{2,1}, \tag{2–2}$$

其中，$\mathrm{u} = (\mathrm{u}_1, \dots, \mathrm{u}_{N_c})$是待从并行磁共振成像数据 $\mathrm{f} = (\mathrm{f}_1, \dots, \mathrm{f}_{N_c})$中重建的多通道图像，$\Theta = (\mathcal{G}, \mathcal{J})$表示深度网络$\mathcal{G}$ 和 $\mathcal{J}$的参数。(2–2)的关键组成部分是非线性组合算子 $\mathcal{J}$ 和稀疏特征编码器 $\mathcal{G}$，我们将在第2.3.3节详细介绍。给定一个由 J 对样本组成的训练数据集 $\{(\mathrm{f}^{[j]}, \hat{\mathrm{u}}^{[j]}) \mid 1 \le \mathrm{j} \le \mathrm{J}\}$，其中 $\mathrm{f}^{[j]} = (\mathrm{f}_1^{[j]}, \dots, \mathrm{f}^{[j]}\,\mathrm{N}_c)$ 和 $\hat{\mathrm{u}}^{[j]} = (\hat{\mathrm{u}}_1^{[j]}, \dots, \hat{\mathrm{u}}^{[j]}\,\mathrm{N}_c)$分别是第 j 个图像数据的部分 k 空间数据，以及由完整 k 空间数据重建得到的真实图像；我们的目标是通过以下双层优化问题学习 Θ （即 $\mathcal{G}$ 和$\mathcal{J}$）：

$$\min_{\Theta} \frac{1}{J}\sum_{j=1}^{J} \ell(\mathbf{u}(\mathbf{f}^{[j]};\Theta), \hat{\mathbf{u}}^{[j]}), \ \text{s.t. } \mathbf{u}(\mathbf{f}^{[j]};\Theta) \text{ solves } (2\text{–}2) \text{ with data } \mathbf{f}^{[j]}, \tag{2–3}$$

其中 $\ell(\mathbf{u}, \hat{\mathbf{u}})$ 用于衡量重建结果$\mathbf{u}$ 与真实值$\hat{\mathbf{u}}$之间的差异。为解决式（2–3）中的低层最小化问题，我们构建了一个带残差学习的近端梯度网络，作为式（2–2）的（近似）求解器。下一小节将介绍该网络的推导细节。

2.3.2 带残差学习的近端梯度网络

如果算子 $\mathcal{J}$ 和 $\mathcal{G}$ 已知，我们可以应用近端梯度下降算法，通过迭代来求得式（2–2）的一个（局部）极小化点的近似值

$$\mathbf{b}_i^{(t)} = \mathbf{u}_i^{(t)} - \rho_t \mathbf{F}^\top \mathbf{P}^\top (\mathbf{P}\mathbf{F}\mathbf{u}_i^{(t)} - \mathbf{f}_i), \tag{2–4a}$$

$$\mathbf{u}_i^{(t+1)} = [\mathrm{prox}_{\rho_t \|\mathcal{G}\circ\mathcal{J}(\cdot)\|_{2,1}}(\mathbf{b}^{(t)})]_i, \quad 1 \le i \le N_c \tag{2–4b}$$

其中 $\mathbf{b}^{(t)} = (\mathbf{b}_1^{(t)}, \dots, \mathbf{b}^{(t)}{}_{N_c})$，$[\mathbf{x}]_i = \mathbf{x}_i \in \mathbb{C}^{mn}$ 表示任意向量 $\mathbf{x} \in \mathbb{C}^{mnN_c}$，$\rho_t > 0$ 为步长，prox_g 为 g 定义的近端算子

$$\mathrm{prox}_g(\mathbf{b}) = \arg\min_{\mathbf{x}} g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{b}\|^2. \tag{2–5}$$

梯度更新步（2–4a）易于计算，并充分利用了根据 MRI 物理推导出的部分 k 空间数据 $\mathbf{f}_i$ 与待重建图像 $\mathbf{u}_i$ 之间的关系。然而，近端更新步（2–4b）存在一些困难：算子 $\mathcal{J}$ 和 $\mathcal{G}$ 均未知，需要从数据中学习；而近端算子 $\mathrm{prox}_{\rho_t\|\mathcal{G}\circ\mathcal{J}(\cdot)\|_{2,1}}$ 很可能不存在闭式解，因而难以计算。假设 $\mathcal{J}$ 和 $\mathcal{G}$ 均由卷积网络参数化，我们采用残差学习技术，利用收缩算子（作为，1-范数 $\|\cdot\|_{2,1}$ 的近端算子），并将（2–4b）转换为显式更新公式。为此，我们将近端步（2–4b）参数化为隐式残差更新：

$$\mathbf{u}_i^{(t+1)} = \mathbf{b}_i^{(t)} + [r(\mathbf{u}_1^{(t+1)}, \cdots, \mathbf{u}_{N_c}^{(t+1)})]_i, \tag{2–6}$$

其中 $r = \tilde{\mathcal{J}} \circ \tilde{\mathcal{G}} \circ \tilde{\mathcal{G}} \circ \tilde{\mathcal{J}}$ 是由以下各项复合而成的残差网络：$\tilde{\mathcal{J}}$、$\tilde{\mathcal{G}}$ 及其伴随算子 $\tilde{\mathcal{J}}$ 和 $\tilde{\mathcal{G}}$。分别学习这四个算子，可以提高网络的容量。为揭示非线性收缩选择在（2–6）中的作用，考虑原始近端更新（2–4b），其中

$$\mathbf{u}^{(t+1)} = \arg\min_{\mathbf{u}} \|\mathcal{G}\circ\mathcal{J}(\mathbf{u})\|_{2,1} + \frac{1}{2\rho_t}\|\mathbf{u} - \mathbf{b}^{(t)}\|^2. \tag{2–7}$$

对于某些采用修正线性单元（ReLU）激活函数的卷积网络 $\mathcal{J}$ 和 $\mathcal{G}$，$\|\mathbf{u}-\mathbf{b}^{(t)}\|^2$ 可近似为 $\alpha\|\mathcal{G}\circ\mathcal{J}(\mathbf{u})-\mathcal{G}\circ\mathcal{J}(\mathbf{b}^{(t)})\|^2$，其中某个 $\alpha>0$ 取决于 $\mathcal{J}$ 和 $\mathcal{G}$ [218]。将此近似代入（2–7），可得

$$\mathcal{G}\circ\mathcal{J}(\mathbf{u}^{(t+1)})=\mathcal{S}_{\alpha_t}(\mathcal{G}\circ\mathcal{J}(\mathbf{b}^{(t)})), \tag{2–8}$$

其中，对于任意向量 $\mathbf{x}=(x_1,\dots,x_n)\in\mathbb{R}^n$，$\mathcal{S}_{\alpha_k}(\mathbf{x})=\mathrm{prox}_{\alpha_k\|\cdot\|_2,1}(\mathbf{x})=[\mathrm{sign}(x_i)\max(|x_i|-\alpha_k,0)]\in\mathbb{R}^n$是软收缩算子。将 (2–8) 代入 (2–6)，可得到 (2–4b) 的显式形式。我们将其与 (2–4a) 一并总结如下：

$$\mathbf{b}_i^{(t)}=\mathbf{u}_i^{(t)}-\rho_t\mathbf{F}^\top\mathbf{P}^\top(\mathbf{P}\mathbf{F}\mathbf{u}_i^{(t)}-\mathbf{f}_i), \tag{2–9a}$$

$$\mathbf{u}_i^{(t)}=\mathbf{b}_i^{(t)}+[\tilde{\mathcal{J}}\circ\tilde{\mathcal{G}}\circ\mathcal{S}_{\alpha_t}(\mathcal{G}\circ\mathcal{J}(\mathbf{b}^{(t)}))]_i,\quad 1\le i\le N_c \tag{2–9b}$$

因此，我们所提出的重建网络由预设的 T 个阶段组成，其中第 t 个阶段执行 (2–9) 中的更新。以零初始值 $\{u^{(0)}\ i\}$ 和部分 k 空间数据 $\{f_i\}$ 作为输入，网络对 $1\le t\le T$ 执行更新 (2–9)，并最终输出 $u^{(T)}$。该网络用作 (2–2) 的求解器，并使用 $u^{(T)}$ 作为真实解 $u(f;\Theta)$ 的近似值。因此，对于每个输入数据 $f^{[j]}$，均以该网络替代 (2–3) 中的约束。

2.3.3 网络架构与训练

我们将 $\mathcal{J}$ 设为一个卷积网络，包含 $N_l=4$ 层，每次线性卷积的卷积核大小为 3×3。前 N_l-1 层各有 $N_f=64$ 个滤波器核，而最后一层有 $N_f=1$ 个。除最后一层外，每层后都接一个 ReLU 激活函数。算子 $\mathcal{G}$ 的设置方式相同，只是 $N_f=32$，卷积核大小为 9×9。算子 $\tilde{\mathcal{J}}$ 和 $\mathcal{G}$ 分别采用对称结构，形式如 $\mathcal{J}$ 和 $\mathcal{G}$。我们将复张量视为大小加倍的实张量，并分别进行卷积。关于网络结构的更多细节见图 2-1和 2-2。实部和虚部卷积算子的输出分别以不同颜色表示。

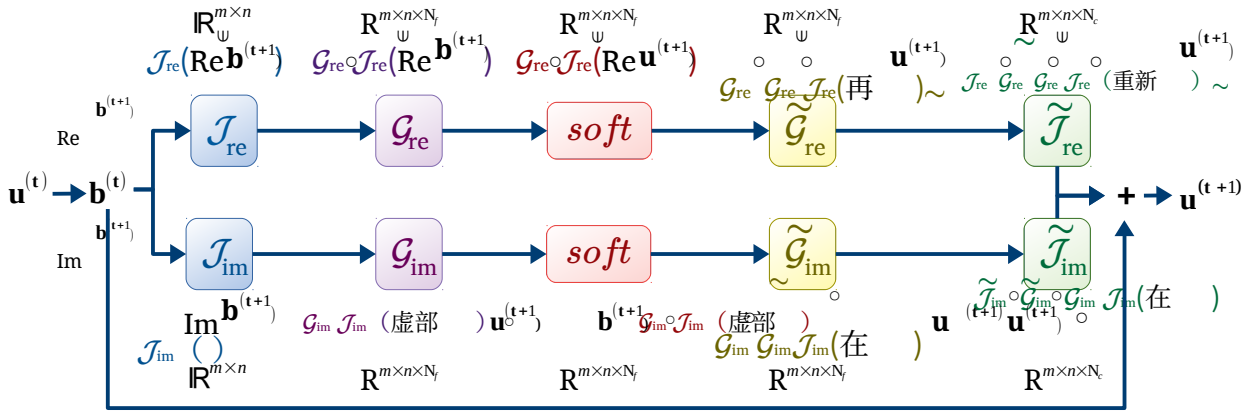

图 2-1。第 $+1$阶段的网络结构范式。$\mathrm{Re}\mathbf{b}^{(t+1)}$、$\mathrm{Re}\mathbf{u}^{(t+1)} \in \mathbb{R}^{m\times n\times N_c}$和 $\mathrm{Im}\mathbf{b}^{(t+1)}$、$\mathrm{Im}\mathbf{u}^{(t+1)} \in \mathbb{R}^{m\times n\times N_c}$ 分别表示$\mathbf{b}^{(t)}$和$\mathbf{u}^{(t)}$ 的实部和虚部。

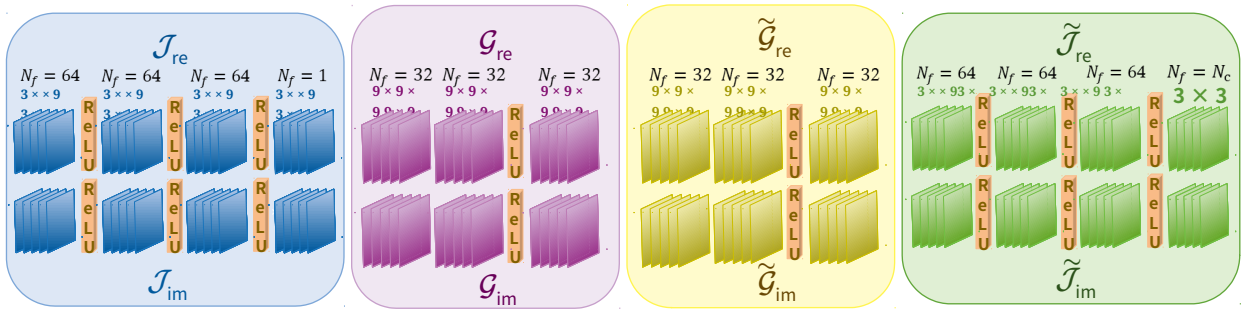

图 2-2。每个卷积算子的结构。网络中实部和虚部的权重不共享，而实部和虚部卷积算子具有相同的结构。

训练数据 $(\mathrm{f}, \hat{\mathrm{u}})$ 由 J 对 $\{(\mathrm{f}^{[j]}\mathrm{i}, \hat{\mathrm{u}}^{[j]}\mathrm{i}) \mid 1 \leq \mathrm{i} \leq \mathrm{N_c},\ 1 \leq \mathrm{j} \leq \mathrm{J}\}$组成。为了提高网络容量，我们允许在不同阶段使用 (2–9) 的不同算子。因此，$\Theta = \{\rho_t, \alpha_t, \mathcal{J}(\mathrm{t})、\mathcal{G}^{(t)}、\tilde{\mathcal{G}}^{(t)}、\tilde{\mathcal{J}}(\mathrm{t}) \mid 1 \leq \mathrm{t} \leq \mathrm{T}\}$ 是需要训练的参数。基于对损失函数 [222, 70, 226], 的分析，可以通过最小化损失函数求解最优参数 Θ：我们将重建结果 u 与 (2–3) 中对应的真实值 $\hat{\mathrm{u}}$ 之间的差异度量 ℓ设定如下：

$$\ell(\mathbf{u}, \hat{\mathbf{u}}) = \|\mathbf{s}(\mathbf{u}) - \mathbf{s}(\hat{\mathbf{u}})\|_2 + \gamma \| |\mathcal{J}(\mathbf{u})| - \mathbf{s}(\hat{\mathbf{u}})\|_2 \tag{2–10}$$

其中，$\mathbf{s}(\mathbf{u}) = (\sum^{N_c}{}_{i=1} |\mathbf{u}_i|^2)^{1/2} \in \mathbb{R}^{mn}$ 是对$\mathbf{u}$的N_c个通道逐点计算平方和平方根，$|\cdot|$ 是逐点模，$\gamma > 0$ 是权重函数。我们还尝试了

表 2-　1. 冠状位 FSPD 数据重建的定量测量。

方法	PSNR	SSIM	RMSE
GRAPPA[65]	24.9251±0.9341	0.4827±0.0344	0.2384±0.0175
SPIRiT[119]	28.3525±1.3314	0.6509±0.0300	0.1614±0.0203
VN[72]	30.2588±1.1790	0.7141±0.0483	0.1358±0.0152
DeepcomplexMRI[191]	36.6268±1.9662	0.9094±0.0331	0.0653±0.0085
pMRI-Net	**37.8475±1.2086**	**0.9212±0.0236**	**0.0568±0.0069**

Table 2-2. 冠状位PD数据重建的定量测量值　　ta.

方法	PSNR	SSIM	RMSE
GRAPPA[65]	30.4154±0.5924	0.7489±0.0207	0.0984±0.0030
SPIRiT[119]	32.0011±0.7920	0.7979±0.0306	0.0824±0.0082
VN [72]	37.8265±0.4000	0.9281±0.0114	0.0422±0.0036
DeepcomplexMRI [191]	41.5756±0.6271	0.9679±0.0031	0.0274±0.0018
pMRI-Net	**42.4333±0.8785**	**0.9793±0.0023**	**0.0249±0.0024**

将第一项替换为 $(1/2)\cdot\|u-\hat{u}\|_2^2$，但在我们的实验中，(2–10) 中给出的形式似乎效果更好。(2–10) 的第二项可以进一步提高重建幅度的准确率。对于任意给定的并行磁共振成像 pMRI $f^{[i]}$，其初始猜测（也是重建网络的输入）设为零填充重建 $F^{-1}f^{[i]}$；多通道图像 $u^{(T)}(f^{[i]};\Theta)$ 是网络 (2–9) 经过 T 个阶段后的输出。此外，$\mathcal{J}(u^{(T)}(f^{[i]};\Theta))$ 是作为副产品重建出的最终单体图像（复值）。

2.4 实验结果

数据。在我们的实验中，从 https://github.com/VLOGroup/mri-variationalnetwork 获取了两组数据序列，分别命名为冠状面质子密度（PD）和冠状面脂肪抑制质子密度（FSPD），以及采样率为 31.56% 的规则笛卡尔采样掩模。两组序列数据均采集自 20 名患者。训练数据由 526 个中心图像切片组成，矩阵大小为320×320，这些数据来自 19 名患者；我们从未纳入训练数据的另一名患者中随机选取 15 个中心图像切片作为测试数据。我们使用零填充重建的绝对值最大值对训练数据进行归一化。

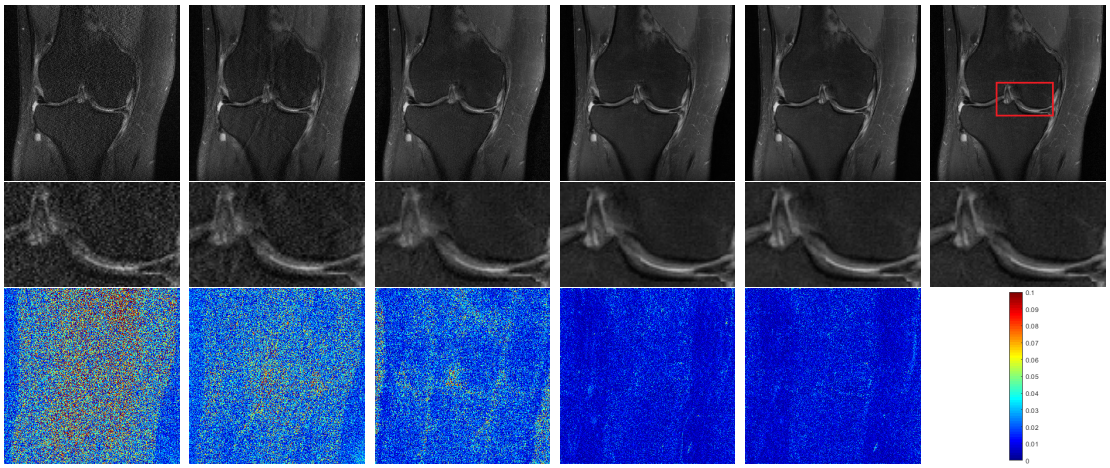

图 2-3。采用规则笛卡尔采样（31.56% 采样率）对冠状位 FSPD 膝关节图像进行重建的结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、pMRI-Net 和真实值。

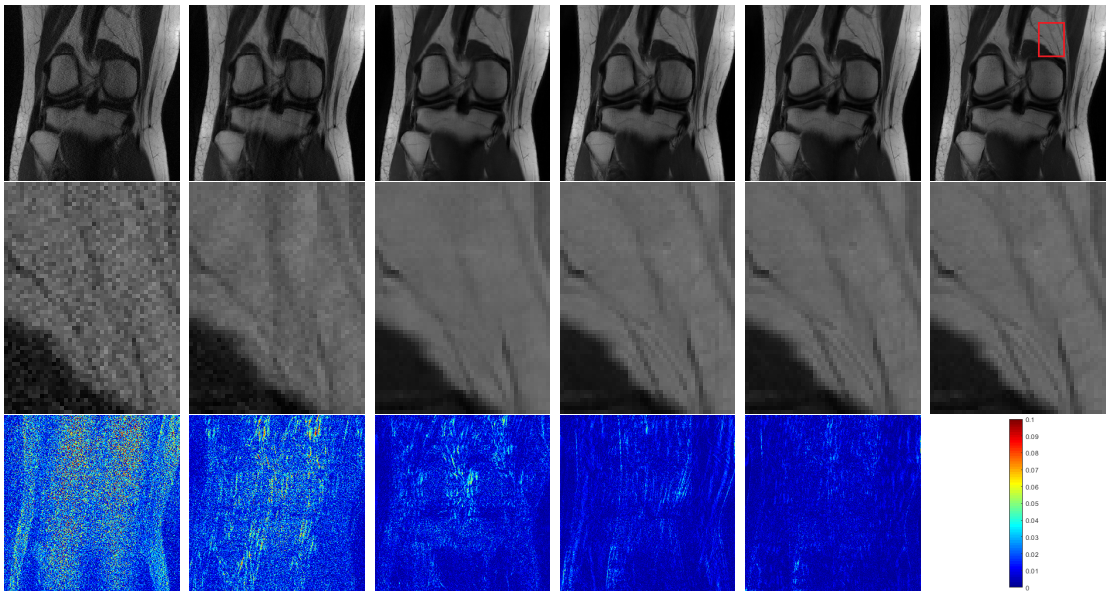

图 2-4。采用规则笛卡尔采样（31.56% 采样率）对冠状位 PD 膝关节图像进行重建的结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、pMRI-Net 和真实值。

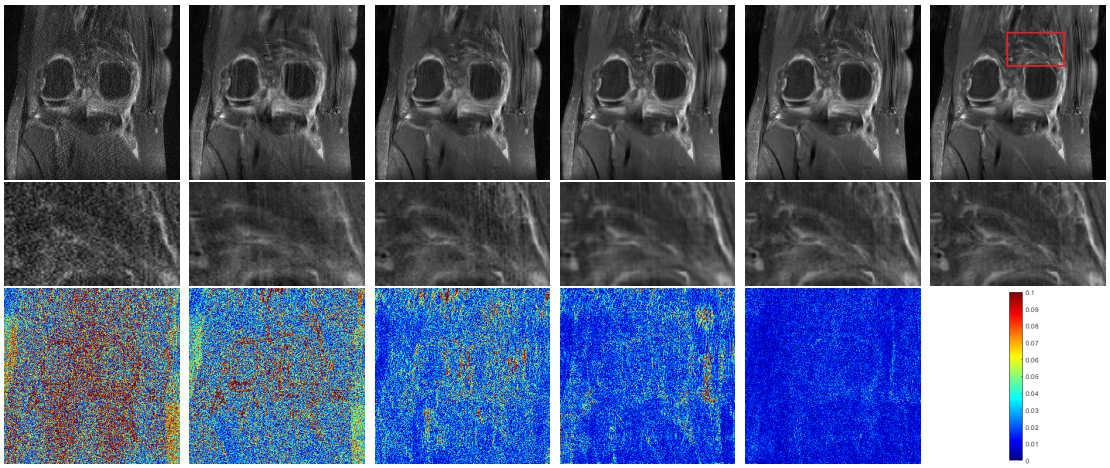

图 2-5. 对采用规则笛卡尔采样（采样率为 31.56%）的冠状位 FSPD 膝关节图像进行的补充实验结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、所提出的方法和真实值。

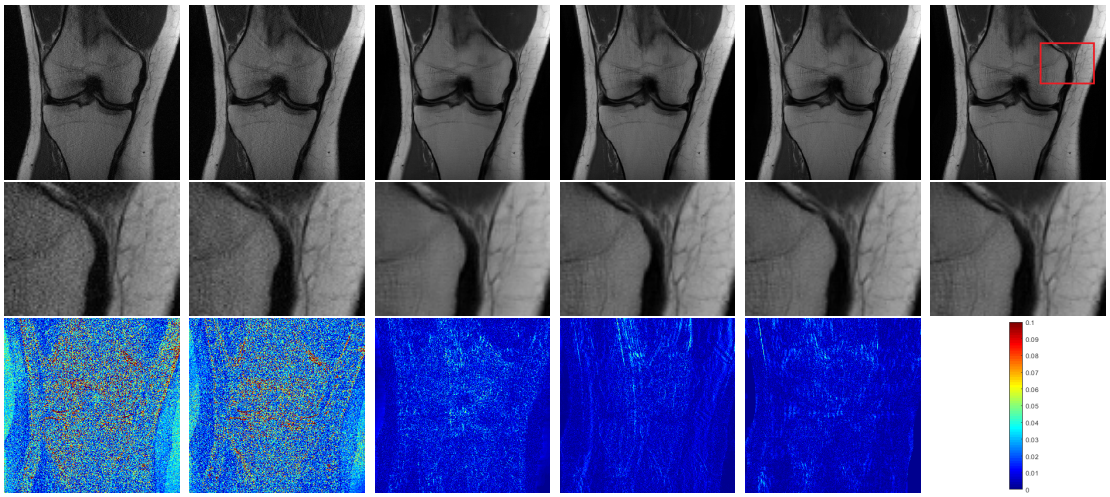

图 2-6. 对采用规则笛卡尔采样（采样率为 31.56%）的冠状位 PD 膝关节图像进行的补充实验结果。从左至右各列依次为：GRAPPA、SPIRiT、VN、DeepcomplexMRI、所提出的方法和真实值。

实现。 所提出的网络使用 $T = 5$ 个阶段。我们采用 Xavier 初始化 [59] 来初始化网络参数，并使用 Adam优化器进行训练。实验中采用大小为 2 的小批次，训练 3000 轮次；训练冠状位 FSPD 数据和 PD 数据的学习率分别为 0.0001 和 0.0005。初始步长 $\rho_0 = 0.1$、阈值参数 $\alpha_0 = 0$ 以及损失函数中的 $\gamma = 10^5$ 。所有实验均

在配备英特尔酷睿 i9-7900 处理器和 Nvidia GTX-1080Ti GPU 的工作站上使用 TensorFlow 实现。

评估。 我们从 PSNR、SSIM [65], 和 [119], RMSE 方面，对传统方法 GRAPPA [72] , SPIRiT [191] 以及深度学习方法 VN [193] DeepcomplexMRI [191] 在 15 张测试冠状位 FSPD 和 PD 膝关节图像上的表现进行评估（$\hat{\mathbf{x}}$ **与真实** $\mathbf{x}^*$ 之间的 RMSE 定义为 $\|\hat{\mathbf{x}} - \mathbf{x}^*\| / \|\mathbf{x}^*\|$）。

实验结果。 平均数值性能及标准差汇总于表 2-1和2-2中。重建图像的比较结果如图 2-3、2-5和图 2-4、2-6 所示，分别对应冠状位 FSPD 和 PD 测试数据。从上到下各行依次为：图像、放大视图以及相对于真实值的逐点绝对误差。在图 2-3中，各方法从左到右各列对应的 PSNR/SSIM/RMSE 如下：GRAPPA：25.6656/0.4671/0.2494，SPIRiT：29.5550/0.6574/0.1594，VN：31.5546/0.7387/0.1333，deepcomplexMRI：38.6842/0.9360/0.0587，pMRI-Net：38.8749/0.9375/0.0574，以及真实值。在图 2-5中，从左到右各列为：GRAPPA：22.2203/0.3596/0.2666，SPIRiT：25.3434/0.5269/0.1861，VN：29.2276/0.7591/0.1190，DeepcomplexMRI：31.3597/0.8231/0.0931，所提出的方法：36.8831/0.9372/0.0493，以及真实值。在图 2-4中，从左到右各列为：GRAPPA：29.9155/0.7360/0.1032，SPIRiT：33.2350/0.8461/0.0704，VN：38.3192/0.9464/0.0393，DeepcomplexMRI：41.2098/0.9713/0.0281，pMRI-Net：42.9330/0.9798/0.0231

，

和真实值。在图2-6中，从左至右各列依次为：GRAPPA：27.3571/0.5338/0.1357，SPIRiT：28.3076/0.5592/0.1216，VN：38.1679/0.9259/0.0391，DeepcomplexMRI：40.9388/0.9678/0.0284，所提出的方法：41.7025/0.9697/0.0260，以及真实值。

尽管训练和测试过程中均未使用线圈灵敏度，所提方法的重建精度仍显著优于 VN；而 VN 使用的是 ESPIRiT 预先计算的线圈灵敏度图 [180], 。这进一步表明，所提方法无需线圈灵敏度信息即可提高准确率。比较 DeepcomplexMRI 中的 10 个

个复数 CNN模块与 pMRI-Net 中的 5 个阶段可见，后者所需的网络参数更少、训练时间更短，同时还能提高重建质量。

在 GRAPPA 和 SPIRiT 的实验中，我们使用的校准卷积核大小为 5×5，校准集中的蒂霍诺夫正则化设为 0.01。SPIRiT 的迭代次数设为 30 次，重建中的蒂霍诺夫正则化设为 10^{-3}。VN 和 DeepcomplexMRI 实验采用默认参数设置。VN 最终恢复的图像是全视野单通道图像，而 DeepcomplexMRI 生成多线圈图像，并使用自适应多线圈组合方法将其合成为单通道图像 [188]。pMRI-Net 同时重建单通道图像$\mathcal{J}(\mathbf{u}^{(T)}(\mathbf{f}; \Theta))$和多通道图像 $\{\mathbf{u}^{(T)}{}_i(\mathbf{f}_i; \Theta)\}$。

2.5 结论

我们采用一种基于学习的多线圈MRI重建方法，无需显式获取线圈灵敏度图，并在压缩感知框架中通过近端梯度方案对网络进行建模。所提出的网络旨在组合逐通道图像的特征，然后从线圈组合图像中提取稀疏特征。实验表明，所提出的“先组合、后正则化”方法性能更优。

第3章 无需线圈灵敏度的联合通道并行 MRI 重建最优控制框架

本研究旨在开发一种新颖的无校准快速并行磁共振成像（pMRI）重建方法，并将其与离散时间最优控制框架相结合。该重建模型旨在学习一种正则化方法，通过利用多线圈图像各通道之间的信息共享来组合通道并提取特征。我们利用图像域和傅里叶空间中的结构化卷积网络，恢复幅度和相位信息。我们开发了一种新颖的变分模型，其可学习目标函数融合了自适应多线圈图像组合算子，以及在图像域和傅里叶空间中的有效图像正则化。我们将重建网络构建为结构化离散时间最优控制系统，从而将参数训练表述为最优控制形式，其中目标函数的参数充当控制变量。我们证明，求解控制问题的拉格朗日方法等价于反向传播，从而确保训练算法的局部收敛。我们针对所提出的方法开展了大量数值实验，并在真实并行磁共振成像数据集上与若干先进的并行磁共振成像重建网络进行比较。数值结果清楚地表明，所提出的方法表现出良好的性能。所提出的方法为高效的联合通道并行磁共振成像重建提供了一种通用的深度网络设计与训练框架。通过学习多线圈图像组合算子，并在图像域和k空间域中进行正则化，所提出的方法实现了高效的并行磁共振成像图像重建网络。

3.1 引言

磁共振成像（MRI）是应用广泛的医学成像技术之一，在临床应用中占据重要地位。在临床应用中，一种称为并行磁共振成像（pMRI）的先进医学磁共振成像技术得到了广泛应用。并行磁共振成像（pMRI）将

多个接收线圈布置在扫描对象周围，并行采集k空间（傅里叶）数据。并行磁共振成像（pMRI）能够缩短数据采集时间，已成为现代磁共振成像应用中的先进技术。为加快扫描过程，并行磁共振成像（pMRI）采用部分数据采集，增大k空间中读出线之间的间距。然而，这会产生混叠伪影，因此需要通过适当的图像重建过程，从部分数据中恢复高质量、无伪影的图像。

并行磁共振成像的图像重建通常采用两种主要方法：第一种是k空间方法，它利用多个接收线圈采集到的数据对缺失的k空间数据进行插值 [45], ，例如广义自校准部分并行采集（GRAPPA） [65] 和空间谐波同时采集（SMASH） [164]。另一种是图像空间方法，通过求解一个将待重建图像与部分k空间数据通过线圈灵敏度联系起来的方程组，消除图像域中的混叠伪影，例如灵敏度编码（SENSE） [142]。传统的基于SENSE的方法离不开线圈灵敏度图，且必须对其进行准确的预估。传统的图像空间并行磁共振成像重建方法遵循基于SENSE的框架，其被表述为一个优化问题，旨在最小化数据保真项与加权正则化项之和。有关该公式的详细解释，请参阅 [94]。

近年来，我们见证了基于深度学习的方法与基于SENSE的并行磁共振成像重建相结合后迅速发展 [94, 115, 173, 71, 120]。其中有两个关键问题需要认真处理。第一个问题是如何选择正则化，包括正则化权重。由于k空间数据严重欠采样，并行磁共振成像重建是一个病态程度极高的逆问题，因此正则化项至关重要。过去几十年来，大多数传统图像重建方法都采用手工设计的正则化项，例如全变分（TV）。近年来，一类模仿经典

优化方案的展开方法应运而生，其中正则化通过深度网络实现，其参数由数据学习得到。然而，学习得到的正则化通常被视为难以解释的黑箱；此外，训练往往需要大量数据，且容易过拟合，尤其是在网络参数过多时 [98, 12, 76]。

第二个问题源于实际中无法获得准确的线圈灵敏度 $\{\mathbf{s}_i\}$ 。不准确的线圈灵敏度图会导致严重偏差，降低重建$\mathbf{v}$的质量。解决这一问题的一种方法是对多线圈图像进行（正则化并）重建，最后一步再采用平方和平方根（RSS）等人工设计的方法，将各通道组合成全身图像。与 RSS 不同，我们提出了一种可学习的多线圈组合算子，用于组合逐通道多线圈图像。然而，如果所有线圈的信噪比都较低，或数据中存在伪影，RSS 会同等加权伪影区域与暗区，导致背景噪声水平升高[187, 38]。

本文在离散时间最优控制框架下解决上述问题，以优化变分并行磁共振成像重建模型。我们概述了该框架的若干主要特点如下。

1. 与大多数对多线圈图像进行正则化和重建的现有方法不同，我们在图像空间和傅里叶空间中均采用正则化，以提高重建质量。

2. 我们的方法提倡采用学习式自适应组合算子，先将多线圈图像合并为一幅视野（FOV）完整的全身图像，然后对该图像进行有效的正则化。这与现有方法形成鲜明对比：现有方法仅在最后一步合并重建后的多线圈图像，而我们的正则化项在每次迭代中都利用组合算子，从而提高了参数效率。

3. 我们采用复值神经网络作为线圈组合算子，在无法获得线圈灵敏度时恢复并行磁共振成像图像的幅度和相位信息

。这种组合方法利用了多个通道共享的线圈信息，这一点不同于大多数手工设计的线圈组合方法。4. 我们提出了一种新颖的深度重建网络，其结构由用于最小化目标函数的离散时间最优控制系统确定。由此得到一种最优控制形式，其中组合算子和正则化算子的参数充当离散动力系统的控制变量。这些参数的最优值通过拉格朗日方法求得，并可使用反向传播实现。

我们考虑了两种临床膝部图像磁共振序列，并在消融研究中验证了所提出的组合算子、不同初始重建结果、复数卷积和域混合网络的有效性能。所提出的网络能够恢复并行磁共振成像图像的幅度和相位信息。上述技术的作用表明，使用我们的方法可以显著提升重建质量和参数效率。为了复现实验，我们的代码可在https://github.com/1lol/pMRI optimal control获取。

本文其余部分的组织结构如下：在第 3.2节中，我们概述了与本文工作相关的并行磁共振成像、跨域重建、复值卷积神经网络以及最优控制启发的深度训练模型的最新进展。我们在第3.3节中详细介绍所提出的问题设置和重建网络架构。在第3.4节中，我们展示了针对各种临床并行磁共振成像数据开展的大量数值实验和分析。第 3.5节对本文进行总结。

3.2 相关工作

近年来，我们见证了医学成像与基于深度学习的方法相结合的快速发展 [80, 82, 81]。大多数现有的基于深度学习的方法都采用端到端神经网络，将部分 k 空间数据映射为重建图像 [191, 227, 145, 123, 228]。这些方法需要大量训练数据，而且所设计的网络被视为黑箱，其

底层机制可能很难解释。为缓解这一问题，研究人员提出了多种展开方法，将现有的优化算法映射为结构化网络，其中网络的每个阶段对应优化算法的一次迭代 [72, 3, 5, 202, 35, 205, 218, 159]。下文将重点介绍并行磁共振成像（pMRI）中基于深度学习的图像重建方法的最新进展。

一些用于并行磁共振成像重建的网络使用 ESPIRiT：变分网络（VN） [72] 被提出用于将梯度下降算法展开为重建网络，并要求将预先计算的灵敏度 $\{\mathbf{s}_i\}$ 作为输入。E2E-VarNet[169] 在级联细化模块中对一种改进的 VN 进行了建模，并学习灵敏度图；其结果表明，该方法在 fastMRI 公共排行榜中名列前茅。MoDL [5] 提出了递归网络中的权重共享策略，通过展开共轭梯度法来学习正则化参数。PI-CNN[227] 被提出作为一种端到端级联网络，在每个阶段将一个 CNN 模块与一个并行成像数据一致性模块串联起来。上述技术使用 ESPIRiT [180] 来估计线圈灵敏度图。VS-Net [48]采用了由 BART [179]预先计算的线圈灵敏度，这些灵敏度利用全采样k空间的中心块，并将变量分裂算法中的优化步骤展开。APIR-Net [217] 提出了一种分层的自校准k空间补全方法，逐步增大 ACS区域。若干方法探索了不同策略，以避免在并行磁共振成像重建中使用预先计算的线圈灵敏度图。Blind-PMRI-Net [127] 通过对灵敏度图和磁共振图像进行正则化，提出了一种并行磁共振成像模型；其网络通过三个子网络交替估计线圈图像、灵敏度和单体图像。De-Aliasing-Net [31] 提出了一种去混叠重建模型，采用分裂Bregman迭代算法，无需显式计算线圈灵敏度。该去混叠网络利用通道间的互相关和空间冗余，从而取得了理想的性能。LINDBERG[190] 探索了一种无校准并行磁共振成像技术，采用自适应稀疏编码来

通过加入联合稀疏正则化以提取理想的跨通道关系，从而精确获得联合稀疏表示。这项工作提出交替更新稀疏表示、灵敏度编码图像和k空间数据。SCDAE[88] 通过卷积和全连接层来估计逐线圈灵敏度图。重建采用基于TV的最小化算法，并通过布雷格曼迭代技术求解。多通道组合通过平方和开根号获得全视野图像。Adaptive-CS-Net [140]是2019年fastMRI挑战赛中的领先方法 [215] ，该方法展开了改进的 ISTA-Net$^+$ [218]。所提出的无校准并行磁共振成像方法与上述相关工作不同，其采用可学习的多线圈组合算子，在迭代过程中自适应地组合更新后的多线圈图像的各个通道。

近年来，跨域方法在医学成像中展现出重要意义 [51, 192,170, 136, 167, 90, 230, 166]。AUTOMAP [228] 由全连接层和后续的卷积层构成，用于学习从欠采样K空间到图像域的变换。GrappaNet [170] 通过引入GRAPPA算子利用物理信息。CD-SFCRF框架 [103] 提出了一种随机全连接图模型，通过利用K空间和空间域中的约束来实现磁共振成像重建。KIKI-net [51] 迭代应用K空间卷积神经网络、图像域卷积神经网络以及交错的数据一致性操作，以实现单线圈图像重建。IKWI-net [192]依次在图像域、K空间、小波域和图像域中采用CNN模块。CDF-Net [136]进一步表明，在空间域与频域之间增加信息交互能够提升性能。其结果表明，特定域网络在恢复组织结构方面各有优势和不足。我们的重建模型受到跨域重建的启发，不同之处在于，我们通过可学习优化算法，利用跨域正则化函数求解重建模型，而不是采用端到端网络。然而，这些黑箱子网络以

端到端的方式进行训练，且参数效率低下，从而降低了训练性能。本文将跨域的思想融入受近端梯度下降启发的网络中，具体而言，我们将跨域思想融入残差网络。针对并行磁共振成像的特点进行精心设计的研究仍然有限。

我们的网络采用复值卷积和激活函数。MRI 数据为复数值，而相位信号也携带重要的病理信息，例如定量磁化率成像中的信息。 [37, 157]大多数现有的基于深度学习的磁共振成像重建方法，都会分别对实部和虚部应用标准实值卷积和非线性激活函数。 [15, 72, 5, 205]采用经过适当正则化设计的复值卷积神经网络有利于优化。一方面，与实值卷积神经网络相比，复数卷积可减少过拟合 [147] ，因为训练参数的数量有所减少。另一方面，由此得到的网络结构更具可解释性。Trabelsi 利用了复数卷积和算法、批归一化以及权重初始化策略等。 [176]Scardapane等。 [158] 将核激活函数这一一般思路拓展到复值神经网络，设计了非参数激活函数。Cole等。 [37] 研究发现，在各种网络架构中，与采用实值卷积的模型相比，采用复值卷积和激活函数的模型重建效果更好。Virtue等。 [185] 开发了复值心形激活函数，并将其应用于复值神经网络，以识别用于磁共振指纹成像的组织参数。 ℂDFNet [44] 提出了一种基于 U-Net 的端到端学习方法，采用复数卷积，随后进行复值批归一化和修正线性单元（ReLU）。Co-VeGAN [183] 开发了复值生成对抗网络，并采用复值激活函数来提升重建性能，同时保留相位信息。DeepcomplexMRI [191] 通过展开一个端到端复值网络来恢复多线圈图像。该网络由若干模块组成

其中，每个模块都会对多线圈图像进行复值卷积以实现去噪，并执行数据一致性操作，将原始采样的 k 空间数据合并到去噪图像的伪全 k 空间中。

在监督学习中，深度残差神经网络可以近似为动力系统经典最优控制问题的离散化，其中训练参数可以视为控制变量 [194, 105, 108]。控制启发式学习算法引入了一类新的网络训练模型，将其与动力系统联系起来。庞特里亚金最大值原理（PMP） [68]被用于研究最优控制的必要最优性条件， [106], [107], 这些工作设计了逐次逼近法（MSA）的离散方法 [36] 及其变体，以求解PMP。神经常微分方程 [28] 通过某些类型的神经网络（如残差网络）对隐藏状态的连续动力学进行建模，其前向传播等价于对常微分方程（ODE）进行离散化后的一步计算。受神经常微分方程 [28],Chen 等人的工作启发， [27] 对基于ODE的深度网络进行了建模，用于磁共振成像重建。本文从拉格朗日乘子法的角度，将优化轨迹建模为离散动力过程。

本研究在 [15] 中使用具有训练后初始值的域混合网络求解线圈灵敏度不可用时的最优控制并行磁共振成像联合通道重建问题，是对前期工作的重大拓展。本研究还开展了更全面的实证研究。

3.3 所提方法

3.3.1 背景

并行磁共振成像以及一般的磁共振成像重建都可以表述为一个逆问题。考虑一个具有 c 个接收线圈的并行磁共振成像系统，以分辨率 $m \times n$ 采集二维磁共振图像（下文中，我们将二维图像 $\mathbf{v} \in \mathbb{C}^{m\times n}$ 与其列向量形式 $\mathbf{v} \in \mathbb{C}^{mn}$ 视为可互换）。令 $\mathbf{P} \in \mathbb{R}^{p\times mn}(p \leq mn)$ 为表示欠采样掩膜的二值矩阵，其中 p 表示k空间中的采样位置，$\mathbf{s}_i \in \mathbb{C}^{mn}$ 为线圈灵敏度，而 $\mathbf{f}_i \in \mathbb{C}^p$ 为

第i个接收线圈采集的部分 k空间数据，其中$i = 1, \dots, c$。部分数据$\mathbf{f}_i$与图像$\mathbf{v}$ 之间的关系为$\mathbf{f}_i = \mathbf{PF}(\mathbf{s}_i \odot \mathbf{v}) + \mathbf{n}_i$，其中$\mathbf{s}_i$ 是第i个线圈的（未知）灵敏度图，$\odot$ 表示两个矩阵的逐元素乘积，$\mathbf{F} \in \mathbb{C}^{mn \times mn}$ 表示将图像映射为其傅里叶系数的（归一化）离散傅里叶变换，而$\mathbf{n}_i$ 是第i个接收线圈在k空间中的未知采集噪声。于是，图像重建的变分模型可表述为如下优化问题：

$$\min_{\mathbf{v}} \sum_{i=1}^{c} \frac{1}{2} \|\mathbf{PF}(\mathbf{s}_i \odot \mathbf{v}) - \mathbf{f}_i\|^2 + R(\mathbf{v}), \tag{3–1}$$

其中$\mathbf{v} \in \mathbb{C}^{mn}$ 是待重建的单幅全身磁共振图像，$R(\mathbf{v})$ 是对图像 $\mathbf{v}$的正则化，且 $\|\mathbf{w}\|^2 := \|\mathbf{w}\|_2^2 = \sum^{n}_{j=1} |w_j|^2$ 表示任意复向量$\mathbf{w} = (w_1, \dots, w_n)^\top \in \mathbb{C}^n$。我们的方法基于均匀笛卡尔 k 空间采样。表 3-1列出了本文使用的符号及其说明。

3.3.2 问题设置

我们提出了一种统一的深度神经网络，通过联合恢复来自各个接收线圈的图像，实现无校准并行磁共振成像重建，且无需了解逐线圈灵敏度分布。记$\mathbf{u}_i$ 为第i个接收线圈的磁共振图像，因此它与全身图像$\mathbf{v}$ 的关系为$\mathbf{u}_i = \mathbf{s}_i \odot \mathbf{v}$。另一方面，图像$\mathbf{u}_i$对应部分 k 空间数据$\mathbf{f}_i$，其关系为$\mathbf{f}_i = \mathbf{PFu}_i + \mathbf{n}_i$，因此数据保真项可表述为最小二乘 $\frac{1}{2}\sum^{c}_{i=1} \|\mathbf{PFu}_i - \mathbf{f}_i\|^2$。我们还需要对图像 $\{\mathbf{u}_i\}$施加适当的正则化R。然而，由于不同接收线圈的灵敏度在图像域内存在物理变化，这些图像具有严重的几何非均匀对比度。因此，与对各个线圈图像（$\mathbf{u}_1, \dots, \mathbf{u}_c$）进行正则化相比，对单幅全身图像$\mathbf{v}$ 进行正则化更为恰当且有效。

为解决如何进行适当正则化的问题，我们提出学习一个非线性算子$\mathcal{J}$，将 $\{\mathbf{u}_i\}$ 合并为图像 $\mathbf{v} = \mathcal{J}(\mathbf{u}) \in \mathbb{C}^{m \times n}$，其中 $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_c) \in \mathbb{C}^{m \times n \times c}$表示通道级多线圈磁共振成像数据，由 $\mathbf{u}_i$（对应$i = 1, \cdots, c$）组成。然后

表 3-1. 本文中使用的一些符号及其含义

表达式	描述
c	接收线圈总数
$\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_c)$	多线圈磁共振成像数据
$\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_c)$	部分 k 空间测量
$\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_c)$	线圈灵敏度图
$\mathbf{v}$	需要重建的全视野图像
$\mathbf{u}^*$	真实多线圈MRI数据
$\mathbf{v}^*$	真实单体MRI数据
$\mathbf{F}$	离散傅里叶变换
$\mathbf{F}^H$	逆离散傅里叶变换
$\mathbf{P}$	欠采样轨迹
$\mathbf{n}$	测量噪声
$\mathbf{g}$	算法展开网络
$\mathbf{g}_0$	初始网络
$\mathcal{J}, \mathcal{G}, \tilde{\mathcal{G}}, \tilde{\mathcal{J}}$	$\mathbf{g}$中的图像空间卷积算子
$\mathcal{K}$	$\mathbf{g}$中的k空间卷积算子
$\mathcal{K}_0$	$\mathbf{g}$中的k空间卷积算子$_0$
h	数据保真项
R	正则化
RSS	平方和的平方根
$t = 1, \cdots, T$	相位数
$k = 1, \cdots, K$	Alg 的迭代次数（1）
$\mathbf{U} = (\mathbf{u}(0), \cdots, \mathbf{u}(T))^\intercal$	预测的多线圈图像集合在每个阶段
$\Theta = (\theta(0), \cdots, \theta(T))^\intercal$	控制变量（参数）集合在每个阶段
$\Lambda = (\lambda(0), \cdots, \lambda(T))^\intercal$	（3–4）的拉格朗日乘子集合

我们对图像$\mathbf{v}$应用适当的 R 。我们还在 $\mathbf{F}\mathbf{u}_i$ 上引入一个k空间 R_f ，以利用傅里叶信息并提升模型性能。

我们将$\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_c) \in \mathbb{C}^{p \times c}$表示为$c$个传感器线圈处的部分 k 空间测量值。假设给定了用于网络训练的N 个数据对 $\{(\mathbf{f}^{(j)}, \mathbf{u}^{*(j)})\}_{j=1}^{N}$ ，其中$\mathbf{u}^{*(j)}$是真实的多线圈磁共振数据，其索引为$j \in \{1, \cdots, N\}$。令 Θ表示通过最小化损失函数 ℓ从网络中学习的参数。我们将网络训练表述为一个双层优化问题，其中下层是在固定可训练参数 Θ 的情况下更新$\mathbf{u}$ ，而上层则是

更新通过最小化损失函数 ℓ从训练数据中学习的 Θ 。

$$\min_{\Theta} \quad \frac{1}{N}\sum_{j=1}^{N} \ell(\mathbf{u}_{\Theta}^{(j)}), \tag{3–2a}$$

$$\text{s.t.} \quad \mathbf{u}_{\Theta}^{(j)} = \arg\min_{\mathbf{u}^{(j)}} \phi_{\Theta}(\mathbf{u}^{(j)}), \tag{3–2b}$$

其中 ϕ_{Θ} 定义如下:

$$\phi(\mathbf{u}) := \frac{1}{2}\sum_{i=1}^{c} \|\mathbf{P}\mathbf{F}\mathbf{u}_i - \mathbf{f}_i\|^2 + R(\mathcal{J}(\mathbf{u})) + R_f(\mathbf{F}\mathbf{u}_i). \tag{3–3}$$

目标函数 ϕ 是用于并行磁共振成像重建的变分模型，其中 Θ 是一个集合，包含从正则化项中学习到的所有参数R $\circ \mathcal{J}$和$\mathbf{R}_f \circ \mathbf{F}$，因此 ϕ 取决于 Θ。问题3–2 是在重建多线圈磁共振成像数据的场景下构建的。最终重建结果$\mathbf{u}_{\Theta}$ 是依赖于网络参数 Θ的前向网络输出。

下层问题（3–2b）中的并行磁共振成像重建深度学习方法可表示并构建为一个离散时间最优控制系统。为简化符号表示，我们使用一个数据样本($\mathbf{f}$，$\mathbf{u}^*$），并省略平均值和数据索引。深度展开方法的架构遵循优化算法的迭代，并求解以下问题的极小化点:

$$\min_{\Theta} \quad \ell\ (\mathbf{u}_{\Theta})\ , \tag{3–4a}$$

$$\text{s.t.} \quad \mathbf{u}(t) = \mathbf{g}(\mathbf{u}(t-1), \theta(t)),\ t = 1, \cdots, T, \tag{3–4b}$$

$$\mathbf{u}(0) = \mathbf{g}_0(\mathbf{f}, \theta(0)), \tag{3–4c}$$

其中，$\Theta = (\theta(0), \cdots, \theta(\mathrm{T}))$ 分别表示所有时间步（阶段）的控制变量 θ(t) 的集合。在 (3–4a) 中，$\mathrm{u}_{\Theta} = \mathrm{u}(\mathrm{T})$ 是整个网络最后一个（第 T 个）阶段的输出图像，因此我们希望找到最优的 u(T)，使其接近 $\arg\min_{\mathrm{u}} \phi_{\Theta}(\mathrm{u})$。(3–4b)、(3–4c) 受用于求解下层约束 (3–2b) 的深度展开算法启发。给定 θ(t)，u(t) 是更新后的重建状态

每个 $t = 0, \cdots, T$ 对应的第 t 个阶段的多线圈数据。g 是受近端梯度算法启发的多相位展开网络，$g(\cdot) \in \mathbb{C}^{m \times n \times c}$ 的输出是各阶段更新后的多线圈磁共振成像数据。网络 g 是从 u(t) 到 $u(t+1)$ 的中间映射，其中 $t = 0, \cdots, T-1$；其结构将在第3.3.4节中介绍，用于最小化变分模型 (3–3)。带有初始控制参数 $\theta(0)$ 的网络 g_0 将部分 k 空间测量值 f 映射为初始重建 u(0)，作为该最优控制系统的输入。

简而言之，我们在离散时间最优控制框架（3–2b）中求解下层问题（2–3）的极小化点（重建结果）。作为（3–4a）约束条件的动力系统（3–4b）、（3–4c）被建模为变分模型（3–3）的最优控制系统。

损失函数 $\ell(\mathbf{u}(T))$ 用于衡量最终状态 $\mathbf{u}(T)$ 与参考图像 $\mathbf{u}^*$ 之间的差异，该参考图像是使用训练数据集中的完整k空间数据获得的。对于所提方法，我们将（3–2a）和（3–4a）中的损失函数设为：

$$\begin{aligned} \ell(\mathbf{u}_\Theta) = \ell(\mathbf{u}(T)) = \sum\nolimits_{i=1}^{c} \gamma \|\mathbf{u}_i(T) - \mathbf{u}_i^*\| + \\ \| |\mathcal{J}(\bar{\mathbf{u}}(T))| - \mathrm{RSS}(\mathbf{u}^*) \| + \eta \|\mathrm{RSS}(\bar{\mathbf{u}}(T)) - \mathrm{RSS}(\mathbf{u}^*)\|, \end{aligned} \tag{3–5}$$

其中，$\mathrm{RSS}(\mathbf{u}^*) = (\sum\nolimits_{i=1}^{c} |\mathbf{u}_i^*|^2)^{1/2} \in \mathbb{R}^{m \times n}$ 表示u^*的c 个通道上的逐点平方和开方，$|\cdot|$ 表示逐点模，而 γ、$\eta > 0$ 为预设的权重参数。

在模型中应用可学习的图像空间正则化 $R \circ \mathcal{J}$ 和 k 空间正则化 $R_f \circ \mathbf{F}$ （3–4）的动机如下所述：(i) 图像域网络能够恢复高空间分辨率，但可能无法抑制某些伪影。频域网络更适合去除高频伪影。(ii) 由于全局线性变换，图像域信息与 k 空间信息是等价的，但在卷积神经网络中加入非线性激活可以有效提升网络学习的功效，并提高重建性能。我们将组合算子 $\mathcal{J}$ 参数化为

卷积神经网络，因为部分 k 空间数据由多个线圈阵列采集，并在通道与线圈图像之间引入了互相关，这与卷积神经网络结构相兼容。

3.3.3 设计变分模型中的正则化

将函数$h(\mathbf{u})$ 记为数据保真项。解决问题 $\min_u h(\mathbf{u}) + R(\mathbf{u})$ 的一种著名传统方法是近端梯度算法 [138]:

$$\mathbf{b}^l = \mathbf{u}^l - \alpha^l \nabla h(\mathbf{u}^l), \tag{3–6a}$$

$$\mathbf{u}^{l+1} = \mathrm{prox}_{\alpha^l R}(\mathbf{b}^l), \tag{3–6b}$$

其中， $\alpha^l > 0$ 为步长，邻近算子 $\mathbf{prox}_{\alpha R}$ 定义如下 :

$$\mathrm{prox}_{\alpha R}(\mathbf{b}) := \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\alpha} \|\mathbf{x} - \mathbf{b}\|^2 + R(\mathbf{x}) \right\}, \tag{3–7}$$

其中 $[\mathbf{x}]_i = \mathbf{x}_i \in \mathbb{C}^{m \times n}$ 对于任意 $\mathbf{x} \in \mathbb{C}^{m \times n \times c}$。所提出的网络结构受（3–6）启发，用于求解（3–2b），该问题可表示为离散动力系统（T 个阶段的迭代过程，其中包括（3–4b）和（3–4c）。我们将第 t 阶段参数化为以下三个步骤:

$$\mathbf{b}_i(t) = \mathbf{u}_i(t) - \rho_t \mathbf{F}^H \mathbf{P}^\top (\mathbf{P}\mathbf{F}\mathbf{u}_i(t) - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{3–8a}$$

$$\bar{\mathbf{u}}_i(t) = [\mathrm{prox}_{\rho_t R(\mathcal{J}(\cdot))}(\mathbf{b}(t))]_i, \quad i = 1, \cdots, c, \tag{3–8b}$$

$$\mathbf{u}_i(t+1) = [\mathrm{prox}_{\rho_t R_f(\mathsf{F}(\cdot))}(\bar{\mathbf{u}}(t))]_i, \quad i = 1, \cdots, c, \tag{3–8c}$$

对于 $t = 0, \cdots, T-1$ 和 $\mathbf{b}(t) = (\mathbf{b}_1(t), \ldots, \mathbf{b}_c(t)) \in \mathbb{C}^{m \times n \times c}$, $\rho_t > 0$ 为步长。F 表示归一化离散傅里叶变换，F^H 表示F的共轭转置（即厄米转置），这里F_H为逆离散傅里叶变换。

步骤（3–8a）通过应用**梯度下降算法**最小化（*3–3*）中的数据保真项来计算 b(t)，该项易于计算。配备联合正则化项 $\mathrm{R}(\mathcal{J}(\cdot))$ 的第一个近端更新步（**3–8b**）将其输入升级为

$b_i(t)$ 到多线圈图像 $\bar{u}_i(t)$。理想情况下，正则化项 $R \circ \mathcal{J}$可以参数化为一个参数能够从数据中自适应学习的深度神经网络；然而，在这种情况下，$\text{prox}_{\rho t}R(\mathcal{J}(\cdot))$ 不具有闭式解，计算起来可能较为困难。作为替代方案，可以将邻近算子 $\text{prox}_{\rho t}R(\mathcal{J}(\cdot))$ 直接参数化为可学习去噪器，并在每次迭代中求解（3–8b）。出于类似原因，（3–8c）中的邻近算子 $\text{prox}_{R_f}(F(\cdot))$ 也可以参数化为卷积神经网络，从而进一步提高 k空间测量的准确率。在（3–8b）和（3–8c）中，正则化项 $R \circ \mathcal{J}$和 $R_f \circ F$ 通过神经网络提取复数特征，并且邻近点可以通过具有残差网络 [73] 结构的去噪网络进行学习。

我们将步骤 (3–8b) 表述为通过残差网络 [73] 执行联合重建，以更新线圈图像：$\bar{u}(t) = b(t) + \mathcal{M}(b(t))$，其中 $\mathcal{M}$ 表示多层卷积神经网络，通过在图像空间中近似近端映射来实现。我们提出先学习一个非线性算子 $\mathcal{J}$，将 $\{b_i\}$ 组合成具有均匀对比度的图像 $z = \mathcal{J}(b_1, \dots, b_c) \in \mathbb{C}^{m\times n}$。然后对 z 应用非线性算子 $\mathcal{G}$，通过 $\mathcal{G}(z) \in \mathbb{C}^{m\times n\times N_f}$ 提取 N_f 维特征。非线性算子 $\mathcal{J}$ 包含四个卷积层，层间使用激活函数，每个卷积层的卷积核大小为 3×3。非线性算子 $\mathcal{G}$ 由三个滤波器尺寸为 9×9的卷积层构成。为提升网络的能力，$\mathcal{J}$ 和 $\mathcal{G}$ 分别作为 $\mathcal{J}$ 和 $\mathcal{G}$ 的伴随算子，二者采用对称结构，参数则分别独立训练。$\mathcal{G} \circ \mathcal{J}$ 旨在发挥编码器的作用，而 $\mathcal{J} \circ \mathcal{G}$ 则发挥解码器的作用。因此，图像域网络可参数化为四个卷积神经网络算子的复合：$\mathcal{M} = \mathcal{J} \circ \mathcal{G} \circ \mathcal{G} \circ \mathcal{J}$，其输出为 $\mathcal{M}(b(t)) \in \mathbb{C}^{m\times n\times c}$。步骤 (3–8b) 从图像域网络输出多线圈数据 $\bar{u} = (\bar{u}_1, \dots, \bar{u}_c) \in \mathbb{C}^{m\times n\times c}$，我们对 $\bar{u}$ 应用组合算子 $\mathcal{J}$，以获得我们期望用于重建的全视野磁共振单通道图像 $v = \mathcal{J}(\bar{u}) \in \mathbb{C}^{m\times n}$。

步骤（3–8c）利用 k空间信息，进一步抑制高频伪影。将（$\bar{\mathbf{u}}(t)$ 来自（3–8b）的输出传递至一个残差网络构成的 k空间域网络

结构 $\mathbf{u}(t+1) = \bar{\mathbf{u}}(t) + \mathbf{F}^H\mathcal{K}\big(\mathbf{F}(\bar{\mathbf{u}}(t))\big)$，其中 $\mathbf{F}^H\mathcal{K}\mathbf{F}$ 是一个卷积神经网络算子，用于细化并进一步提高各线圈 k空间数据的准确率。该卷积神经网络 $\mathcal{K}$ 由四个卷积层组成，最后一个卷积层的卷积核数量与线圈图像的通道数相匹配。

因此，(3–8)按以下方案进行：

$$\mathbf{b}_i(t) = \mathbf{u}_i(t) - \rho_t\mathbf{F}^H\mathbf{P}^\top(\mathbf{P}\mathbf{F}\mathbf{u}_i(t) - \mathbf{f}_i), \quad i = 1, \cdots, c, \tag{3–9a}$$

$$\bar{\mathbf{u}}_i(t) = \mathbf{b}_i(t) + \mathcal{M}(\mathbf{b}_i(t)), \quad i = 1, \cdots, c, \tag{3–9b}$$

$$\mathbf{u}_i(t+1) = \bar{\mathbf{u}}_i(t) + \mathbf{F}^H\mathcal{K}\big(\mathbf{F}(\bar{\mathbf{u}}_i(t))\big), \quad i = 1, \cdots, c, \tag{3–9c}$$

对于 $t = 0, \cdots, T-1$。现在，我们可以通过结合（$\mathbf{g}$）中描述的函数（3–4）与（3–9）推导出该函数：

$$\begin{aligned}\mathbf{u}_i(t+1) = {} & (\mathcal{N} - \rho_t\mathbf{F}^H\mathbf{P}^\top\mathbf{P}\mathbf{F} - \mathcal{N}(\rho_t\mathbf{F}^H\mathbf{P}^\top\mathbf{P}\mathbf{F}))\mathbf{u}_i(t) \\ & + \mathbf{u}_i(t) + \rho_t\mathbf{F}^H\mathbf{P}^\top\mathbf{f}_i - \mathcal{N}(\rho_t\mathbf{F}^H\mathbf{P}^\top\mathbf{f}_i) && \text{(3–10a)} \\ = {} & \mathbf{g}(\mathbf{u}_i(t)), && \text{(3–10b)}\end{aligned}$$

其中 $\mathcal{N} = \mathcal{M} + \mathbf{F}^H\mathcal{K}\mathbf{F} + \mathbf{F}^H\mathcal{K}\mathbf{F}\mathcal{M}$，函数 $\mathbf{g}$ 将 $\mathbf{u}_i(t)$ 映射为 $\mathbf{u}_i(t+1)$，其中 $t = 0, \cdots, T-1$ 由（3–10a）定义。

所提出的重建网络由预设的 T 个阶段组成。网络的初始输入设计为 $\mathrm{u}(0) = \mathrm{g}_0(\mathrm{f}, \theta(0)) := \mathrm{F}^H(\mathrm{f} + \mathcal{K}_0(\mathrm{f}))$，其中 $\mathcal{K}_0$ 是一个卷积神经网络算子，在残差学习中作用于 f，以插值缺失数据并生成伪全 k 空间。网络以选定的初始值 $\{\mathrm{u}_i(0)\}$ 和部分 k 空间数据 $\{\mathrm{f}_i\}$ 为输入，在第 t 个阶段执行更新（3–9），其中 $\mathrm{t} = 0, \cdots, \mathrm{T}-1$；最终，整个网络重建出线圈图像 $\mathrm{u}(\mathrm{T})$ 和 $\mathcal{J}(\bar{\mathrm{u}}(\mathrm{T}))$，后者是作为副产品重建得到的最终单体图像（复值）。

图 3-1 展示了我们整个网络流程的流程图。初始重建可抑制由欠采样造成的伪影。图 3-2 展示了各阶段这些卷积神经网络结构的流程图。我们采用复值卷积

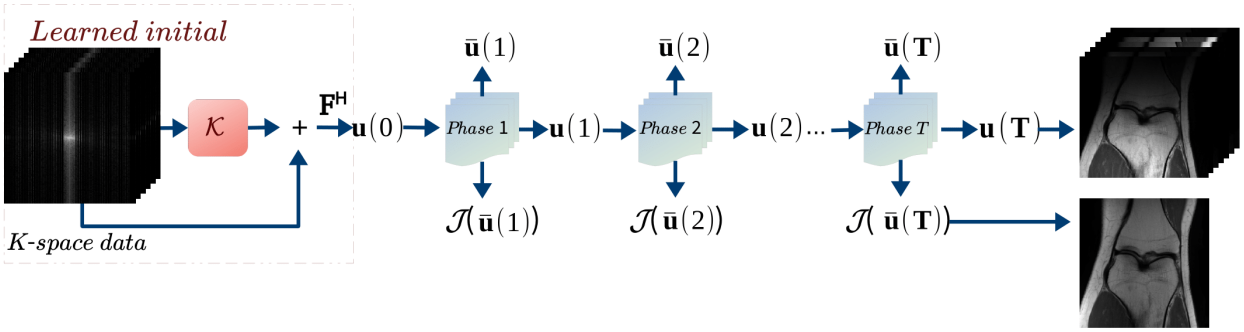

图 3-1. 所提出的框架范式适用于所有阶段。

，在其中执行复数之间的乘法，并使用逐分量复值激活函数 $\mathbb{C}\mathrm{ReLU}(a + ib) = \mathrm{ReLU}(a) + i\mathrm{ReLU}(b)$，如[37]所建议。Landweber 更新步骤（3–9a）在增强图像空间与k空间之间的交互方面发挥作用。可学习步长 ρ_t 控制收敛的速度和稳定性。在图像空间去噪步骤（3–9b）中，算子 $\mathcal{J}$ 提取所有通道的特征，从而提升空间分辨率，并在通道合并图像中恢复组织细节。具有通道整合功能的卷积神经网络 $\mathcal{M}$ $\mathcal{J}$在每两个阶段之间以共享权重进行 T 次处理，因此能够高效地学习跨通道的空间特征。然而，振荡伪影可能会被误认为是真实特征，并因此被锐化。在k空间去噪更新步骤（3–9c）中，k空间网络 $\mathcal{K}$ 与低频信息相兼容，因此能够去除高频伪影并恢复图像结构 [51, 192]。因此，在不同域中迭代网络 $\mathcal{M}$ 和 $\mathcal{K}$ ，利用各自的作用，使其性能相互补偿、缺点相互抵消。我们通过消融研究评估混合域重建的效果。此外，（3–9b）和（3–9c）中规定的去噪网络会在每次迭代中细化并更新线圈图像。该迭代过程使$\mathcal{J}$（¯u(t)）的重建质量逐步提升。

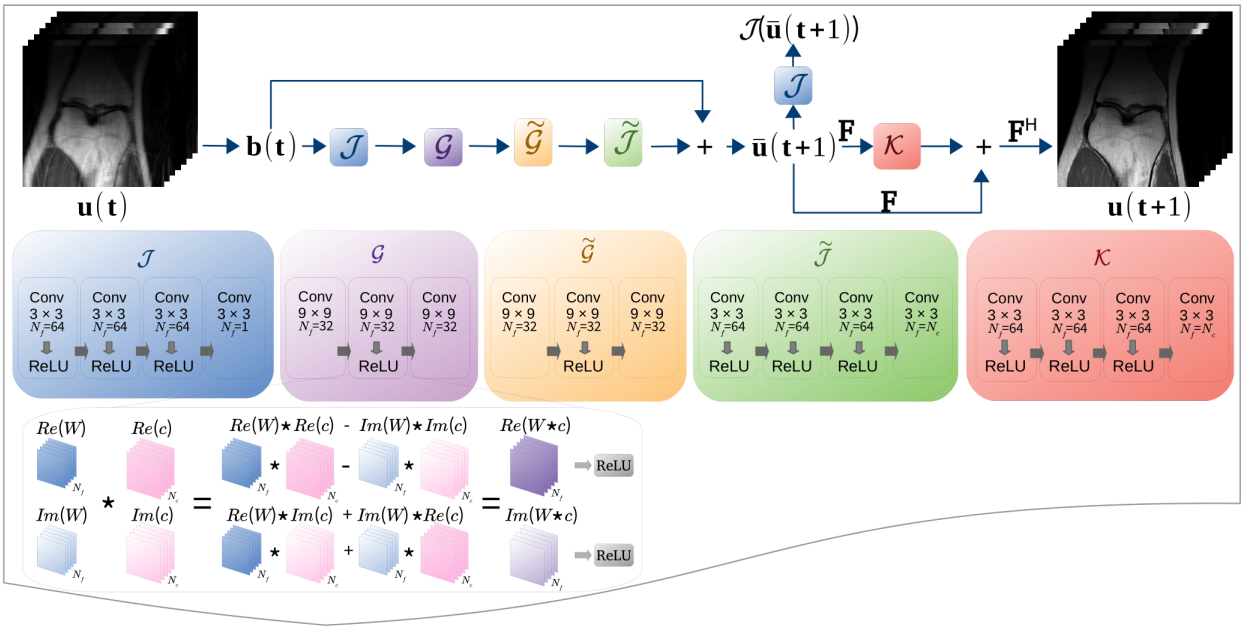

图3-2。第$t+1$阶段所提出的框架范式。$W=Re(W)+iIm(W)$ 表示复数卷积滤波器核。

3.3.4 从拉格朗日乘子法（MLM）的角度看网络训练

需要从（3–4）求解的网络参数为 $\Theta=\{\theta(t): t=0,\cdots,T\}$，其中 $\theta(t)=\{\rho_t,\mathcal{J}_t,\mathcal{G}_t,\tilde{\mathcal{G}}_t,\tilde{\mathcal{J}}_t,\mathcal{K}_t\}$ 当 $t=1$时，$\cdots$，T以及 $\theta(0)=\mathcal{K}_0$。

控制问题（3–4）可以使用 MLM 求解。相应的拉格朗日函数为

$$\begin{aligned}L(\mathbf{U},\Theta;\Lambda)=&\ell(\mathbf{u}(T))+\langle\lambda(0),\mathbf{u}(0)-\mathbf{g}_0(\mathbf{f},\theta(0))\rangle\\&+\sum_{t=1}^{T}\langle\lambda(t),\mathbf{u}(t)-\mathbf{g}(\mathbf{u}(t-1),\theta(t))\rangle,\end{aligned}\tag{3–11}$$

其中，$\mathbf{U}=(\mathbf{u}(0),\cdots,\mathbf{u}(\mathrm{T}))^{\mathrm{T}}$是所有状态 $\mathbf{u}(t)$ 的集合，$\Lambda=(\lambda(0),\cdots,\lambda(\mathrm{T}))$ 是式（3–4）的拉格朗日乘子。该算法采用迭代方案更新 Θ^k；对于每个训练轮次 $\mathbf{k}=0,\cdots,\mathbf{K}$，更新 $\Theta^k=(\theta^k(0),\cdots,\theta^k(\mathrm{T}))$、$\mathbf{U}^k=(\mathbf{u}^k(0),\cdots,\mathbf{u}^k(\mathrm{T}))$ 和 $\Lambda^k=(\lambda^k(0),\cdots,\lambda^k(\mathrm{T}))$。

如果$(\mathbf{U}^*, \Theta^*; \Lambda^*)$ 使拉格朗日函数（3–11）最小化，则根据一阶最优性条件，有

$$\begin{cases} \partial_{\mathbf{U}} L(\mathbf{U}^*, \Theta^*; \Lambda^*) = 0 & \text{(3–12a)} \\ \partial_{\Theta} L(\mathbf{U}^*, \Theta^*; \Lambda^*) = 0 & \text{(3–12b)} \\ \partial_{\Lambda} L(\mathbf{U}^*, \Theta^*; \Lambda^*) = 0 & \text{(3–12c)} \end{cases}$$

•固定 $\Theta = \Theta_k$，定义 $(\mathbf{U}_k, \Lambda_k) := \arg\min_{\mathbf{U},\Lambda} L(\mathbf{U}, \Theta_k; \Lambda)$，则根据关于 $\lambda(t)$ 最小化 L 的一阶最优性条件，对于 $t = 0, \cdots, T$，$(\mathbf{U}_k, \Lambda_k)$ 应满足

$$\begin{aligned} &\partial_{\lambda(0)} \langle \lambda(0), \mathbf{u}(0) - \mathbf{g}_0(\mathbf{f}, \theta^k(0)) \rangle \Big|_{(\mathrm{u}^k(0), \lambda^k(0))} = 0 \\ &\Longrightarrow \ \mathbf{u}^k(0) = \mathbf{g}_0(\mathbf{f}, \theta^k(0)) \end{aligned} \tag{3–13}$$

$$\begin{aligned} &\partial_{\lambda(t)} \langle \lambda(t), \mathbf{u}(t) - \mathbf{g}(\mathbf{u}(t-1), \theta^k(t)) \rangle \Big|_{(\mathrm{u}^k(t), \lambda^k(t))} = 0 \\ &\Longrightarrow \ \mathbf{u}^k(t) = \mathbf{g}(\mathbf{u}^k(t-1), \theta^k(t)), \ t = 1, \cdots, T. \end{aligned} \tag{3–14}$$

那么，根据关于$\mathbf{u}(t)$ 最小化 L 的一阶最优性条件，对于 $t = T$，可得

$$\begin{aligned} &\partial_{\mathrm{u}(T)} [\ell(\mathbf{u}(T)) + \langle \lambda(t), \mathbf{u}(T) \rangle] \Big|_{(\mathrm{u}^k(t), \lambda^k(t))} = 0 \\ &\Longrightarrow \ \lambda^k(T) = -\partial_{\mathrm{u}(T)} \ell(\mathbf{u}^k(T)), \end{aligned} \tag{3–15}$$

对于 $t = 0, \cdots, T-1$:

$$\begin{aligned} &\partial_{\mathrm{u}(t)} [\langle \lambda(t), \mathbf{u}(t) \rangle - \langle \lambda(t+1), \mathbf{g}(\mathbf{u}(t), \theta^k(t+1)) \rangle] \Big|_{(\mathrm{u}^k(t), \lambda^k(t))} = 0 \\ &\Longrightarrow \ \lambda^k(t) = \langle \lambda^k(t+1), \partial_{\mathrm{u}(t)} \mathbf{g}(\mathbf{u}^k(t), \theta^k(t+1)) \rangle, \end{aligned} \tag{3–16}$$

•固定 $(\mathbf{U}_k, \Lambda_k)$ 以更新 Θ，我们计算梯度 $\partial_{\Theta} L(\mathbf{U}_k, \Theta; \Lambda_k)$:

$$\begin{aligned} &\partial_{\theta(t)} L(\mathbf{U}^k, \theta(t); \Lambda^k) = \partial_{\theta(t)} [-\langle \lambda^k(t), \mathbf{g}(\mathbf{u}^k(t-1), \theta(t)) \rangle] \\ &= -\langle \lambda^k(t), \partial_{\theta(t)} \ \mathbf{g}(\mathbf{u}^k(t-1), \theta(t)) \rangle, t = 1, \cdots, T, \end{aligned} \tag{3–17a}$$

$$\partial_{\theta(0)} L(\mathbf{U}^k, \theta(0); \Lambda^k) = -\langle \lambda^k(0), \partial_{\theta(0)} \mathbf{g}_0(\mathbf{f}, \theta(0)) \rangle. \tag{3–17b}$$

定理 3.1。 $\partial_{\Theta} L(\boldsymbol{U}^k, \Theta; \ \Lambda^k) = \partial_{\Theta} \ell(\mathbf{u}^k(T)(\Theta))$。

证明。 首先证明以下结论对$t = 0, \cdots, T$成立：

$$\lambda^k(t) = -\partial_{\mathrm{u}(t)}\ell(\mathbf{u}^k(T)) \tag{3–18}$$

由（3–15），当 $t = T$ 时，（3–18）成立。假设当 $t = \tau \in \{1, \cdots, T\}$ 时，（3–16）成立，则有

$$\begin{aligned}\lambda^k(\tau-1) &= \langle \lambda^k(\tau), \partial_{\mathrm{u}(\tau-1)}\, \mathbf{g}(\mathbf{u}^k(\tau-1), \theta^k(\tau))\rangle \\ &= \langle -\partial_{\mathrm{u}(\tau)}\ell(\mathbf{u}^k(T)), \partial_{\mathrm{u}(\tau-1)}\mathbf{u}^k(\tau)\rangle \qquad (3–19a) \\ &= -\partial_{\mathrm{u}(\tau-1)}\ell(\mathbf{u}^k(T)) \qquad (3–19b)\end{aligned}$$

因此，（3–18）对 $t = \tau - 1$ 成立。根据数学归纳法，（3–18）对所有$t = 0, \cdots, T$ 均成立。

因此，（3–17a）可化简为

$$\begin{aligned}\partial_{\theta(t)}L(\mathbf{U}^k, \theta(t); \Lambda^k) &= -\langle \lambda^k(t), \partial_{\theta(t)}\, \mathbf{g}(\mathbf{u}^k(t-1), \theta(t))\rangle \\ &= \langle \partial_{\mathrm{u}(t)}\ell(\mathbf{u}^k(T)), \partial_{\theta(t)}\mathbf{u}^k(t)\rangle \qquad (3–20a) \\ &= \partial_{\theta(t)}\ell(\mathbf{u}^k(T)),\ t = 1, \cdots, T. \qquad (3–20b)\end{aligned}$$

此外，（3–17b）给出 $\partial_{\theta(0)}\mathrm{L}(\mathrm{U}^k, \theta(0);\ \Lambda^k) = \partial_{\theta(0)}\ell(\mathrm{u}^k(\mathrm{T}))$。因此，我们推导得出 $\partial_{\Theta}\mathrm{L}(\mathrm{U}^k, \Theta;\ \Lambda^k) = \partial_{\Theta}\ell(\mathrm{u}^k(\mathrm{T})(\Theta))$。 □

该定理进一步表明：由于梯度相同，使用随机梯度下降算法或其变体（如 Adam）[93] 来最小化损失函数 ℓ 等价于在L上执行相同的算法。使用 MLM 的网络训练算法可总结于算法 1。

3.4 实验结果

3.4.1 数据集

我们实验中的数据由一个15通道膝关节线圈阵列采集，使用了两种脉冲序列：一种是带有（FSPD）的质子密度加权，另一种是不带

算法 1： 使用 MLM 进行网络训练

超参数： K （迭代次数）
初始化： 初始猜测 $\theta^0(t) \in \Theta^0, t = 0, \cdots, \quad T$
对于 $k = 0$ **到** $K - 1$，**执行**
  设 $\mathbf{u}^k(0) = \mathbf{g}_0(\mathbf{f}, \theta(0))$
  对于 $t = 1$ **到** T，**执行**
    $\mathbf{u}^k(t) = \mathbf{g}(\mathbf{u}^k(t-1), \theta^k(t))$
  end
  集合 $\lambda^k(T) = -\partial_u \ell(\mathbf{u}^k(T))$
  对于 $t = T - 1$ **到** 0 **执行**
    $\lambda^k(t) = \langle \lambda^k(t+1), \partial_u \mathbf{g}(\mathbf{u}^k(t), \theta^k(t+1)) \rangle$
  end
  对于 $t = 0$ **到** T **执行**
    设定 $\theta^{k+1}(t) = \arg\min_{\theta(t)} L(\mathbf{U}^k, \theta(t); \Lambda^k)$.
  end
结束
输出： $\theta^K(t), t = 0, \cdots, \quad T$.

（PD）脂肪抑制，采集方向为冠状位，数据来自 https://github.com/VLOGroup/mri-variationalnetwork。我们使用了规则笛卡尔采样掩模，采样率为31.56%，如图3-4右下角所示。两种序列的数据均包含20名患者的图像。我们从19名患者中选取了27–28张中心图像切片，共计526张图像，每张图像的尺寸为 320×320，作为训练数据集；另从一名未纳入训练数据集的患者中选取15张中心图像切片，作为测试数据集。

3.4.2 实现

我们在 15 幅测试用冠状位 FSPD 和冠状位 PD 膝关节图像上，采用规则笛卡尔采样，对经典方法 GRAPPA [65], SPIRiT [119], 和深度学习方法 VN [72], De-AliasingNet [31], DeepcomplexMRI [191]及 Adaptive-CS-Net[140] 进行评估，评估指标为峰值信噪比（PSNR）、结构相似性指数（SSIM） [193] 和相对误差均方根。以下方程用于计算重建图像 $\mathbf{v} = |\mathcal{J}(\bar{\mathbf{u}})|$ 与真实值单体图像 $\mathbf{v}^*$之间的 SSIM、PSNR 和 RMSE：

$$SSIM = \frac{(2\mu_v\mu_{v^*} + C_1)(2\sigma_{vv^*} + C_2)}{(\mu_v^2 + \mu_{v^*}^2 + C_1)(\sigma_v^2 + \sigma_{v^*}^2 + C_2)}, \tag{3–21}$$

其中， $\mu_{\mathbf{v}}$、 $\mu_{\mathbf{v}^*}$ 为像素强度的局部均值， $\sigma_{\mathbf{v}}$、 $\sigma_{\mathbf{v}^*}$ 表示标准差， $\sigma_{\mathbf{v}\mathbf{v}^*}$ 为$\mathbf{v}$ 与$\mathbf{v}^*$之间的协方差；$C_1 = (k_1 L)^2$,和$C_2 = (k_2 L)^2$ 是两个用于避免分母为零的常数，且$k_1 = 0.01$、$k_2 = 0.03$。L 为线圈图像幅度的最大像素值。

$$PSNR = 20 \log_{10} \left(\max(|\mathbf{v}^*|) / \frac{1}{N} \|\mathbf{v}^* - |\mathcal{J}(\bar{\mathbf{u}})|\|^2 \right), \tag{3–22}$$

其中 N 是真实值幅度中的像素总数 .

$$RMSE = \|\mathbf{v}^* - |\mathcal{J}(\bar{\mathbf{u}})|\| / \|\mathbf{v}^*\|. \tag{3–23}$$

多线圈重建 $\mathbf{u}$ 与真实值$\mathbf{u}^*$ 之间的相对误差定义为

$$RMSE = \sqrt{\sum_{i=1}^{C} \|\mathbf{u}_i^* - \mathbf{u}_i\|^2 / \sum_{i=1}^{C} \|\mathbf{u}_i^*\|^2}. \tag{3–24}$$

所有实验均在 TensorFlow [2] 中实现并测试，运行于配备 3.3GHz Intel Core i9 CPU 和具有 11GB 显卡内存的 Nvidia GTX-1080Ti GPU 的 Windows 工作站上。所提出的网络中的参数采用 Xavier 初始化 [60] 来初始化 $\theta^0(\mathrm{t})$， $\mathrm{t} = 0, \cdots, \mathrm{T}$。实际上，使用 Adam 等任意随机梯度下降算法 [93] 求解算法 1 中的第 13 行，其性能与使用相同算法最小化关于 $\theta(\mathrm{t})$的损失函数并更新$\theta^{k+1}(\mathrm{t})$相同，这是因为梯度等价： $\partial_{\Theta} \mathrm{L}(\mathrm{U}^k, \Theta; \Lambda^k) = \partial_{\Theta} \ell(\mathrm{u}^k(\mathrm{T})(\Theta))$，这一点由定理 3.1 证明。TensorFlow 提供了用于自动微分的优化 API，有助于构建高性能输入流水线并保持较高的 GPU利用率。为实现更稳定的梯度计算，我们将第 13 行替换为使用 Adam算法最小化 ℓ 。网络经过总计 $\mathrm{K} = 700$ 个训练轮次，以更新 Θ^k。我们采用指数衰减学习率 0.0001,$\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$， 并在 Adam优化器中使用批量大小为 2 的小批量。实部和虚部的初始步长均设为 $\rho_0 = 1$ ， 并且我们选择

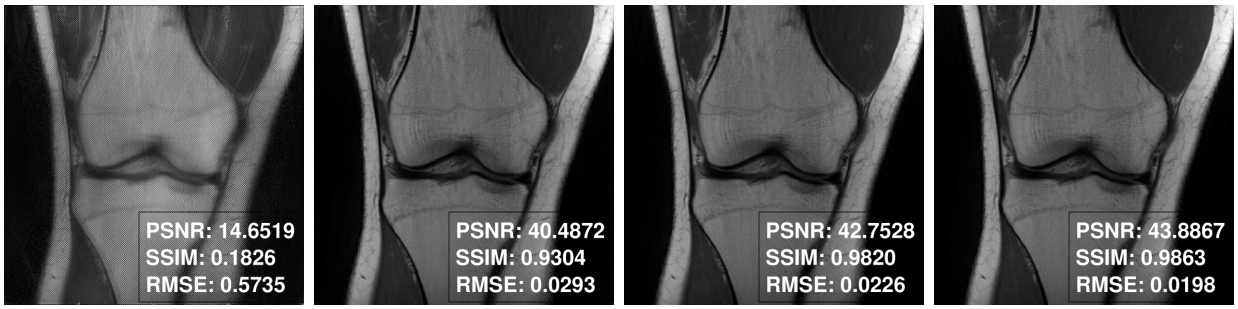

图 3-3. 所提方法在$t = 1$、2、3、4 阶段的重建的通道合并图像 $\mathcal{J}(\bar{\mathbf{u}}(t))$。

表 3-2. FSPD数据和 PD 数据在各阶段的测试平均峰值信噪比。

阶段编号	FSPD数据的峰值信噪比	PD 数据的峰值信噪比
1	16.2007	15.5577
2	33.7201	40.5228
3	36.9523	42.6623
4	40.7101	44.8120

$\gamma = 10^{-3}$，$\eta = 10^{-4}$ 在（3–5）中。所提出的方法网络采用 $T = 4$ 实现，且网络中训练的参数 $\mathcal{M}$ 在每两个阶段之间共享。

3.4.3 与现有方法的比较

所提方法与若干最先进方法的平均数值表现及标准差汇总于表 3-3。所提方法在 PSNR/SSIM/RMSE 方面取得了最佳重建质量。由于网络 $\mathcal{M}$在每两个阶段共享参数，整个网络的可学习参数减少了 1/4 以上，因此我们的方法具有较高的参数效率。我们在表中列出了训练后的参数数量和推理时间 3-3。图 3-3 和表 3-2 表明，随着t 增大，重建性能逐步提升。GRAPPA 和 SPIRiT 采用的校准核大小为 5×5，校准中的蒂霍诺夫正则化设为 0.01。重建中的蒂霍诺夫正则化设为10^{-3}，以实现 SPIRiT，共进行了 30 次迭代。VN 和 DeepcomplexMRI 的训练与测试中，网络及参数设置均按其论文和代码中的说明进行。我们通过将滤波器数量增加至 64，对 De-AliasingNet 进行了修改，迭代次数为 20

次，以提升性能并确保公平比较。在 CDF-Net 中，我们将三个频率信息引导 U-Net 的初始层均设为 128 个通道，深度设为 4。我们选择 σ 作为 Softplus 激活函数。我们实现的 Adaptive-CS-Net 总共包含 15 个重建块，其中 2 个使用大小为 3×3 的 16 个滤波器，3 个使用大小为 5×5 的 32 个滤波器，2 个使用大小为 5×5 的 64 个滤波器；这些模块混合使用 2 种和 3 种尺度，其设置受图形处理器容量所允许的最大容差限制。我们将数据一致性作为训练数据的先验知识输入网络，并将中心切片与相邻切片一起作为输入。为确保公平比较，CDF-Net 和 Adaptive-CS-Net 均采用了复数卷积和激活函数 $\mathbb{C}$ReLU。

VN 应用了预先计算的线圈灵敏度，并输出全身图像。De-AliasingNet、DeepcomplexMRI、CDF-Net 和 Adaptive-CS-Net 均输出多线圈图像，且不需要线圈灵敏度图。DeepcomplexMRI 使用自适应线圈组合方法 [188], ，其他网络均采用平方和开根号。CDF-Net 和所提网络执行跨域重建，其他基于学习的方法则在图像域执行重建。参考的并行磁共振成像方法与所提出的方法之间的比较见图 3-4 ，质子密度图像和脂肪抑制质子密度图像见图 3-5。第一行显示重建图像和参考图像，第二行和第三行显示红框区域（在最右侧的真实图像上标出）对应的放大感兴趣区域，第四行显示对应的逐点绝对误差图，最后一行显示对应的数值及规则笛卡尔采样（采样率为 31.56%）掩模。我们观察到，基于深度学习的方法在重建精度方面显著优于 GRAPPA 和 SPIRiT 等经典方法。在基于学习的方法中，VN 和 De-AliasingNet 重建的图像比其他方法更模糊，在某些复杂组织中会丢失清晰细节；其他方法在细节上仅有轻微差异。

表3-3. 参考方法在冠状位FSPD与PD数据上的重建图像定量评估及各方法的重建时间。时间（秒）指各方法的测试时间。

方法	FSPD数据 PSNR	SSIM	RMSE	PD 数据 PSNR	SSIM	RMSE	时间	参数
GRAPPA [65]	24.9251±0.9341	0.4827±0.0344	0.2384±0.0175	30.4154±0.5924	0.7489±0.0207	0.0984±0.0030	280s	N/A
SPIRiT [119]	28.3525±1.3314	0.6509±0.0300	0.1614±0.0203	32.0011±0.7920	0.7979±0.0306	0.0824±0.0082	43s	N/A
越南 [72]	30.2588±1.1790	0.7141±0.0483	0.1358±0.0152	37.8265±0.4000	0.9281±0.0114	0.0422±0.0036	0.16秒	0.13 M
De-AliasingNet [31]	36.1017±1.2981	0.8941±0.0269	0.0697±0.0119	41.2151±0.7872	0.9711±0.0033	0.0285±0.0015	0.92秒	0.34 M
DeepcomplexMRI [191]	36.5706±1.0215	0.9008±0.0190	0.0654±0.0049	41.5756±0.6271	0.9679±0.0031	0.0274±0.0018	1.04秒	1.45 M
CDF-Net [136]	40.0405±1.5518	0.9520±0.0182	0.0443±0.0070	43.7943±1.9888	0.9879±0.0026	0.0215±0.0042	0.47秒	3.44 M
Adaptive-CS-Net [140]	40.1846±1.4780	0.9534±0.0175	0.0435±0.0065	44.1131±1.5596	0.9878±0.0026	0.0206±0.0027	1.57秒	5.59 M
所提出的方法	**40.7101±1.5357**	**0.9619±0.0144**	**0.0408±0.0051**	**44.8120±1.3185**	**0.9886±0.0023**	**0.0189±0.0018**	0.52秒	2.92 M

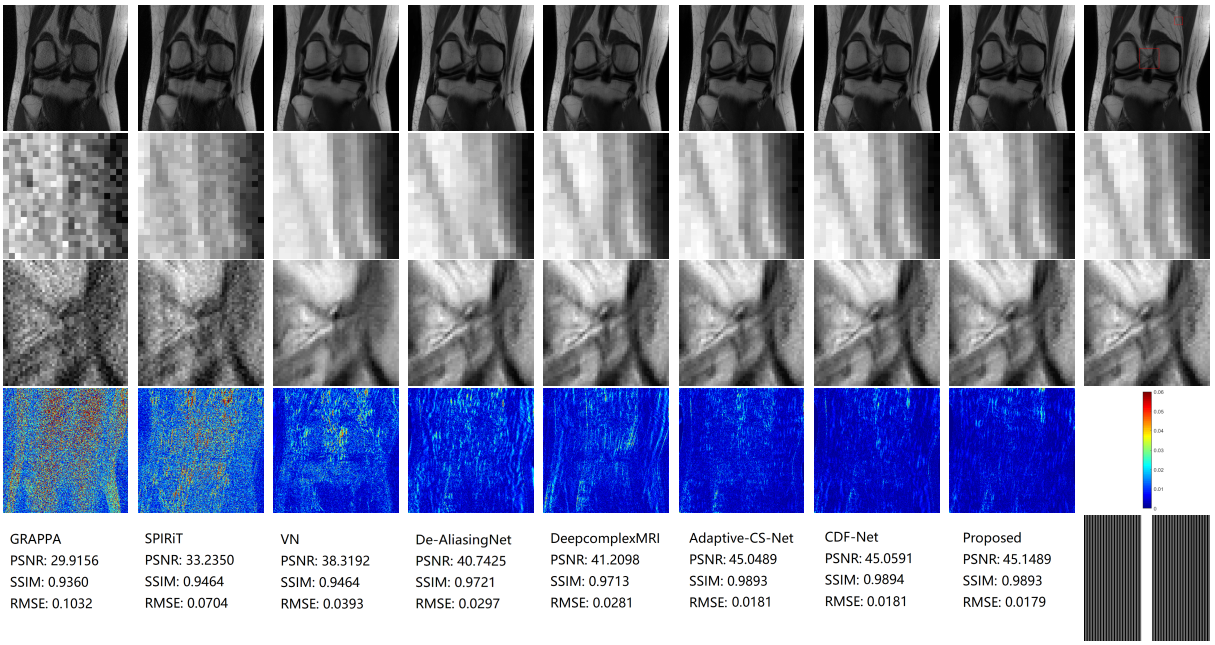

图 3-4. 冠状位 PD 膝关节图像上各重建方法的定性比较结果。

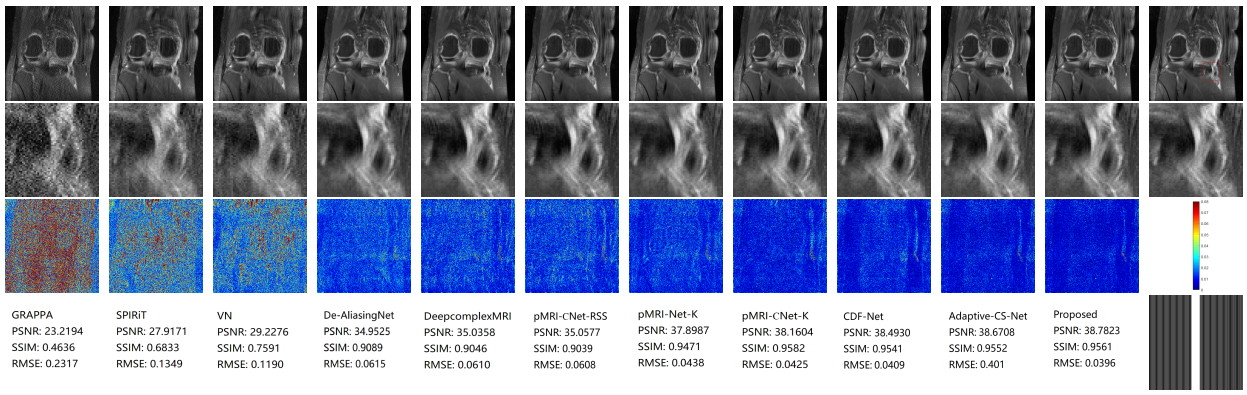

图 3-5. 冠状位 FSPD 膝关节图像上各重建方法的定性比较结果。

表 3-4. 与重建多线圈图像的方法相比，PD 数据中$\mathbf{u}$ 与$\mathbf{u}^*$ 之间的均方误差。

方法	$\mathbf{u}$ 与$\mathbf{u}^*$之间的均方根误差
De-AliasingNet [31]	0.1573
DeepcomplexMRI [191]	0.1253
CDF-Net [136]	0.1196
Adaptive-CS-Net [140]	0.1096
并行磁共振成像-ℂNet-K (3–26)	0.0538
所提出的方法 (3–5)	0.0505

3.4.4 消融研究

本节介绍的实验仅在图像域中进行，其中 (3–9) 被替换为：

$$\mathbf{b}_i(t) = \mathbf{u}_i(t) - \rho_t \mathbf{F}^H \mathbf{P}^\top (\mathbf{P}\mathbf{F}\mathbf{u}_i(t) - \mathbf{f}_i), \tag{3–25a}$$

$$\mathbf{u}_i(t+1) = \mathbf{b}_i(t) + \tilde{\mathcal{J}} \circ \tilde{\mathcal{G}} \circ \mathcal{S}_{\alpha_t}(\mathcal{G} \circ \mathcal{J}(\mathbf{b}_i(t))). \tag{3–25b}$$

$\mathcal{S}_{\alpha_t}$ 表示软收缩算子，我们对实部和虚部均设置 $\alpha_0 = 0$ 。该模型的详细信息已在会议论文 [15] 中说明。近端算子在残差更新中学习（3–25b），且仅在图像域中执行。

我们开展了一系列实验，以检验所提出的网络中若干重要组件的作用。表 3-6 展示了所提出的消融研究对比。（各变体的标签说明见表 3-5。未标注损失函数的实验均采用（3–27）进行训练。表 3-6中的所有实验均通过（3–25）实现，只有最后一个“所提出的方法”例外，该方法展开了（3–9）。为明确各组件的使用情况，我们将表 3-5 中所示的标签附加到经过消融的 pMRI 网络的不同变体上。例如，采用实值卷积和激活函数并进行零填充初始化的网络记为 pMRI-Net-ZF，等等。具有三种初始化方式的网络完整结构如图 3-6所示。对于 $t = 1, \cdots, T$，每个阶段均遵循我们先前工作中介绍的算法 [15]。

表 3-5. 所提出的并行磁共振成像重建网络各变体的标签。

标签	含义
-Net	实值卷积/激活
-ℂ网络	复值卷积/激活
-RSS	用平方和平方根代替 $\mathcal{J}$
-ZF	零填充作为初始值 $\mathbf{u}(0)$
-SP	SPIRiT 作为初始值$\mathbf{u}(0)$
-K	$\mathbf{F}^H(\mathbf{f} + \mathcal{K}(\mathbf{f}))$ 作为初始值 $\mathbf{u}(0)$

表3-6. 使用所提出的方法的不同变体对冠状位FSPD与PD数据进行重建的定量评估。

方法	FSPD数据 PSNR	SSIM	RMSE	PD 数据 PSNR	SSIM	RMSE	阶段 T	参数
并行磁共振成像-ℂ网络-平方和开根号	36.4887±0.9787	0.9002±0.0197	0.0661±0.0051	41.2897±0.8430	0.9281±0.0357	0.0285±0.0037	5	5.03 M
pMRI-Net-ZF	37.8475±1.2086	0.9212±0.0236	0.0568±0.0069	42.4333±0.8785	0.9793±0.0023	0.0249±0.0024	5	5.03 M
pMRI-Net-SP	38.0205±0.8125	0.9291±0.0183	0.0555±0.0057	42.7435±0.4856	0.9754±0.0047	0.0239±0.0019	5	5.03 M
并行磁共振成像-CNet-ZF	38.1157±1.3776	0.9277±0.0257	0.0552±0.0085	42.7859±1.1285	0.9818±0.0026	0.0241±0.0045	5	5.03 M
并行磁共振成像-CNet-SP	38.3239±1.1305	0.9282±0.0269	0.0539±0.0075	42.8924±0.9336	0.9760±0.0054	0.0237±0.0034	5	5.03 M
pMRI-Net-K	38.8717±1.1330	0.9389±0.0209	0.0504±0.0057	42.9060±0.8765	0.9802±0.0028	0.0236±0.0028	4	4.11 M
并行磁共振成像-CNet-K (3–26)	38.9661±1.4382	0.9421±0.0177	0.0498±0.0056	43.2604±0.7610	0.9833±0.0022	0.0226±0.0022	4	4.11 M
并行磁共振成像-CNet-K	39.3360±1.1854	0.9497±0.0208	0.0477±0.0048	43.5653±0.8265	0.9844±0.0022	0.0217±0.0010	4	4.11 M
所提出的方法 (3–5)	**40.7101±1.5357**	**0.9619±0.0144**	**0.0408±0.0051**	**44.8120±1.3185**	**0.9886±0.0023**	**0.0189±0.0018**	4	2.92 M

在消融实验中，我们考虑两种类型的训练数据集，并在第 3.4.4 节为每种数据集设计了合适的损失函数。第一种情况下，真实值为$\mathbf{u}^*$，我们设定损失函数为：

$$\begin{aligned}\ell(\mathbf{u}) = \sum_{i=1}^{C} \gamma \|\mathbf{u}_i - \mathbf{u}_i^*\| + \| |\mathcal{J}(\mathbf{u})| - \mathrm{RSS}(\mathbf{u}^*)\| \\ + \beta \|\mathrm{RSS}(\mathbf{u}_i(0)) - \mathrm{RSS}(\mathbf{u}_i^*)\|,\end{aligned} \tag{3–26}$$

为简便起见， $\mathbf{u}$ 表示$\mathbf{u}(T)$。如果训练数据集由真实值单体图像$\mathbf{v}^* \in \mathbb{R}^{m\times n}$ = 平方和开根号$(\mathbf{u}^*)$组成，那么我们使用以下损失函数测试网络的性能：

$$\ell(\mathbf{u}) = \gamma \|\mathrm{RSS}(\mathbf{u}) - \mathbf{v}^*\| + \| |\mathcal{J}(\mathbf{u})| - \mathbf{v}^*\| + \beta \|\mathrm{RSS}(\mathbf{u}(0)) - \mathbf{v}^*\|. \tag{3–27}$$

损失函数列于表3-6 中，并且3-4 是在实验之后得到的。我们在实现中设置$\gamma = 1$、$\beta = 10^{-3}$ 在（3–26）和（3–27）中。

3.4.4.1 组合算子与平方和平方根

为了验证学习得到的非线性组合算子 $\mathcal{J}$的有效性，我们通过将 $\mathcal{J}$ 替换为广泛用于组合ℂ的 RSS，对 pMRI-ℂNet-ZF 进行了修改

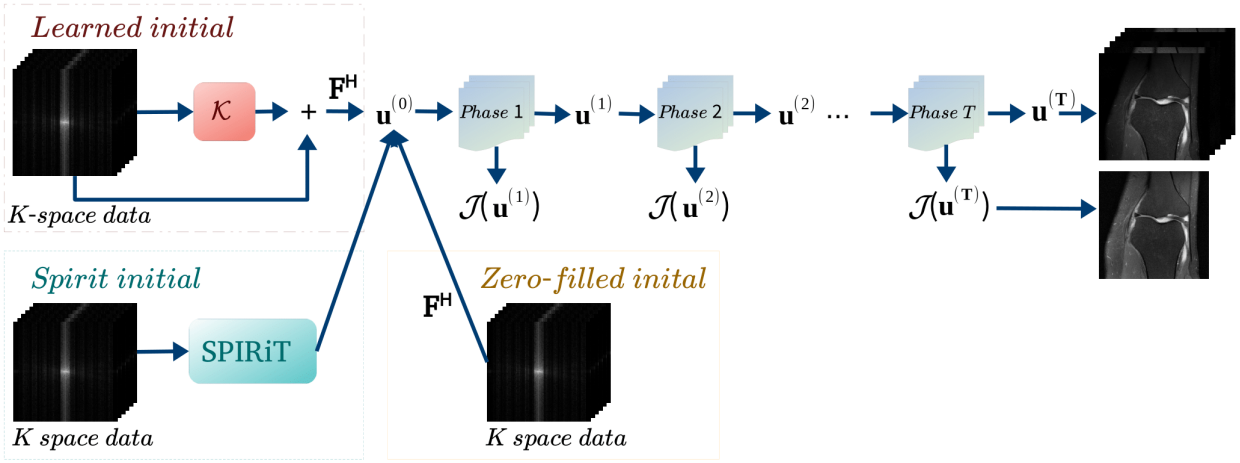

图3-6. 采用三种不同初始重建方法（包括零填充初始值、SPIRiT 重建初始值和学习得到的初始值）时各阶段的框架范式。

~ ~多线圈图像以生成单体图像。其他算子 $\mathcal{G}$、$\mathcal{G}$ 和 $\mathcal{J}$ 仍用于执行复数卷积。具体而言，RSS 的输出：$(\sum^{c}_{i=1}|\mathbf{u}_i(t)|^2)^{1/2} \in \mathbb{R}^{mn}$是一个单通道实值图像，它同时输入非线性算子 $\mathcal{G}$的实部和虚部，因此输出被拆分为复数值。我们将采用组合算子 $\mathcal{J}$ 的网络称为 pMRI-Net-ZF/pMRI-ℂNet-ZF，将采用 RSS 的网络称为 pMRI-ℂNet-RSS；其余设置均与之前相同。

这两类网络的重建图像和平均评估结果分别如图3-5和表3-6所示。pMRI-Net-ZF 和 pMRI-ℂNet-ZF 的表现优于 pMRI-ℂNet-RSS，峰值信噪比分别平均提高了 1.36 分贝和 1.63 分贝。这表明，与应用平方和开根号相比，学习得到的组合算子 $\mathcal{J}$ 表现更佳。

3.4.4.2 初始化

重建网络输入$\mathbf{u}(0)$的选择也会影响最终的重建质量。我们没有直接将部分 k 空间数据$\mathbf{f}$ 用作网络输入，而是采用了三种不同的输入$\mathbf{u}(0)$：(i) 零填充重建$\mathbf{F}^H\mathbf{f}$；(ii) SPIRiT [119]重建；(iii) $\mathbf{F}^H(\mathbf{f} + \mathcal{K}(\mathbf{f}))$，可以将$\mathbf{f} + \mathcal{K}(\mathbf{f})$视为插值伪全 k 空间。

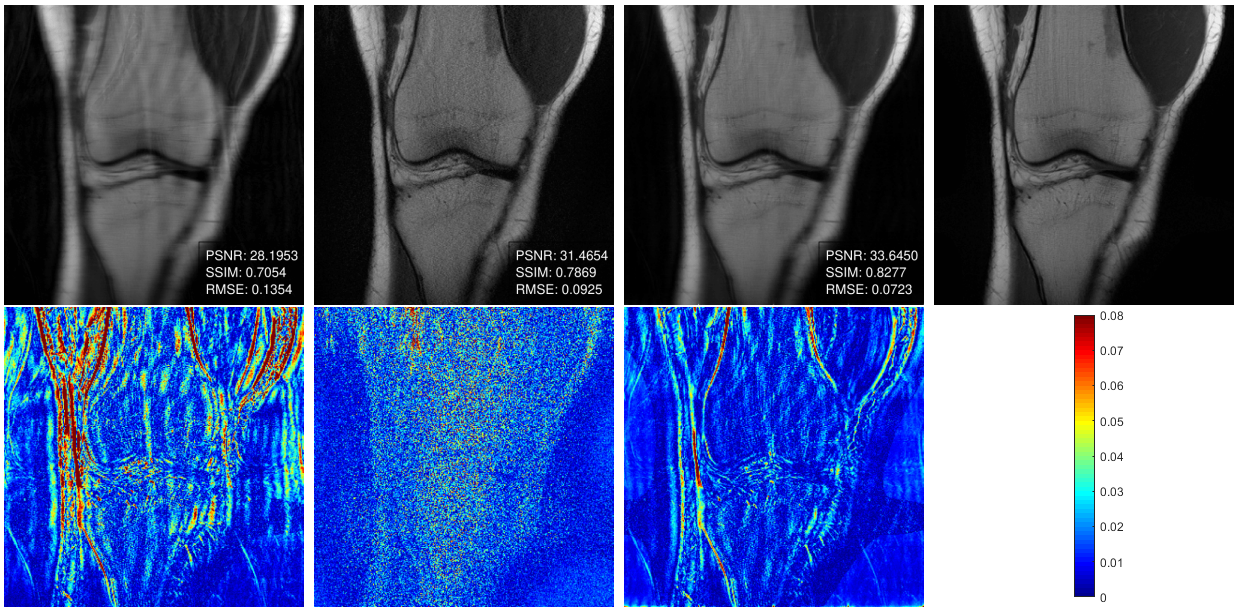

图 3-7。第一行（从左到右）显示初始值 $\bar{\mathbf{u}}(0)$ 的平方和开根号。

从表 3-6可以看出，与零填充初始值相比，SPIRiT 初始值略有改进，而学习得到的初始值相比另外两种初始化方式取得了最高的重建质量。图 3-7 展示了 pMRI-Net-ZF/pMRI-$\mathbb{C}$Net-ZF、pMRI-Net-SP/pMRI-$\mathbb{C}$Net-SP、pMRI-Net-K/pMRI-$\mathbb{C}$Net-K 的三种初始值，以及冠状位 PD 膝关节图像上的参考图像。第二行显示了它们的逐点误差图和色标。。我们观察到，与零填充相比，SPIRiT 和学习得到的初始值都具有更高的空间分辨率。SPIRiT 是一种经典的 k 空间方法；与零填充和学习得到的初始值相比，该初始值在减少混叠伪影和保留边缘方面表现更好，但 SPIRiT 会引入更多噪声。

与零填充相比，得益于k空间网络 $\mathcal{K}$，学习得到的初始值保留了主要组织的结构特征。与SPIRiT初始值相比，学习得到的初始值减少了图像空间中的分辨率噪声。基于学习的初始值在零填充和SPIRiT之间取得了均衡的性能，避免了这两种初始值各自的缺点。

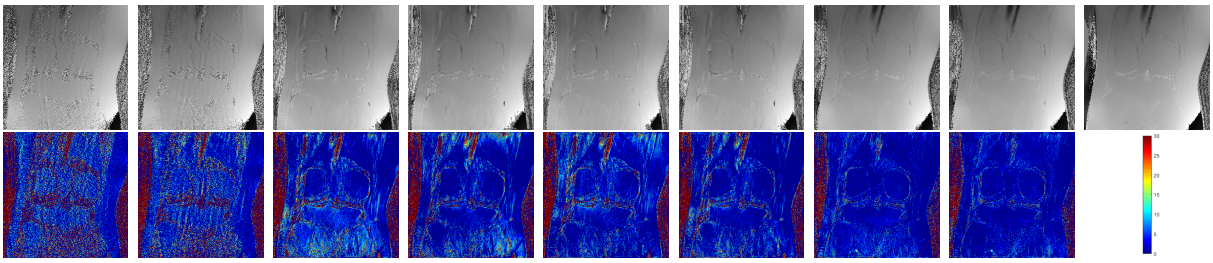

图 3-8。不同方法重建的线圈图像中一个线圈的相位信息与所提出的方法和参考图像的比较。

3.4.4.3 复数卷积

在本实验中，我们比较了-Net和-ℂNet。表3-6 中的定量评估表明，与所提出的实值网络相比，所提出的复值网络在峰值信噪比/结构相似性指数/均方根误差方面表现优异。表3-4 表明，复数卷积能够提取每个$\mathbf{u}_i$ 中的特征，并获得最低的均方根误差。图3-8展示了重建线圈图像中一个通道的相位图像。第一行显示了GRAPPA、SPIRiT、De-AliasingNet、DeepcomplexMRI、CDF-Net、Adaptive-CS-Net、pMRI-ℂNet-K（采用损失函数3–26）、所提方法以及参考图像中一个线圈的相位信息。第二行显示了相应的逐点误差图和色标，最大误差为 30°。这些结果表明，复值网络在更新多线圈图像和保留每个通道的相位信息方面发挥着重要作用。

3.4.4.4 与所提出的图像域重建和域混合重建的比较

pMRI-ℂNet-K 与所提出网络的主要区别在于，pMRI-ℂNet-K 仅在图像域中进行迭代（3–25b），而所提方法则进行迭代（3–9），从而实现域混合重建。式（3–25b）中的去噪网络为$\tilde{\mathcal{J}} \circ \tilde{\mathcal{G}} \circ \mathcal{S}_{\alpha_t}$（$\mathcal{G} \circ \mathcal{J}$），而在所提出的网络中，我们使用 $\mathcal{M} + \mathbf{F}^H \mathcal{K} \mathbf{F} + \mathbf{F}^H \mathcal{K} \mathbf{F} \mathcal{M}$，其中 $\mathcal{M} = \mathcal{J} \circ \mathcal{G} \circ \mathcal{G} \circ \mathcal{J}$是图像域网络。所提方法中省去了软阈值算子$\mathcal{S}_{\alpha_t}$，因为我们发现，加入软阈值并不会使结果产生明显差异。

表3-6 和表3-4 中的平均重建结果表明，域混合方法的性能更佳。与 pMRI-ℂNet-K 相比，我们所提方法的峰值信噪比提高了 0.598 分贝，结构相似性指数提高了 0.003，均方根误差降低了 0.014；质子密度数据集的结果见表3-6 。

3.5 结论

本文为免校准并行磁共振成像重建模型提出了一种离散时间最优控制框架。我们采用卷积组合算子合并多线圈图像的通道，并对通道合并图像应用参数化正则化函数，以重建逐通道多线圈图像。所提方法的灵感来自近端梯度算法。近端点由两个去噪网络学习，分别在图像域和k空间域中运行。我们将重建网络构建为结构化离散时间最优控制系统，由此得到参数训练的最优控制形式，为并行磁共振成像重建提供了一种可解释且性能优异的深度架构。我们从拉格朗日乘子法的角度设计网络训练。我们证明了拉格朗日乘子法等价于反向传播，并可采用基于随机梯度下降的算法，获得满足最优控制问题必要条件的解。重建结果具有较高的感知质量，展现了所提出的pMRI-Net的优越性能。

第4章：用于多样化数据集磁共振成像重建的基于优化的元学习模型

本研究旨在元学习框架下开发一种具有泛化能力的磁共振成像重建模型。元学习中的标准基准测试因面对多样的任务分布而受到挑战。所提出的网络在变分模型中学习正则化函数，并利用异构数据集重建具有不同欠采样率或采样模式的磁共振图像，这些欠采样率或采样模式可能出现在训练数据中，也可能未曾出现。我们提出一种由可学习优化算法（LOA）诱导的展开网络，用于求解磁共振成像重建的非凸非光滑变分模型。在该模型中，可学习正则化函数包含一个任务不变的公共特征编码器，以及由浅层网络表示的任务特定学习器。为训练网络，我们将训练数据分为训练集和验证集两部分，并引入双层优化算法。下层优化在任务特定学习器参数固定的情况下，利用训练数据集训练特征编码器的任务不变参数；上层则在验证数据集上优化任务特定学习器的参数。与在已见压缩感知比率上通过传统监督学习训练的网络相比，峰值信噪比平均提高了1.5分贝。我们测试了元训练后在未见任务上的快速适应结果；与直接在未见压缩感知比率上训练的传统学习流程相比，平均峰值信噪比提高了1.22分贝，同时训练时间缩短了一半。对于未见过的采样模式，与传统学习相比，平均峰值信噪比提高了1.87分贝。我们提出了一种元学习框架，包括基础网络架构、正则化设计以及基于双层优化的训练。该网络继承了LOA的收敛性质和变分模型的可解释性。通过专门设计的正则化和基于双层优化的训练算法，提升了泛化能力。

4.1 引言

深度学习方法在各种图像重建问题中都展现出良好的性能。然而，深度学习模型通常针对特定任务进行训练，并要求训练样本遵循相应的分布。具体而言，源域/训练样本与目标域/测试样本需要来自同一分布 [102, 148, 11, 9]。实际中，这些数据集通常采集自不同来源，具有显著的异质性，因此在现实应用中，样本可能遵循相关但不同的分布 [53, 43]。因此，利用此类数据集稳健且高效地训练深度神经网络，在理论上十分重要，在基于深度学习的方法的应用中也具有实际意义。

元学习提供了一种独特范式，可实现稳健且高效的神经网络训练 [133, 54, 102, 156, 207, 9]。元学习也称为学会学习，旨在通过涵盖相关任务分布的学习回合经验，快速学习未见任务。在多任务场景中，给定一组任务，元学习已被证明是提取任务无关知识、提升该组中新任务学习性能的有效工具[174, 78]。我们利用元学习的这一特性来训练网络：磁共振成像训练数据通过不同的欠采样模式（例如笛卡尔掩模、径向掩模、泊松掩模）、不同的欠采样率以及不同的扫描参数设置采集，从而产生不同的对比度（例如 T1加权、T2加权、质子密度（PD）和液体衰减反转恢复）。这些数据具有很强的异质性，可视为来自不同任务。因此，我们的目标是在元学习框架下开发一种稳健且具有良好泛化能力的图像重建方法，利用此类大规模异质数据进行磁共振成像重建。

我们的方法可概述如下。首先，我们提出一个变分模型，将图像重建表述为一个非凸非光滑优化问题。在

我们的变分模型中，正则化项被参数化为一个结构化深度神经网络，其网络参数可在训练过程中学习。随后，我们提出一种具有严格收敛保证的可学习优化算法（LOA），用于求解该优化问题。接着，我们严格遵循此 LOA 构建一个深度重建网络；具体而言，该 LOA诱导网络的每个阶段都恰好对应 LOA 的一次迭代。这一方法的灵感源自 [32], ，但本研究开发的 LOA 在计算上比 [32]中的方法更高效：本研究中的保护变量仅在必要时更新，因此在保留收敛保证的同时，能够显著降低计算成本。

其次，为了提高网络鲁棒性并缓解过拟合问题，我们将正则化的网络参数明确划分为两部分：任务不变部分和任务特定部分。前者从训练数据中不同任务的图像提取共同的先验信息，并学习正则化的任务不变分量。另一方面，后者由另一个网络得出，该网络针对不同任务挖掘正则化的适当任务特定分量（也称为元知识）。该网络的超参数也在训练过程中学习。此外，我们将可用数据划分为两个集合：训练集和验证集。随后，我们引入双层优化模型来学习网络参数。具体而言，下层问题（也称为内层问题）在固定任务特定部分的条件下，利用训练数据集寻找正则化的任务不变部分；上层（外层）问题则利用验证数据集，寻找正则化参数中最优的任务特定部分。这种方法极大地提高了所学习正则化的鲁棒性，这意味着经过训练的 LOA 诱导的深度重建网络能够很好地泛化到未见任务。

正如第 4.5 节的数值实验所示，在 MRI图像重建中，所提出的框架利用具有各种欠采样轨迹和欠采样率的多样化数据集，显著提升了图像质量。原因在于

有效的正则化能够整合来自多样化数据集中各种训练样本的共同特征和先验信息，但这些信息需要与特定任务（即欠采样轨迹和欠采样率）所获得的数据保真项进行适当权衡。我们的贡献可概括如下：

1. 受 LOA 启发的网络架构——我们的网络架构严格遵循一种具有保证收敛性的 LOA。因此，与现有展开网络相比，该网络更具可解释性、参数效率更高且更加稳定。

2. 正则化的自适应设计——我们的自适应正则化器由任务不变部分和任务特定部分组成，两者均可通过数据进行适当训练。

3. 提升网络鲁棒性和泛化能力——我们将网络参数训练过程构建为双层优化，以训练数据作为下层输入，以验证数据作为上层输入，从而提升该过程的鲁棒性。这种方法也增强了训练后网络的泛化能力，使其能够快速适应采用全新、未见过的采样轨迹进行的图像重建，并生成高质量的重建图像。

本文其余部分安排如下。在第4.2节中，我们将讨论基于优化的元学习以及用于磁共振成像重建的深度展开网络的相关工作。我们将在第4.3节提出元学习模型和神经网络，并在第4.4节介绍实现细节。第4.5 节给出所提方法的数值结果。第4.6节总结全文。

4.2 相关工作

近年来，元学习方法结合不同技术，在各个领域展现出了颇具前景的成果 [78]。元学习技术可分为三类 [209, 101, 85]：基于度量的方法 [95, 184, 163]、基于模型的方法 [131, 150, 143, 64], 以及基于优化的方法 [54, 149, 110]。基于优化的方法通常被表述为双层优化问题，并且对于更广泛的任务分布表现出相对更强的泛化能力。我们主要关注

本文中的基于优化的元学习。关于元学习更全面的文献综述和发展情况，读者可参阅近期的综述 [78, 85]。

基于优化的元学习方法已广泛应用于各种深度学习领域 [54, 8, 149, 110, 134, 55, 63, 56, 211]。这些元学习方法中的网络训练问题通常被表述为领导变量和跟随变量的双层优化。双层优化的约束条件是：对于每个固定的领导变量，跟随变量都是下层问题的最优解；而双层优化的最终目标是在此约束下，找到使上层目标函数最小化的最优领导变量（通常也包括相应的最优跟随变量）。在许多现有的基于优化的元学习应用中，例如模型无关元学习（MAML） [54], ，以及大量采用类似策略、旨在提升泛化能力的 MAML 后续研究[101, 156, 56, 63, 134, 186, 208, 210]，下层问题通过一步或几步梯度下降来近似求解。深度双层学习 [89] 旨在获得比仅在单个任务上训练时更好的泛化能力，并能够良好地泛化到另一个任务。该模型通过优化正则化损失函数，从训练集中寻找网络参数，并确定超参数，使网络在验证数据集上表现良好。

当未见任务与元训练任务处于不一致的领域时，如 [29],所示，元学习器的泛化表现会受到影响。这一现象部分源于对已见元训练任务的元过拟合，在 [210]中被认定为记忆化问题。 [210] 提出了一种基于信息论的元正则化器，通过调节任务适应过程中的信息依赖来解决记忆化问题。

MetaReg [9] 将整个网络解耦为特征网络和任务网络，其中元正则化项仅应用于任务网络。首先，他们使用元训练集更新任务网络的参数，以获得

域对齐的任务网络，然后在元测试集上更新元正则化项的参数，以学习跨域泛化。与 MetaReg 不同，特征评论网络 [109] 利用元正则化项来构建域不变特征提取网络。元正则化被设计为一个特征评论网络，以提取的特征作为输入。通过最小化新的元正则化损失来更新特征提取网络的参数。特征评论网络中的辅助参数则通过最大化相较于非元方法的性能提升来学习。为有效评估元学习器的性能，研究人员在更贴近现实的设置下开发了若干新基准测试 [152, 177, 213] ，这些基准测试可在多样化的视觉域中良好运行。正如 [177],所述，在多模态或异构数据集中实现对未见任务的泛化，仍是现有元学习方法面临的一项挑战。

上述方法针对学习了正则化函数的分类网络进行领域泛化，以实现跨域泛化。所提出的方法旨在解决逆问题，并构建了一种自适应正则化器，不仅能够学习任务之间的通用参数，还能够学习任务感知参数。所设计的自适应正则化器有助于提升深度模型的泛化能力，使训练良好的模型能够在已见任务和未见任务的异构数据集上均表现出良好性能。

4.3 所提出的方法

4.3.1 预备知识

我们首先介绍压缩感知磁共振成像（CS-MRI）的背景、图像重建问题，以及用于解决图像重建问题的学习型优化算法。CS-MRI通过对k空间（傅里叶空间）测量值进行欠采样来加速MRI数据采集。欠采样k空间测量值与图像之间的关系如下式所示 [67]:

$$\mathbf{y} = \mathbf{P}\mathcal{F}\mathbf{x} + \mathbf{n}, \tag{4–1}$$

其中，$\mathbf{y} \in \mathbb{C}^p$ 表示k空间中的测量值，共有p 个采样数据点；$\mathbf{x} \in \mathbb{C}^{N\times 1}$ 是待重建的磁共振图像，包含N 个像素；$\mathcal{F} \in \mathbb{C}^{N\times N}$ 是二维离散傅里叶变换（DFT）矩阵；$\mathbf{P} \in \mathbb{R}^{p\times N}$（$p < N$）是表示k空间采样轨迹的二值矩阵。$\mathbf{n}$ 是k空间中的采集噪声。

求解 $\mathbf{x}$ 由（含噪声的）欠采样数据 $\mathbf{y}$ 根据（4–1）得到的结果是一个病态问题。缓解病态问题的有效策略之一，是在重建中引入先验信息。变分方法是实现这一目标的最有效方法之一。该方法的一般框架是最小化由数据保真项和正则化项组成的目标函数，如下所示：

$$\bar{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{P}\mathcal{F}\mathbf{x} - \mathbf{y}\|^2 + R(\mathbf{x}), \tag{4–2}$$

其中，第一项是数据保真度，用于确保重建的$\mathbf{x}$ 与测量数据$\mathbf{y}$一致；第二项$R(\mathbf{x})$是正则化项，用于为待重建图像引入先验知识。在传统的变分方法中，$R(\mathbf{x})$是手工设计的函数，例如全变分（TV） [19]。优化技术的进步使得人们能够借助理论论证，开发出更有效的算法来求解变分模型。然而，手工设计的正则化项可能过于简单，无法捕捉细微细节，也难以达到临床诊断质量。

近年来，我们见证了深度学习在解决各种逆问题方面取得的巨大成功，但这些基于深度学习的方法的可解释性和泛化能力仍是主要关注点。作为对通用黑箱型深度神经网络（DNN）的改进，受优化算法启发的若干类可学习优化算法（LOA）构建了神经网络，称为展开网络。这类网络将迭代算法展开为多阶段网络，并在实证研究中展现出良好的求解精度和效率 [116, 111, 157, 124, 223, 162, 24, 6]。然而，其中许多网络只是对迭代算法似是而非的模仿，因此缺乏变分模型这一理论基础，也没有任何收敛保证。

鉴于深度学习已取得的巨大成功以及目前可用的大量训练数据，我们可以将正则化项参数化为一个从训练样本中学习的深度卷积神经网络（CNN）。LOA诱导的重建方法已成功应用于压缩感知磁共振成像（CS-MRI），通过可学习正则化项解决逆问题：

$$\arg\min_{x} \frac{1}{2}\|\mathbf{P}\mathcal{F}\mathbf{x} - \mathbf{y}\|^2 + R(\mathbf{x};\Theta). \tag{4–3}$$

其中，$R(\mathbf{x};\Theta)$ 是以参数Θ构成的深度网络形式的正则化项。根据 $R(\mathbf{x};\Theta)$ 的具体参数化形式以及用于展开的优化方案，近年来人们提出了多种展开网络。例如，变分网络（VN）[72]被提出用于展开梯度下降算法，并将正则化项参数化为线性滤波器与非线性卷积神经网络的组合。MoDL [5] 提出了一种递归网络中的权重共享策略，通过展开共轭梯度法来学习正则化参数。ADMM-Net[172]模拟了著名的交替方向乘子法；其正则化项设计为以分段线性函数替代 L_1-范数。ISTA-Net [218] 将正则化项视为卷积网络的 L_1-范数。该网络通过迭代展开多个阶段，每个阶段都模拟迭代收缩阈值算法（ISTA）的一次迭代 [112, 10]。然而，这些网络只是表面上模拟了相应的优化方案，与原始优化方法或变分模型缺乏直接联系，也不具备任何收敛保证。在本研究中，我们首先针对（4–3）提出一种可学习优化算法（LOA），并对其进行全面的收敛性分析；随后，严格遵循 LOA 的迭代方案，构建由 LOA 诱导的网络。

4.3.2 LOA诱导重建网络

本节首先介绍一种学习型优化算法（LOA），用于求解（4–3），其中正则化网络参数 Θ 是固定的。由于 Θ 在（4–3）中是固定的，我们

暂时在下面的 LOA 推导中省略这一点，并为简化记号，将 $R(\mathbf{x};\Theta)$ 写作 $R(\mathbf{x})$。

在本研究中，为了将稀疏性与学习特征结合起来，我们将函数参数化为 $R(\mathbf{x},\,\Theta) = \kappa \cdot r(\mathbf{x},\,\Theta)$，其中 $\kappa > 0$ 是一个权重参数，需要根据具体任务（例如噪声水平、欠采样率等）适当选取；而 r是一个正则化项，其参数化形式为神经网络的复合，并可适用于广泛的成像应用。具体而言，我们将 r 参数化为$\ell_{2,1}$ 范数与可学习特征提取算子 $\mathbf{g}(\mathbf{x})$的复合。也就是说，我们将式 r （4–11）中的

$$r(\mathbf{x}) := \|\mathbf{g}(\mathbf{x})\|_{2,1} = \sum_{j=1}^{m} \|\mathbf{g}_j(\mathbf{x})\|. \tag{4–4}$$

这里，“:=”表示“定义为”。$\mathbf{g}(\cdot) = (\mathbf{g}_1(\cdot), \ldots, \mathbf{g}_m(\cdot))$, $\mathbf{g}_j(\cdot) = \mathbf{g}_j(\cdot;\,\theta)$ 对于$j = 1, \cdots, m$，被参数化为卷积神经网络（CNN）；而 θ 是$r(\cdot;\,\theta)$ 中已学习并固定的网络参数，如上文所述。我们暂且也将κ视为已学习并固定的 θ，并将在下一小节讨论如何学习二者。我们在$\mathbf{g}$中使用式（4–20）所定义的平滑激活函数，使$\mathbf{g}$ 成为平滑但非凸的函数。由于$\|\cdot\|_{2,1}$是非光滑的，r 因此是一个非光滑非凸函数。

由于式（4–11）中的最小化问题是非凸且非光滑的，因此我们需要推导一种高效的 LOA 来求解该问题。这里，我们首先考虑对任意固定的 $\ell_{2,1}$ 范数进行平滑，其中 $\mathbf{g}(\mathbf{x})$ 是

$$r_\varepsilon(\mathbf{x}) = \sum\nolimits_{j=1}^{m} \sqrt{\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2} - \varepsilon. \tag{4–5}$$

我们将$R_\varepsilon = \kappa \cdot r_\varepsilon$记为。此处推导的 LOA 受到近端梯度下降算法的启发，并通过迭代以下步骤来求解平滑后的问题：

$$\mathbf{z}_{t+1} = \mathbf{x}_t - \alpha_t \nabla f(\mathbf{x}_t) \tag{4–6a}$$

$$\mathbf{x}_{t+1} = \mathrm{prox}_{\alpha_t R_{\varepsilon_t}}(\mathbf{z}_{t+1}), \tag{4–6b}$$

其中，ε_t 表示特定迭代 t 时的平滑参数 ε，近端算子定义为 $\text{prox}_{\alpha g}(b) := \arg\min_x (1/2)\|x - b\|^2 + \alpha g(x)$，见式（4–6b）。从式（4–5）可以很快看出，随着 ε 逐渐减小，$R_\varepsilon \to R$，因此我们有意在算法2的第16行将 $\varepsilon_t \to$ 推至 0。接着，可以直接证明，对所有 x 和 $\varepsilon > 0$，均有 $R_\varepsilon(x) \le R(x) \le R_\varepsilon(x) + \varepsilon$。根据算法2，第16行会自动减小 ε，迭代序列将收敛到原始非光滑非凸问题（4–11）的解——附录 4.7 中的收敛性分析对此作了精确说明。

由于 R_{ε_t} 是一个涉及深度神经网络的复值函数，因此其近端算子没有闭式解，无法在子问题 in(4–6b) 中轻易计算。为克服这一困难，我们考虑用以下方式近似R_{ε_t}：

$$\hat{R}_{\varepsilon_t}(\mathbf{x}) = R_{\varepsilon_t}(\mathbf{z}_{t+1}) + \langle \nabla R_{\varepsilon_t}(\mathbf{z}_{t+1}), \mathbf{x} - \mathbf{z}_{t+1} \rangle + \frac{1}{2\beta_t}\|\mathbf{x} - \mathbf{z}_{t+1}\|^2.$$

Then，我们更新 $\mathbf{u}_{t+1} = \text{prox}_{\alpha_t \hat{} R_{\varepsilon t}}(\mathbf{z}_{t+1})$ 以替代 (4–6b)；因此，我们得到

$$\mathbf{u}_{t+1} = \mathbf{z}_{t+1} - \tau_t \nabla R_{\varepsilon_t}(\mathbf{z}_{t+1}), \text{ where } \tau_t = \frac{\alpha_t \beta_t}{\alpha_t + \beta_t}. \tag{4–8}$$

如果 LOA 2 中的第5行条件满足，我们将迭代 $\mathbf{x}_{t+1} = \mathbf{u}_{t+1}$。为保证算法的收敛，我们引入 ϕ_{ε_t}（其中$\phi_{\varepsilon_t} := f + R_{\varepsilon_t}$）在 x 处的标准梯度下降：

$$\mathbf{v}_{t+1} = \arg\min_x \langle \nabla f(\mathbf{x}_t), \mathbf{x} - \mathbf{x}_t \rangle + \langle \nabla R_\varepsilon(\mathbf{x}_t), \mathbf{x} - \mathbf{x}_t \rangle + \frac{1}{2\alpha_t}\|\mathbf{x} - \mathbf{x}_t\|^2, \tag{4–9}$$

由此可得

$$\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t \nabla \phi_{\varepsilon_t}(\mathbf{x}_t), \tag{4–10}$$

作为收敛的保障措施。如果 LOA 中第9行的条件2 不成立，我们将通过乘以 α_t 来缩小 ρ，重复有限次后即可满足第9行的条件。具体而言，如果$\mathbf{x}_{t+1} = \mathbf{u}_{t+1}$ 满足 $\phi_{\varepsilon_t}(\mathbf{u}_{t+1}) \le \phi_{\varepsilon_t}(\mathbf{v}_{t+1})$，则设为$\mathbf{x}_{t+1} = \mathbf{v}_{t+1}$；否则，设为$\mathbf{x}_{t+1} = \mathbf{v}_{t+1}$。然后重复此过程。

我们的算法总结于算法2。含未知参数的先验项本身具有精确的残差更新，从而改进了学习和训练过程 [74]。第5行的条件检查旨在确保其处于能量下降方向。一旦第5行的条件不成立，过程就转到$\mathbf{v}_{t+1}$，而第12行的线搜索保证可以在有限步内找到合适的步长，使函数值下降。从第3行到第14行，我们考虑在 ϕ_{ε_t} 固定的情况下，求解一个最小化问题。第15行根据约简准则更新 ε_t 的值。关于该机制的详细分析和深入的收敛论证见附录4.7。相应的展开网络严格遵循算法2 ，因此具有相同的收敛性质。与 LDA [32], 每次迭代都计算两个候选值$\mathbf{u}_{t+1}$和$\mathbf{v}_{t+1}$ ，再选择函数值较小者不同，我们提出了上述第5行的准则来更新$\mathbf{x}_{t+1}$，这有望节省计算候选值$\mathbf{v}_{t+1}$ 所需的额外时间，并有望减少两个候选值之间频繁切换的情况。此外，本工作提出的平滑方法比对偶空间中的平滑方法 [32] 更为直接，同时仍能保证可证明收敛性，如定理4.1所示。

所提出的 LOA诱导网络是一种多阶段网络，其架构严格遵循所提出的 LOA（算法2），每个阶段都对应算法的一次迭代。具体而言，我们构建了一个深度网络，记为F_Θ，它针对用户指定的 T 次迭代严格遵循算法 2。其中， Θ表示 F_Θ中的可学习参数集合，包括正则化网络参数 θ、权重 κ以及算法 2的其他算法参数。因此，对于任意输入的欠采样$\mathbf{k}$空间测量值$\mathbf{y}$，$F_\Theta(\mathbf{y})$ 会执行 LOA（算法 2）共 T 次迭代，并生成该最小化问题（4–11）的近似解：

$$F_\Theta(\mathbf{y}) \approx \arg\min_{\mathbf{x}} \left\{ \phi_\Theta(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}, \mathbf{y}) + R(\mathbf{x}; \Theta) \right\}. \tag{4–11}$$

算法2： 具有可证明收敛性的算法展开方法

输入： 初始值 $\mathbf{x}_0$、 $0 < \rho$、 $\gamma < 1$ 和 ε_0、a、$\sigma > 0$。最大总阶段数 T 或容差 $\epsilon_{\text{tol}} > 0$。

对于 $t = 0, 1, 2, \ldots, T-1$ **执行**

  $\mathbf{z}_{t+1} = \mathbf{x}_t - \alpha_t \nabla f(\mathbf{x}_t)$

  $\mathbf{u}_{t+1} = \mathbf{z}_{t+1} - \tau_t \nabla R_{\varepsilon_t}(\mathbf{z}_{t+1})$,

  若 $\|\nabla \phi_{\varepsilon_t}(\mathbf{x}_t)\| \le a\|\mathbf{u}_{t+1} - \mathbf{x}_t\|$ 且 $\phi_{\varepsilon_t}(\mathbf{u}_{t+1}) - \phi_{\varepsilon_t}(\mathbf{x}_t) \le -\frac{1}{a}\|\mathbf{u}_{t+1} - \mathbf{x}_t\|^2$ **则**

    设置 $\mathbf{x}_{t+1} = \mathbf{u}_{t+1}$,

  else

    $\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t \nabla \phi_{\varepsilon_t}(\mathbf{x}_t)$,

    if $\phi_{\varepsilon_t}(\mathbf{v}_{t+1}) - \phi_{\varepsilon_t}(\mathbf{x}_t) \le -\frac{1}{a}\|\mathbf{v}_{t+1} - \mathbf{x}_t\|^2$ 成立 **则**

      设置 $\mathbf{x}_{t+1} = \mathbf{v}_{t+1}$,

    else

      更新 $\alpha_t \leftarrow \rho \alpha_t$，然后 **前往** 8,

    结束 if

  结束 if

  如果 $\|\nabla \phi_{\varepsilon_t}(\mathbf{x}_{t+1})\| < \sigma\gamma\varepsilon_t$，设置 $\varepsilon_{t+1} = \gamma\varepsilon_t$；**否则**，设置 $\varepsilon_{t+1} = \varepsilon_t$。

  如果 $\sigma\varepsilon_t < \epsilon_{\text{tol}}$，终止。

结束循环

输出： $\mathbf{x}_t$。

这里我们使用“≈”，因为F_Θ 仅执行有限步优化算法来近似求解。值得强调的是，只要f（可能是非凸的）连续可微且具有 Lipschitz连续梯度，这种方法就可以轻松应用于更广泛的图像重建问题。在下一小节中，我们将提出一种基于元学习的方法，用于对网络参数 Θ进行鲁棒训练。

4.3.3 网络训练的双层优化算法

在本节中，我们研究 LOA诱导网络 F_Θ 的参数训练问题。具体而言，我们开发了一种双层优化算法，从多样化数据集中训练网络参数 Θ ，以提高网络鲁棒性和泛化能力。

请注意，LOA诱导网络F_Θ 严格遵循算法2，该算法旨在求解包含可学习正则化项R（x； Θ）的变分模型（4–11）。如第4.3.2节所示，我们设计了R（x； Θ） $= \kappa \cdot r$（**x**； Θ），其中r经过学习以捕捉

所有不同任务中底层共同特征的内在属性。为应对不同训练/验证数据集中的巨大差异，我们引入任务特定参数 ω_i，以近似第i个任务适当的 κ。具体而言，对于第i个任务，权重 κ设为 $\sigma(\omega_i) \in (0, 1)$，其中 $\sigma(\cdot)$ 是 Sigmoid 函数。因此，$\kappa = \sigma(\omega_i)$ 根据第i个任务特定的采样率或模式，确定r的适当权重。参数 ω_i 将在下文的超参数调优过程中与 Θ一同优化。

假设我们获得了用于训练和验证的 $\mathcal{M}$ 数据对 $\{(\mathrm{y}_m, \mathrm{x}^*_m)\}_{m=1}^{\mathcal{M}}$，其中 y_m 是观测值，在我们的设定中即为部分 k 空间数据，而 x^*_m 是对应的真实图像。随后将这些数据对采样为$\mathcal{N}$ 任务 $\{\mathcal{D}_{\tau_i}\}_{i=1}^{\mathcal{N}}$，其中每个 $\mathcal{D}_{\tau_i}$ 表示特定任务 τ_i中的数据对集合。在每个任务 τ_i中，我们进一步将数据划分为任务特定的训练集$\mathcal{D}^{tr}_{\tau_i}$和验证集 $\mathcal{D}^{val}_{\tau_i}$。我们的基础网络架构完全遵循上一节提出的 LOA（算法2），其包含可学习参数 θ 以及第 i 个任务的任务特定参数 ω_i。更具体地，对于任务 τ_i 中索引为 j 的一个数据样本，记为 $(\mathrm{y}^{(i)}\mathrm{j}\,, \mathrm{x}*^{(i)}\mathrm{j}\,)$，我们为任务 τ_i提出如下算法展开网络

$$F_{\theta,\omega_i}(\mathbf{y}_j^{(i)}) \approx \arg\min_{\mathrm{x}} f(\mathbf{x}, \mathbf{y}_j^{(i)}) + \sigma(\omega_i) r(\mathbf{x}; \theta), \tag{4–12}$$

其中 θ 表示具有任务不变表示的不同任务间可学习的通用参数，而 ω_i 是任务 τ_i的任务特定参数。权重 $\sigma(\omega_i)$表示与特定任务r 相关联的权重 τ_i。在我们所提出的网络中，Θ 是任务 (θ, ω_i) 的集合，其中任务$i = 1, \cdots \mathcal{N}$。我们将 ω 记作集合 $\{\omega_i\}_{i=1}^{\mathcal{N}}$。该网络的详细架构见第4.3.2节。我们定义任务特定损失函数

$$\ell_{\tau_i}(\theta, \omega_i; \mathcal{D}_{\tau_i}) := \sum_{j=1}^{|\mathcal{D}_{\tau_i}|} \ell\big(F_{\theta,\omega_i}(\mathbf{y}_j^{(i)}), \mathbf{x}*_j^{(i)}\big), \tag{4–13}$$

其中 $|\mathcal{D}_{\tau_i}|$ 表示 $\mathcal{D}_{\tau_i}$ 的基数，并且

$$\ell\big(F_{\theta,\omega_i}(\mathbf{y}_j^{(i)}), \mathbf{x}*_j^{(i)}\big) := \frac{1}{2}\|F_{\theta,\omega_i}(\mathbf{y}_j^{(i)}) - \mathbf{x}*_j^{(i)}\|^2. \tag{4–14}$$

为防止所提出的模型对训练数据过拟合，我们提出了一种新颖的学习框架，将网络训练表述为一个双层优化问题，以学习 ω 和 θ 在（4–12）中的形式：

$$\min_{\omega=\{\omega_i: i\in[N]\}} \quad \sum_{i=1}^{\mathcal{N}} \ell_{\tau_i}(\theta(\omega), \omega_i; \mathcal{D}_{\tau_i}^{val}) \tag{4–15a}$$

$$\text{s.t.} \qquad \theta(\omega) = \arg\min_{\theta} \sum_{i=1}^{\mathcal{N}} \ell_{\tau_i}(\theta, \omega_i; \mathcal{D}_{\tau_i}^{tr}). \tag{4–15b}$$

在（4–15）中，下层优化在固定任务特定参数 ω_i 的条件下，利用训练数据集学习特征编码器的任务不变参数 θ；上层则调整任务特定参数 $\{\omega_i\}$，使任务不变参数θ 也能在验证数据集上稳健地执行。为简单起见，我们省略求和，并重新定义 $\mathcal{L}(\theta, \omega; \mathcal{D}) := \sum_{i=1}^{\mathcal{N}} \ell_{\tau_i}(\theta, \omega; \mathcal{D})$，然后将（4–15）简要改写为

$$\min_{\omega} \mathcal{L}(\theta(\omega), \omega; \mathcal{D}^{val}) \qquad \text{s.t.} \quad \theta(\omega) = \arg\min_{\theta} \mathcal{L}(\theta, \omega; \mathcal{D}^{tr}). \tag{4–16}$$

接着，根据 [125]，我们将（4–16）松弛为单层约束优化，其中下层问题由其一阶必要条件取代

$$\min_{\omega} \mathcal{L}(\theta(\omega), \omega; \mathcal{D}^{val}) \qquad \text{s.t.} \quad \nabla_\theta \mathcal{L}(\theta, \omega; \mathcal{D}^{tr}) = 0. \tag{4–17}$$

通过惩罚项，可进一步将其近似为无约束问题，形式如下：

$$\min_{\theta,\omega} \big\{\widetilde{\mathcal{L}}(\theta, \omega; \mathcal{D}^{tr}, \mathcal{D}^{val}) := \mathcal{L}(\theta, \omega; \mathcal{D}^{val}) + \frac{\lambda}{2}\|\nabla_\theta \mathcal{L}(\theta, \omega; \mathcal{D}^{tr})\|^2\big\}. \tag{4–18}$$

我们在每次迭代中采用小批量数据集上损失函数的随机梯度。在我们的模型中，每个批次需要包含多个任务的数据对；因此，我们在训练时提出了跨任务小批量。在每次训练

迭代中，我们会在每个任务 τ_i上随机抽取训练数据对 $\mathcal{B}_{\tau_i}^{tr} = \{(\mathrm{y}^{(i)}\mathrm{j},\ \mathrm{x}*^{(i)}\mathrm{j}) \in \mathcal{D}_{\tau_i}^{tr}\}_{=1}^{\mathcal{J}^{tr}}\mathrm{j}$和验证数据对 $\mathcal{B}_{\tau_i}^{val} = \{(\mathrm{y}^{(i)}\mathrm{j},\ \mathrm{x}*^{(i)}\mathrm{j}) \in \mathcal{D}_{\tau_i}^{val}\}^{\mathcal{J}}\mathrm{val}\,\mathrm{j}_{=1}$。随后，每次训练迭代中使用的整体训练和验证小批量 $\mathcal{B}^{tr}$ 和 $\mathcal{B}^{val}$ 由从所有任务中抽取的数据对组成；即 $\mathcal{B}^{tr} = \bigcup^{\mathcal{N}}\mathrm{i}_{=1}\{\mathcal{B}_{\tau_i}^{tr}\}$ 和 $\mathcal{B}^{val} = \bigcup^{\mathcal{N}}\mathrm{i}_{=1}\{\mathcal{B}_{\tau_i}^{val}\}$。因此，在每次迭代中，我们分别使用 $\mathcal{N}\cdot\mathcal{J}^{tr}$ 和 $\mathcal{N}\cdot\mathcal{J}^{val}$ 个数据对进行训练和验证。为求解最小化问题（4–17），我们采用算法3中总结的随机小批量交替方向法，该方法改编自 [125]。

如 [125], 所分析，这种惩罚型方法具有线性时间复杂度，无需计算低层的海森矩阵；由于每次训练迭代只需存储中间的 θ、ω，因此空间复杂度为常数，适合求解大规模双层优化问题。算法3利用一阶必要条件，将双层优化问题松弛为单层约束优化问题。该问题与原问题并不等价，但求解起来容易得多且效率更高。在算法3的内循环（第5–9行）中，我们继续针对 θ 最小化转换后的单层优化函数（4–18），迭代K步，然后交替更新ω 一次，直到第5行中容差为 δ 的条件不再满足。第5行中条件的基本思想源于一阶必要条件：我们希望~将 $\mathcal{L}$ 的梯度推向0。此外，在外循环（第2–11行）的第11行，我们减小容差 δ。结合第5行和第11行，可确保每次~内循环终止时，$\mathcal{L}$ 关于 θ 和 ω 的梯度都越来越接近0。参数 δ_{tol} 用于控制整个算法的准确率；当 δ 足够小时（即 $\delta \leq \delta_{tol}$），外循环便会终止。此外，λ 是（4–18）中第二个约束项的权重；开始时，我们将 λ 设得较小，以实现快速初始收敛，随后逐渐增大其值，以强化该约束。

算法3： 用于求解的随机小批量交替方向罚方法
问题（4–17）

输入 $\mathcal{D}_{\tau_i}^{tr}$，$\mathcal{D}_{\tau_i}^{val}$，$\delta_{tol} > 0$。
初始化 θ，ω，δ，$\lambda > 0$ 和 $\nu_\delta \in (0,1)$，$\nu_\lambda > 1$。
当 $\delta > \delta_{tol}$ **时执行**
  采样跨任务训练批次 $\mathcal{B}^{tr} = \bigcup_{i=1}^{\mathcal{N}} \{(\mathbf{y}^{(i)}{}_j, \mathbf{x}*^{(i)}{}_j) \in \mathcal{D}_{\tau_i}^{tr}\}_{j=1:\mathcal{J}^{tr}}$
  采样跨任务验证批次 $\mathcal{B}^{val} = \bigcup_{i=1}^{\mathcal{N}} \{(\mathbf{y}^{(i)}{}_j, \mathbf{x}*^{(i)}{}_j) \in \mathcal{D}_{\tau_i}^{val}\}_{j=1:\mathcal{J}^{val}}$
  当 $\|\nabla_\theta$ **时** $\widetilde{\mathcal{L}}(\theta, \omega;\ \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 + \|\nabla_\omega \widetilde{\mathcal{L}}(\theta, \omega;\ \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 > \delta$ **执行**
    对于 $k = 1, 2, \ldots, K$（内循环）**执行**
      $\theta \leftarrow \theta - \rho_\theta^k \nabla_\theta \widetilde{\mathcal{L}}(\theta, \omega;\ \mathcal{B}^{tr}, \mathcal{B}^{val})$
    结束循环
    $\omega \leftarrow \omega - \rho_\omega \nabla_\omega \widetilde{\mathcal{L}}(\theta, \omega;\ \mathcal{B}^{tr}, \mathcal{B}^{val})$
  结束 while 循环
  更新 $\delta \leftarrow \nu_\delta \delta$，$\lambda \leftarrow \nu_\lambda \lambda$
结束 while 循环
输出： θ，ω。

4.4 实现

4.4.1 特征提取算子

我们将特征提取算子 $\mathbf{g}$ 设为一个标准的l层卷积神经网络，其分量级非线性激活函数为 φ 且不含偏置，如下所示：

$$\mathbf{g}(x) = \mathbf{w}_l * \varphi \cdots\ \varphi(\mathbf{w}_3 * \varphi(\mathbf{w}_2 * \varphi(\mathbf{w}_1 * x))), \tag{4–19}$$

其中 $\{\mathbf{w}_q\}_{q=1}^{l}$ 表示由 d 个具有相同空间卷积核尺寸的卷积核组成的卷积权重，而 $*$表示卷积操作。此处，φ 构造为下文定义的平滑修正线性单元：

$$\varphi(x) = \begin{cases} 0, & \text{if } x \le -\delta, \\ \frac{1}{4\delta}x^2 + \frac{1}{2}x + \frac{\delta}{4}, & \text{if } -\delta < x < \delta, \\ x, & \text{if } x \ge \delta, \end{cases} \tag{4–20}$$

其中，在我们的实验中，前缀参数 δ 设为 0.001。特征提取算子的默认配置如下：特征提取

算子 $\mathbf{g}$ 由 $l = 3$ 个卷积层组成，所有卷积均使用 4 个卷积核，空间尺寸为 3×3。

4.4.2 设置

由于我们的方法引入了算法展开网络，因此算法迭代与神经网络阶段（或模块）之间存在一一对应关系。前向传播的每个阶段都可以视为一次算法迭代，这促使我们模仿优化算法的迭代过程，并采用阶梯式训练策略 [32]。在第一阶段，我们先使用一个阶段开始训练网络参数；待损失函数收敛后，再增加阶段（每次增加一个阶段），并继续训练过程。重复这一过程，直到增加更多模块后损失函数不再下降时停止。每次使用基于随机梯度下降的优化器 Adam 将损失函数最小化 100 个训练轮次/迭代次数，Adam [93] 的参数为 $\beta_1 = 0.9$，$\beta_2 = 0.999$，初始学习率设为 10^{-3}，小批量大小设为 8。使用 Xavier 初始化器 [61] 初始化所有卷积的权重。初始平滑参数 ε_0 设为 0.001，随后与其他网络参数一同学习。展开网络的输入 $\mathbf{x}_0$ 通过零填充策略[13]获得。深度展开网络使用 Tensorflow 工具箱 [1] 以 Python 编程语言实现。

4.5 数值实验

4.5.1 数据集

为验证所提方法的性能，我们使用的数据来自多模态脑肿瘤分割挑战赛 2018 [128], ，其中训练数据集包含从 285 名患者采集的四种模态（T1、$T1_c$、T2 和液体衰减反转恢复），验证数据集包含 66 名患者的图像，每个体积的大小为$240 \times 240 \times 155$。每种模态包含两种类型的胶质瘤：75 个低级别胶质瘤（LGG）体积和 210 个高级别胶质瘤（HGG）体积。我们的实现使用了两种模态（T1加权成像和 T2图像）的 HGG 磁共振成像，并从

训练数据集的每种模态中选择 30 名患者来训练我们的网络。在验证数据集中，我们随机选取 15 名患者作为验证数据，并从训练数据集中选取 6 名患者作为测试数据；这些患者与训练集和验证集中的患者互不重叠。我们将二维图像裁剪为 160×180，取每个体积中心的 10 个相邻切片，最终得到 300 张训练图像、150 张验证图像和 60 张测试图像。这里提到的数据量针对单个任务；但由于我们采用了多任务训练，因此每个数据集中的图像总数应乘以任务数。对于每个二维切片，我们将空间强度除以最大像素值进行归一化。

4.5.2 实验设置

所有实验均通过 TensorFlow 在一台 Windows 工作站上完成，该工作站配备 3.3GHz Intel Core i9 CPU 和一块显存为 11 GB 的 Nvidia GTX-1080Ti GPU [2]。所提出的网络中的参数采用 Xavier 初始化 [60]。我们使用四个与四种不同压缩感知比率（10%、20%、30% 和 40%）协同关联的任务对元学习网络进行训练，并使用具有这四种比率对应的相同掩模的测试数据集，对训练充分的模型进行测试。每种压缩感知比率使用 300 个训练数据，训练数据集中共计 1200 幅图像。T1加权成像和 T2加权成像的磁共振重建结果分别见表 4-1 和 4-2。对应的重建图像见图4-2和 4-1。我们还使用径向掩模对未见过的 15%、25% 和 35% 比率的未见任务，以及使用比率为 10%、20%、30% 和 40% 的随机笛卡尔掩模，对训练充分的元学习模型进行了测试。对于未见任务，我们针对不同采样率的不同掩模，分别重新训练任务特定参数，同时固定任务不变参数 θ。在这些实验中，我们只需学习 ω_i，以适应径向掩模下的三种未见压缩感知比率和笛卡尔掩模下的四种常规压缩感知率。实验训练所用数据和迭代次数较少，采用 100 幅磁共振图像并训练 50 轮。例如，要重建压缩感知比率为 15% 的磁共振图像

对于径向掩模，我们固定参数 θ，并使用 100 个原始数据点训练 50 轮，重新训练任务特定参数ω，然后使用更新后的 ω，在测试数据集上进行测试；测试数据集中的原始测量数据由径向掩模以 15% 的压缩感知径向采样率采集。径向掩模对应的结果见表 4-3 和 4-4，T1加权成像和 T2加权成像的结果分别见图 4-3 和 4-4。笛卡尔掩模对应的结果列于表 4-5，重建图像见图 4-5。

我们将所提出的元学习方法与传统监督学习进行了比较。传统监督学习每次使用一个任务进行训练，并且只学习任务不变参数 θ，不学习任务特定参数 ω_i。传统学习的前向网络展开了算法2，包含 11 个阶段，与元学习相同。我们合并了训练集和验证集，得到 450 张图像，用于传统监督学习的训练。训练批次大小设为 25，总共训练 2000 轮；而元学习训练了 100 轮，批量大小为 8。元学习和传统学习均使用相同的测试集，以评估这两种方法的性能。

我们针对七种不同的压缩感知比率（10%、20%、30%、40%、15%、25% 和 35%），在两种随机欠采样模式——径向采样掩模和笛卡尔采样掩模——下，从两个方面比较了元学习与传统网络。元学习网络和传统学习网络的参数均通过 Adam优化器进行训练 [93],，并且二者都学习了前向展开的任务不变参数 θ。传统方法的网络训练在卷积数量、卷积神经网络卷积核的深度和大小、阶段数以及参数初始化器等方面，采用了与元学习网络相同的网络配置。这两种方法在训练过程中的主要区别在于，元学习通过利用从算法 ω_i 3中学习到的任务特定参数，针对多任务执行学习；同时，通过展开算法 2的前馈网络学习任务之间的共同特征，而传统学习则解决任务特定的

问题：通过算法 2简单地展开前向网络，其中训练和测试均在同一任务上进行。为考察元学习的泛化能力，我们在两种具有不同轨迹的采样掩模下，针对来自不同分布的磁共振图像，对训练充分的元学习模型进行了测试。传统学习的训练和测试采用相同的压缩感知比率；也就是说，如果传统方法以 10% 的压缩感知比率进行训练，那么测试时也会使用压缩感知比率为 10% 的数据集，依此类推。

由于磁共振图像以复数值表示，我们为每个卷积神经网络应用了复数卷积 [191] ；也就是说，每个卷积核都由实部和虚部组成。g 中使用了三次卷积，每次卷积都包含四个滤波器，空间卷积核尺寸为 3×3。在算法2中，若将终止条件设为 $\epsilon_{\text{tol}} = 1 \times 10^{-3}$，则总共可实现 11 个阶段；除步长外，各阶段的参数均共享。对于算法2中的超参数，我们选择了初始可学习步长 $\alpha_0 = 0.01$、$\tau_0 = 0.01$、$\varepsilon_0 = 0.001$，并将预设值设为 $\mathrm{a} = 10^5$、$\sigma = 10^3$、$\rho = 0.9$ 和 $\gamma = 0.9$。选择这些参数的原则基于算法的收敛性和计算的有效性。参数 $0 < \rho < 1$ 是线搜索过程中步长的缩减率，用于保证收敛。第 15 步的参数 $0 < \gamma < 1$ 是 ε的缩减率。在算法 1 中，从第 2 步到第 14 步，平滑级别 ε保持固定。当平滑函数的梯度足够小时，我们将 ε 按比例因子 γ 缩小，以寻找原始非光滑非凸问题的近似聚点。我们选择较大的 a，以便有更多迭代次数 k 使 u_{k+1}满足第 5 步中的条件，从而减少需要计算 v_{k+1}的迭代次数。此外，计算 u_{k+1} 的方案符合残差学习架构，该架构已被证明能有效降低训练误差。

在算法 3中，我们将 $\nu_\delta = 0.95$，并将参数 δ 初始化为 $\delta_0 = 1 \times 10^{-3}$，并在值 $\delta_{tol} = 4$处停止。35×10^{-6}，总共进行了 100 个训练轮次。为了

训练传统方法时，我们设置了 2000 个训练轮次，并采用与训练元学习方法相同的阶段数、卷积数和卷积核尺寸。初始值 λ设置为 1×10^{-5} 和 $\nu_\lambda = 1.001$。

我们使用峰值信噪比（PSNR） [77], 结构相似性指数（SSIM） [193], 和归一化均方误差（NMSE） [141]这三项指标评估测试数据集上的重建结果。以下公式用于计算重建图像$\mathbf{x}$ 与真实值$\mathbf{x}^*$之间的 PSNR、SSIM 和 NMSE。PSNR 可由均方误差（MSE）推导得出，其中

$$PSNR(\mathbf{x}, \mathbf{x}^*) = 20 \log_{10} \left(\frac{\max(|\mathbf{x}^*|)}{\sqrt{MSE(\mathbf{x}, \mathbf{x}^*)}} \right), \tag{4–21}$$

其中N 是真实值中像素的总数，MSE 定义为$MSE(\mathbf{x}, \mathbf{x}^*) = {}^1{}_N \|\mathbf{x}^* - \mathbf{x}\|^2$。

–

$$SSIM(\mathbf{x}, \mathbf{x}^*) = \frac{(2\mu_x\mu_{x^*} + C_1)(2\sigma_{xx^*} + C_2)}{(\mu_x^2 + \mu_{x^*}^2 + C_1)(\sigma_x^2 + \sigma_{x^*}^2 + C_2)}, \tag{4–22}$$

其中 μ_x, μ_{x^*}表示局部均值， σ_x, σ_{x^*}表示标准差， σ_{xx^*}表示$\mathbf{x}$ 与$\mathbf{x}^*$之间的协方差， $C_1 = (k_1 L)^2, C_2 = (k_2 L)^2$ 是两个用于避免零分母的常数，且$k_1 = 0.01, k_2 = 0.03$。L 是磁共振图像的最大像素值。

$$NMSE(\mathbf{x}, \mathbf{x}^*) = \frac{\|\mathbf{x} - \mathbf{x}^*\|_2^2}{\|\mathbf{x}\|_2^2}, \tag{4–23}$$

其中，NMSE 用于衡量平均相对误差。有关上述三项指标的详细信息，请参阅 [77, 193, 141]。

4.5.3 径向掩模中不同压缩感知比率的实验结果

本节评估训练充分的元学习和传统学习方法的性能。表4-1、4-2 和4-5 报告了平均数值性能及其标准差，以及相关的去尺度化任务特定元知识的定量结果 σ（ω_i）。通过对径向掩模进行的实验，我们观察到，对于T1加权成像

脑图像，与传统方法相比，元学习在全部四种压缩感知比率下的平均峰值信噪比值提高了1.54分贝；对于T2脑图像，元学习的平均峰值信噪比提高了1.46分贝。由于元学习的一般设置旨在利用每个独立任务提供的信息，而每个任务都对应一个独立的采样掩码，并且这些掩码可能包含互补的采样点，因此每个任务的重建性能都能从其他任务中受益。较小的压缩感知比率会抑制重建精度，这是由于原始测量中的欠采样轨迹较为稀疏；而即使在训练数据量不足的情况下，元学习也展现出解决这一问题的良好潜力。

一般而言，监督学习要求训练数据具有相同或相似的分布；异质数据呈现出不同的特征结构变化，这会妨碍卷积神经网络高效提取特征。在我们的实验中，以不同压缩感知率采样得到的原始测量数据呈现出不同程度的不完整性；这些欠采样测量并不属于相同的分布，但彼此相关。不同的采样掩码如图下方所示4-2 和4-3，而且它们可能互补地采样了某些点，也就是说，某些未被采样率为40% 的掩码采样的点，可能被其他掩码采集到。在我们的实验中，不同的采样掩码利用各自采样点所包含的信息，这意味着四个重建任务相互促进，从而实现了高效的性能。因此，这解释了为什么在采样率较高时，元学习仍优于传统学习。

元学习为监督学习拓展了一种新范式，其目的是快速学习多个任务。元学习只需针对可与四个不同任务共享的共同特征学习一次任务不变参数，并且每个 $\sigma(\omega_i)$ 都会根据“学会学习”的原则提供任务特定权重参数。在传统学习中，由于任务不变参数无法泛化到其他任务，网络参数需要针对四种不同的掩模分别训练四次，

这会耗费大量时间。从表4-1和 4-2中可以看出，较小的压缩感知比率需要较大的 $\sigma(\omega_i)$。事实上，在我们的模型（4–11）中，任务特定参数充当任务特定正则化项的加权约束，而表格表明，压缩感知比率越低，正则化所需的权重就越大。

图 4-2 和 4-1 分别展示了传统方法与元学习方法的定性比较，显示了 T1加权成像和 T2加权成像中同一层面的重建磁共振图像。我们用红框标出了高级别胶质瘤的放大细节。我们观察到，传统学习的结果较为模糊，并且会丢失清晰的边缘，在压缩感知比率较低时尤其如此。从逐点误差图中可以看出，与传统学习相比，元学习能够减少噪声，尤其是在一些细节丰富且复杂的区域。

对于前一项实验中使用的 T1加权图像，我们还使用三分之二的训练数据和验证数据测试了元学习的性能，并将其记为“元学习”。对于传统学习，网络也使用了前一项实验中三分之二的训练样本进行训练。测试数据集与之前相同。结果如表 4-1 所示，其中我们将减少数据量的实验记为“元学习*”和“传统方法*”。这些实验表明，减少训练数据量会降低测试数据的准确率；但元学习* 仍优于传统学习*，甚至优于传统学习。

为验证所提出的 LOA 2的重建性能，我们将所提出的传统学习方法与 ISTA-Net $^+$ [218]进行了比较。ISTA-Net 是一种用于磁共振成像重建的先进深度展开网络。我们在 T1加权图像上，使用与传统学习方法相同的训练数据集和测试数据集重新训练了 ISTA-Net$^+$ 。为进行公平比较，在训练过程中，我们采用了与传统学习方法相同数量的卷积核、每次卷积使用的相同卷积核维度，以及相同的阶段数。测试数值结果列于表 4-1 中，磁共振重建图像

显示于图 4-2中。在图 4-1、4-2中，图片（从上到下）依次展示了 T1脑图像重建结果、放大细节、带有色标的逐点误差，以及元学习和传统学习在四种不同压缩感知比率（从左到右分别为 10%、20%、30%、40%）下对应的**径向** 掩模。传统学习* 和元学习* 分别使用传统学习方法和元学习方法训练所用数据集的三分之二进行训练。我们可以观察到，展开算法 2 的传统学习方法在所有压缩感知比率下均优于 ISTA-Net+ 。从相应的逐点绝对误差可以看出，传统学习方法取得了低得多的误差和显著更好的重建质量。

4.5.4 不同采样模式下不同未见压缩感知比率的实验结果

在本节中，我们测试所提出模型对未见任务的泛化能力。我们固定训练良好的任务不变参数 θ ，仅训练 ω_i ，针对径向掩模的采样率 15%、25% 和 35%，以及笛卡尔掩模的采样率 10%、20%、30% 和 40%进行训练。在该实验中，每种压缩感知比率仅使用 100 个训练数据点，并总共训练 50 轮次。对于径向掩模下重建的 T1 和 T2 脑图像，平均评估值及标准差分别列于表4-3 和4-4 中；表4-5 展示了应用随机笛卡尔采样掩码后重建 T2 脑图像的定性表现。在表4-3和4-4中，元学习采用 10%、20%、30% 和 40% 的压缩感知比率进行训练，并在未见过的 15%、25% 和 35% 比率下进行测试。传统方法则采用相同的 15%、25% 和 35% 压缩感知比率进行常规训练和测试。在 T1 图像重建结果中，与传统方法相比，元学习在压缩感知比率为 15% 时的峰值信噪比提高了 1.6921 分贝，在 25% 时提高了 1.6608 分贝，在 35% 时提高了 0.5764 分贝；这表明，较低压缩感知比率下的重建质量提升幅度大于较高比率下的提升幅度。采用不同采样掩模的 T2 重建结果也呈现出类似趋势。

定性比较见图4-3、4-4和4-5 ，分别展示了径向掩模下采用未见压缩感知比率测试的 T1 和 T2 图像，以及笛卡尔掩模下采用常规压缩感知比率测试的 T2 图像。在图4-3、4-4中，元学习采用 10%、20%、30% 和 40% 的压缩感知比率进行训练，并在三个不同的未见压缩感知比率 15%、25% 和 35% 下进行测试（从左至右）。传统学习采用相同的 15%、25% 和 35% 压缩感知比率进行训练和测试。右上方图像为真实的全采样图像。右上方图像为真实的全采样图像。在4-5中，图片从上到下依次展示 T2 脑图像重建结果、放大细节、带色标的逐点误差，以及元学习和传统学习在四种不同压缩感知比率 10%、20%、30% 和 40% 下对应的**笛卡尔**掩模（从左至右）。右上方图像为真实的全采样图像。在采用径向掩模的实验中，元学习优于传统学习，尤其是在压缩感知比率为 15% 时——可以观察到，红框中的细节区域保留了边缘，且更接近真实图像，而传统方法的重建结果较为模糊，在一些复杂组织中丢失了细节。逐点误差图也表明，元学习具有抑制噪声的能力。

使用笛卡尔掩模进行训练比使用径向掩模更困难，尤其是对于传统学习而言；由于网络仅应用三次卷积，且每次卷积都使用四个卷积核，因此网络并不算深。表 4-5 表明，对于 T2 脑图像，元学习的平均性能比传统方法提高了约 1.87 分贝。这些结果进一步表明，元学习具有参数效率方面的优势；即使使用训练数据量较少的浅层网络，其性能也远优于传统学习。

上述数值实验结果表明，元学习能够快速适应新任务，并且面对由异质、多样化数据构成的广泛任务时，具有更稳健的泛化能力。元学习可以被视为一种

表4-1. 不同采样率的**径向** 掩模对应的 T1脑图像重建图像的定量评估。

压缩感知比率	方法	PSNR	SSIM	NMSE	$\sigma(\omega_i)$
10%	ISTA-Net+ [218]	21.2633 ± 1.0317	0.5487 ± 0.0440	0.1676 ± 0.0253	
	传统方法*	21.6947 ± 1.0264	0.5689 ± 0.0404	0.1595 ± 0.0240	
	传统方法	21.7570 ± 1.0677	0.5650 ± 0.0412	0.0259 ± 0.0082	
	元学习*	22.9633 ± 1.0969	0.5962 ± 0.0415	0.0194 ± 0.0065	0.9339
	元学习	23.2672 ± 1.1229	0.6101 ± 0.0436	0.0184 ± 0.0067	0.9218
20%	ISTA-Net+ [218]	26.2734 ± 1.0115	0.7068 ± 0.0364	0.0944 ± 0.0155	
	传统方法*	26.4639 ± 1.0233	0.7107 ± 0.0357	0.0924 ± 0.0154	
	传统方法	26.6202 ± 1.1662	0.7121 ± 0.0397	0.0910 ± 0.0169	
	元学习*	27.9381 ± 1.1121	0.7541 ± 0.0360	0.0063 ± 0.0023	0.8150
	元学习	28.2944 ± 1.2119	0.7640 ± 0.0377	0.0058 ± 0.0022	0.7756
30%	ISTA-Net+ [218]	28.8309 ± 1.3137	0.7492 ± 0.0407	0.0708 ± 0.0142	
	传统方法*	29.2923 ± 1.3194	0.7522 ± 0.0399	0.0671 ± 0.0136	
	传统方法	29.5034 ± 1.4446	0.7557 ± 0.0408	0.0657 ± 0.0143	
	元学习*	30.8691 ± 1.5897	0.8310 ± 0.0394	0.0033 ± 0.0015	0.6359
	元学习	31.1417 ± 1.5866	0.8363 ± 0.0385	0.0031 ± 0.0014	0.6501
40%	ISTA-Net+ [218]	30.7282 ± 1.5482	0.8008 ± 0.0428	0.0572 ± 0.0127	
	传统方法*	31.3761 ± 1.5892	0.8035 ± 0.0420	0.0532 ± 0.0121	
	传统方法	31.4672 ± 1.6390	0.8111 ± 0.0422	0.0029 ± 0.0014	
	元学习*	32.7330 ± 1.6386	0.8623 ± 0.0358	0.0022 ± 0.0010	0.6639
	元学习	32.8238 ± 1.7039	0.8659 ± 0.0370	0.0022 ± 0.0010	0.6447

表4-2。不同采样率的**径向**掩模对应的T2脑图像重建结果的定量评估。

压缩感知比率	方法	PSNR	SSIM	NMSE	$\sigma(\omega_i)$
10%	传统方法	23.0706 ± 1.2469	0.5963 ± 0.0349	0.2158 ± 0.0347	
	元学习	24.0842 ± 1.3863	0.6187 ± 0.0380	0.0112 ± 0.0117	0.9013
20%	传统方法	27.0437 ± 1.0613	0.6867 ± 0.0261	0.1364 ± 0.0213	
	元学习	28.9118 ± 1.0717	0.7843 ± 0.0240	0.0122 ± 0.0030	0.8742
30%	传统方法	29.5533 ± 1.0927	0.7565 ± 0.0265	0.1023 ± 0.0166	
	元学习	31.4096 ± 0.9814	0.8488 ± 0.0217	0.0069 ± 0.0019	0.8029
40%	传统方法	32.0153 ± 0.9402	0.8139 ± 0.0238	0.0770 ± 0.0128	
	元学习	33.1114 ± 1.0189	0.8802 ± 0.0210	0.0047 ± 0.0015	0.7151

一种高效的方法，通过利用从较简单任务中提取的特征来解决困难任务。

表4-3。与不同采样率的**径向**掩模相关的T1脑图像重建图像的定量评估。

压缩感知比率	方法	PSNR	SSIM	NMSE	$\sigma(\omega_i)$
15%	传统方法	24.6573 ± 1.0244	0.6339 ± 0.0382	0.1136 ± 0.0186	
	元学习	26.3494 ± 1.0102	0.7088 ± 0.0352	0.0090 ± 0.0030	0.9429
25%	传统方法	28.4156 ± 1.2361	0.7533 ± 0.0368	0.0741 ± 0.0141	
	元学习	30.0764 ± 1.4645	0.8135 ± 0.0380	0.0040 ± 0.0017	0.8482
35%	传统方法	31.5320 ± 1.5242	0.7923 ± 0.0420	0.0521 ± 0.0119	
	元学习	32.1084 ± 1.6481	0.8553 ± 0.0379	0.0025 ± 0.0011	0.6552

表4-4。不同采样率的**径向** 掩模对应的T2脑图像重建结果的定量评估。

压缩感知比率	方法	PSNR	SSIM	NMSE	$\sigma(\omega_i)$
15%	传统方法	24.8921 ± 1.2356	0.6259 ± 0.0285	0.1749 ± 0.0280	
	元学习	26.7031 ± 1.2553	0.7104 ± 0.0318	0.0205 ± 0.0052	0.9532
25%	传统方法	29.0545 ± 1.1980	0.7945 ± 0.0292	0.1083 ± 0.0173	
	元学习	30.0698 ± 0.9969	0.8164 ± 0.0235	0.0093 ± 0.0022	0.8595
35%	传统方法	31.5201 ± 1.0021	0.7978 ± 0.0236	0.0815 ± 0.0129	
	元学习	32.0683 ± 0.9204	0.8615 ± 0.0209	0.0059 ± 0.0014	0.7388

表4-5. 随机**笛卡尔**掩模在不同采样率下 T2脑图像重建的定量评估。

压缩感知比率	方法	PSNR	SSIM	NMSE	$\sigma(\omega_i)$
10%	传统方法	20.8867 ± 1.2999	0.5082 ± 0.0475	0.0796 ± 0.0242	
	元学习	22.0434 ± 1.3555	0.6279 ± 0.0444	0.0611 ± 0.0188	0.9361
20%	传统方法	22.7954 ± 1.2819	0.6057 ± 0.0412	0.0513 ± 0.0157	
	元学习	24.7162 ± 1.3919	0.6971 ± 0.0380	0.0329 ± 0.0101	0.8320
30%	传统方法	24.2170 ± 1.2396	0.6537 ± 0.0360	0.0371 ± 0.0117	
	元学习	26.4537 ± 1.3471	0.7353 ± 0.0340	0.0221 ± 0.0068	0.6771
40%	传统方法	25.3668 ± 1.3279	0.6991 ± 0.0288	0.1657 ± 0.0265	
	元学习	27.5367 ± 1.4107	0.7726 ± 0.0297	0.0171 ± 0.0050	0.6498

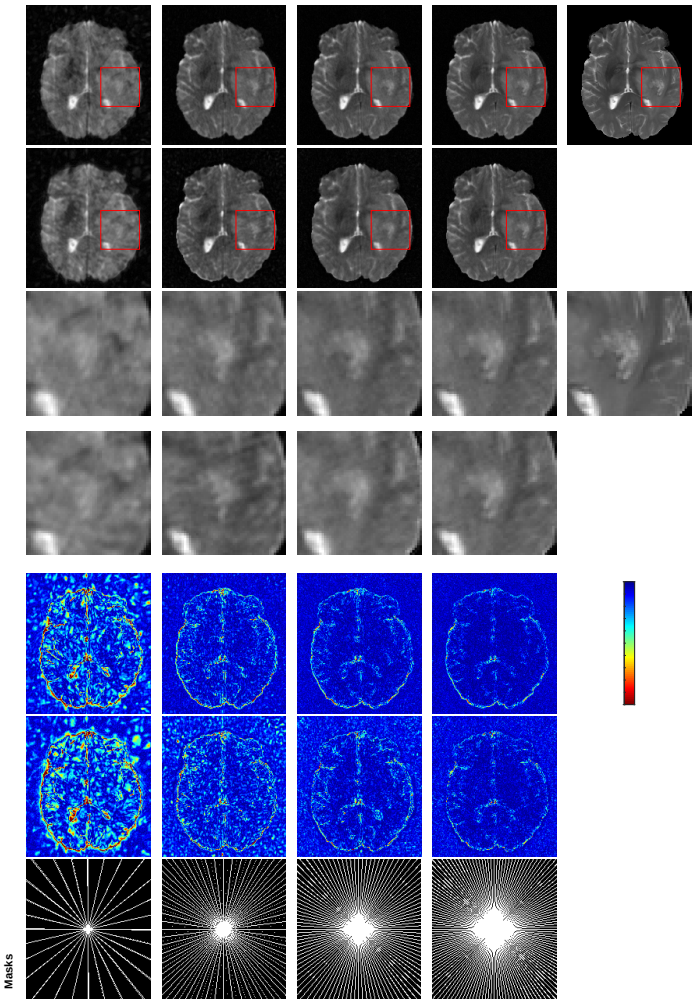

图4-1。从上到下：T2脑图像重建结果、放大细节、带色标的逐点误差以及相应的**径向**掩模。右上角的图像为真实值全采样图像。

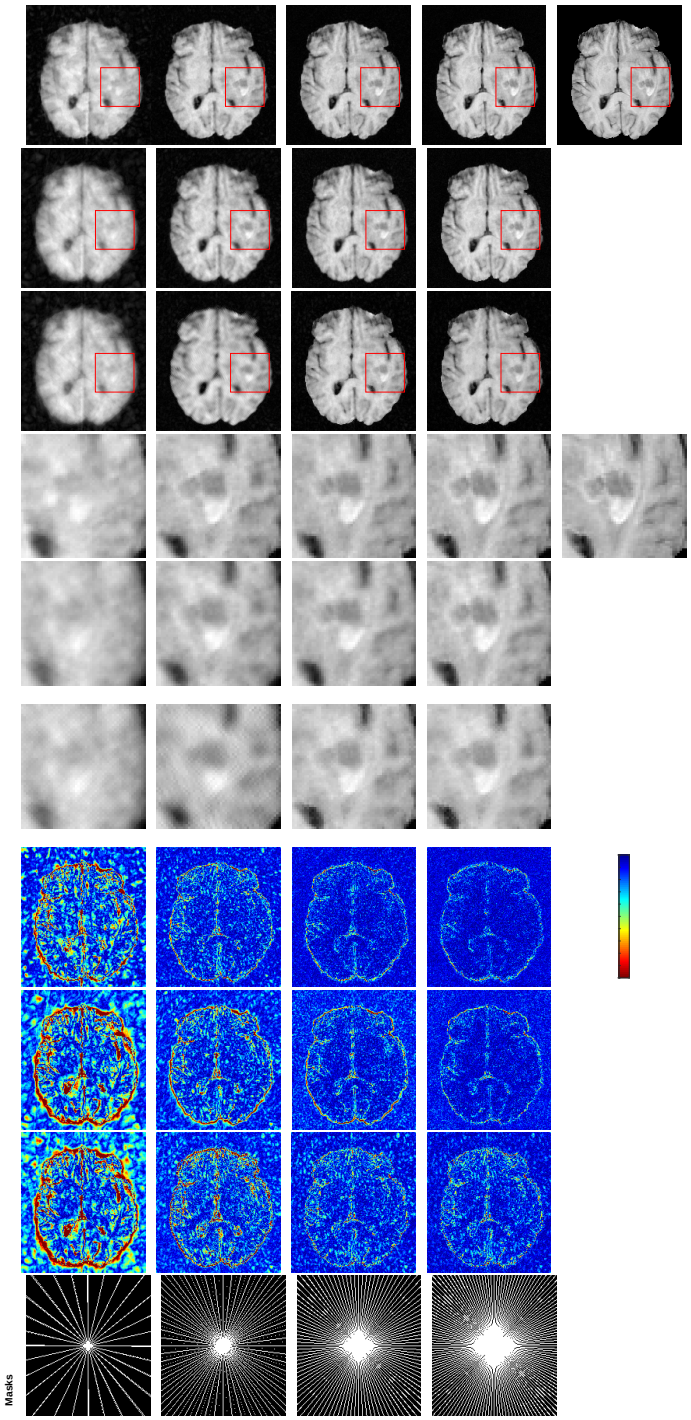

图4-2。从上到下：元学习、传统学习和 ISTA-Net+ 的重建结果、放大细节、带色标的逐点误差，以及相应的**径向**掩模。

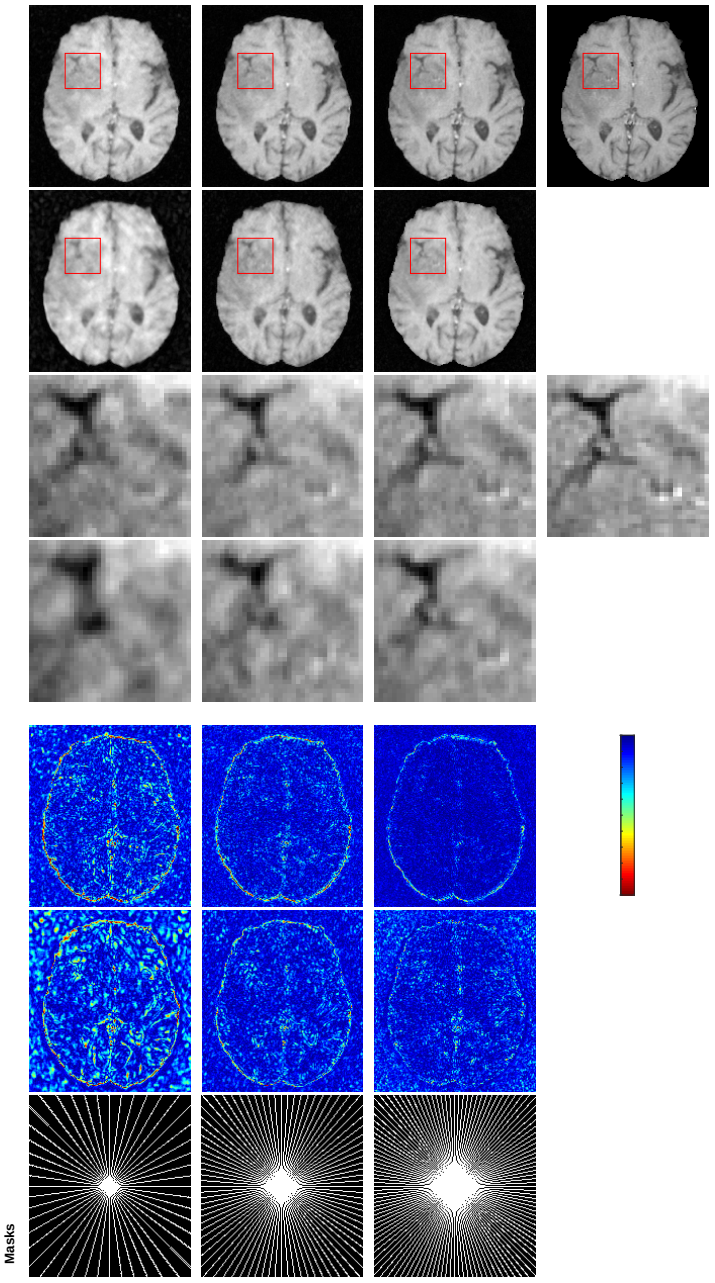

图4-3。从上到下依次展示 T1脑图像重建结果、放大细节、带有色标的逐点误差，以及相应的**径向**掩模。

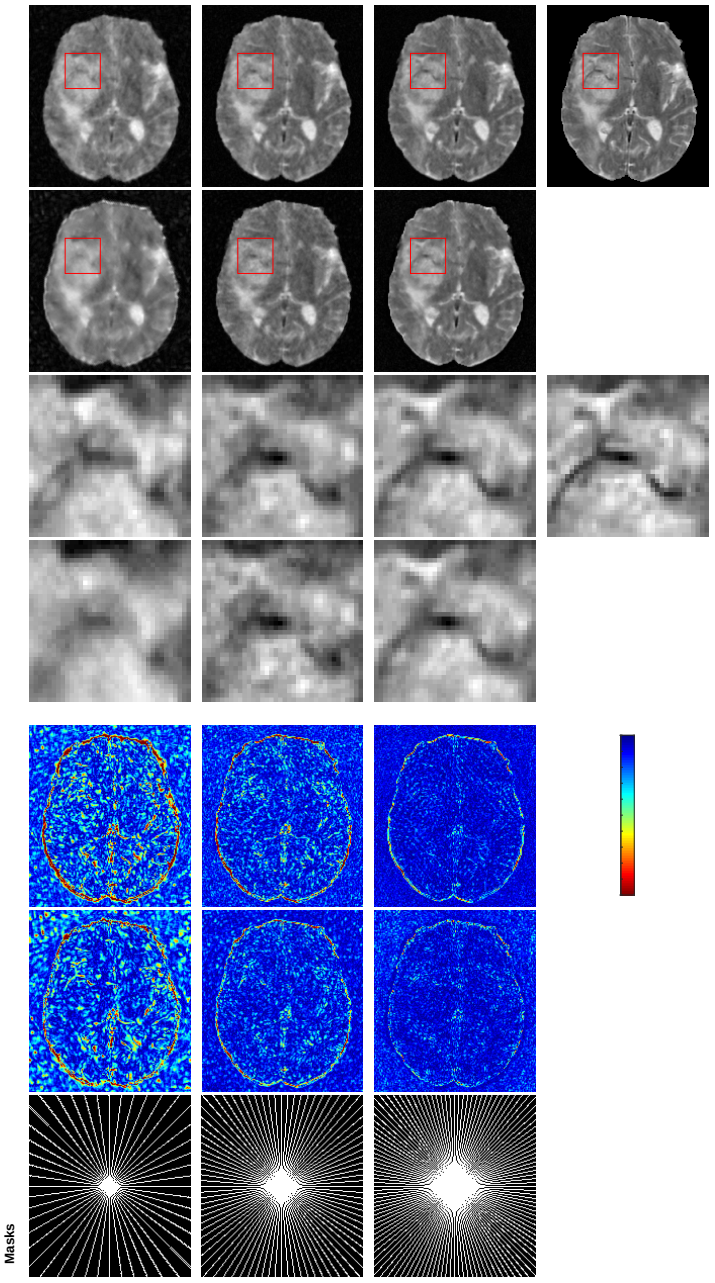

图 4-4. 图片（从上到下）展示了 T2脑图像重建结果、放大细节、带有色标的逐点误差以及相应的**径向** 掩模。

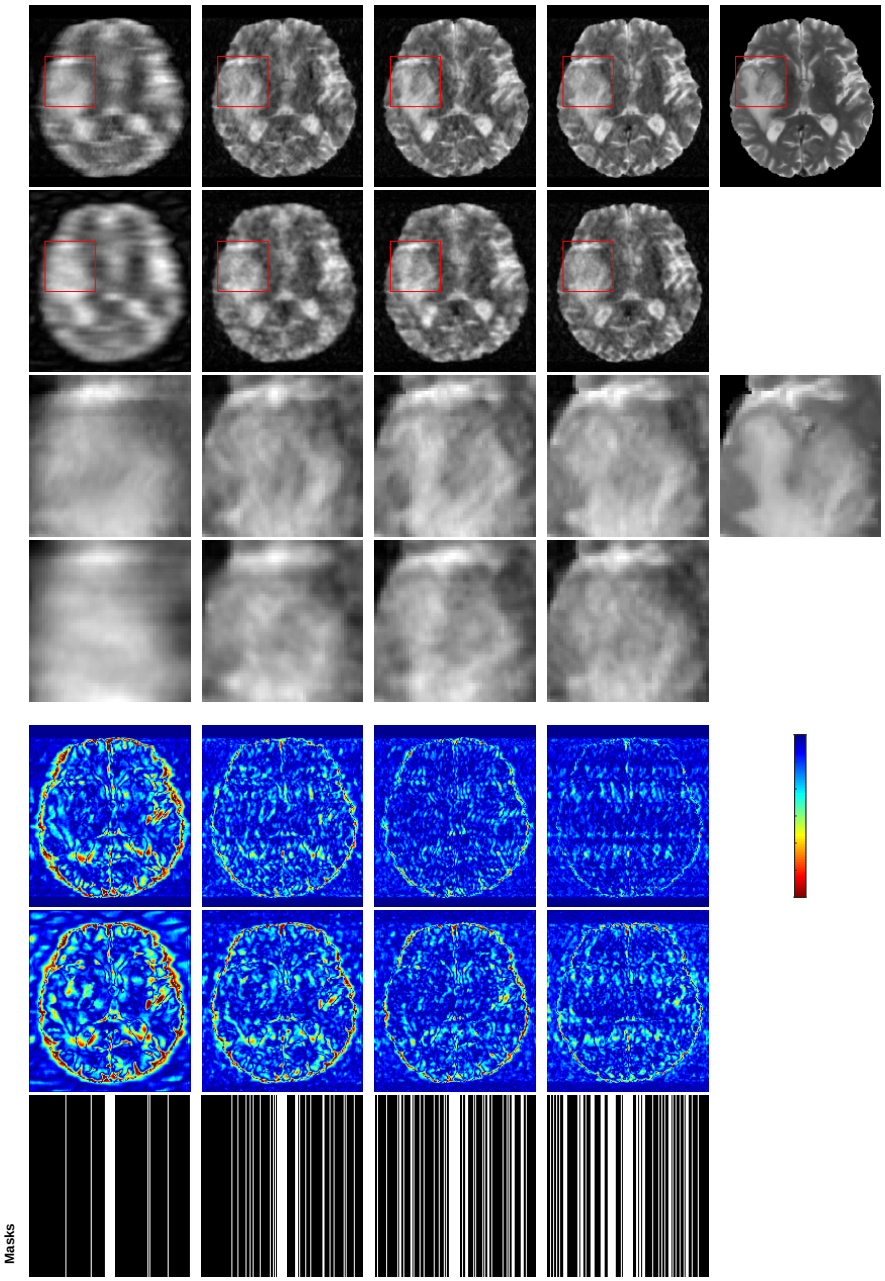

图4-5。从上到下依次展示了T2脑图像重建结果、放大细节、带有色标的逐点误差，以及相应的**笛卡尔**掩模。

接下来，我们在图 2 中通过实验展示算法 4-6的收敛性。这表明，随着阶段数的增加，目标函数值 ϕ 逐渐降低，而测试数据的峰值信噪比值稳步上升，这说明学习得到的算法确实如我们所期望的那样，正在最小化学习得到的函数。

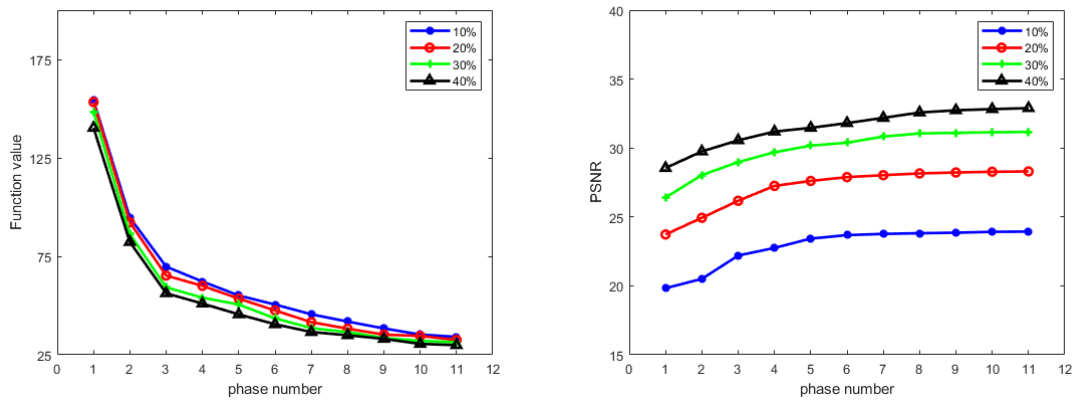

图 4-6. 使用径向掩模，以四种不同的压缩感知比率进行 T1加权磁共振成像图像重建时，算法2的收敛行为。**左**：目标函数值 ϕ 随相位数的变化。**右**：峰值信噪比值随相位数的变化。

4.5.5 未来工作与开放挑战

基于深度优化的元学习技术展现出了很强的泛化能力，但仍有若干尚待解决的挑战，可在未来的工作中加以讨论并有望解决。一个主要问题是记忆化问题，因为基学习器需要针对大量阶段进行优化，而且训练算法包含多个梯度步；此外，计算在时间和内存开销方面都非常昂贵。除了通过不同轨迹重建磁共振成像之外，医学成像的另一个潜在应用是多模态重建与合成。通过多模态采集获取解剖结构图像能够增强诊断信息，也可以将其构建为一个可从元学习中获益的多任务问题。

4.6 结论

本文提出了一种通过元学习进行 MRI 重建的新型深度模型。所提方法能够协同解决多任务，且经过充分训练的模型能够很好地泛化到新任务。我们的基线网络通过展开 LOA 构建，继承了其收敛性质，提高了可解释性，并提升了所设计网络的参数效率

结构。指定的自适应正则化器由任务不变学习器和任务特定元知识构成。网络训练遵循双层优化算法：上层针对验证数据最小化任务特定参数 ω ，下层在固定ω的情况下针对训练数据最小化任务不变参数 θ 。所提方法是首个通过对变分模型中的自适应正则化进行元训练来解决逆问题的模型。我们将不同采样轨迹及各种采样模式下欠采样原始数据的恢复视为不同任务。在多个 MRI 数据集上进行的大量数值实验表明，所提方法能够很好地泛化至各种采样轨迹，并能快速适应未见过的轨迹和采样模式。与传统监督学习相比，无论对于已见还是未见的 k 空间轨迹，重建图像的质量都更高。

4.7 收敛性分析

We 在本研究中，对f和$\mathbf{g}$ 作出以下假设 ：

- (a1)：f 可微且（可能）非凸，并且 ∇f 是L_f-Lipschitz连续的。
- (a2)：$\mathbf{g}$ 的每个分量均可微且（可能）非凸，并且 $\nabla \mathbf{g}$ 是L_g-Lipschitz连续的。
- (a3)：$\sup_{x\in\mathcal{X}} \|\nabla \mathbf{g}(\mathbf{x})\| \leq M$ ，其中$M > 0$为某个常数。
- (a4)： ϕ 是强制的，且 $\phi^* = \min_{x\in\mathcal{X}} \phi(\mathbf{x}) > -\infty$。

首先，我们在引理 [32] 中给出$r(\mathbf{x})$ 的 Clark次微分，并在引理1中证明r_ε 的梯度是Lipschitz连续的2。

引理1。 设$r(\mathbf{x})$ 定义于(4–4)中；则，r 在$\mathbf{x}$ 处的 *Clarke* 次微分为

$$\partial r(\mathbf{x}) = \{\sum_{j\in I_0} \nabla \mathbf{g}_i(\mathbf{x})^\top \mathbf{w}_j + \sum_{j\in I_1} \nabla \mathbf{g}_j(\mathbf{x})^\top \frac{\mathbf{g}_j(\mathbf{x})}{\|\mathbf{g}_j(\mathbf{x})\|} \;\Big|\; \mathbf{w}_j \in \mathbb{R}^d,\ \|\Pi(\mathbf{w}_j; \mathcal{C}(\nabla \mathbf{g}_i(\mathbf{x})))\| \leq 1,\ \forall j \in I_0\}, \tag{4–24}$$

其中 $I_0 = \{j \in [m] \mid \|\mathbf{g}_j(\mathbf{x})\| = 0\}, I_1 = [m] \setminus I_0,$且 $\Pi(\mathbf{w}; \mathcal{C}(\mathbf{A}))$ 是将 $\mathbf{w}$ 投影到$\mathcal{C}(\mathbf{A})$ 上的投影，其中 $\mathcal{C}(\mathbf{A})$ 表示$\mathbf{A}$的列空间。

引理 2. r_ε 的梯度是 *Lipschitz*连续的，其常数为 $m(L_g + {}^2M_\varepsilon{}^2)$。 ——

证明。 由$r_\varepsilon(\mathbf{x}) = \sum^m{}_{j=1}(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2)^{\frac{1}{2}} - \varepsilon$，可得

$$\nabla r_\varepsilon(\mathbf{x}) = \sum_{j=1}^{m} \nabla \mathbf{g}_j(\mathbf{x})^\top \mathbf{g}_j(\mathbf{x})(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2)^{-\frac{1}{2}}. \tag{4–25}$$

对于任意$\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$，我们首先定义$h(\mathbf{x}) = \mathbf{g}_j(\mathbf{x})(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2)^{-\frac{1}{2}}$，于是 $\|h(\mathbf{x})\| < 1$。

$$\|h(\mathbf{x}_1) - h(\mathbf{x}_2)\| \tag{4–26a}$$

$$= \left\| \frac{\mathbf{g}_j(\mathbf{x}_1)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}} - \frac{\mathbf{g}_j(\mathbf{x}_2)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} \right\| \tag{4–26b}$$

$$= \left\| \frac{\mathbf{g}_j(\mathbf{x}_1)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}} - \frac{\mathbf{g}_j(\mathbf{x}_1)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} + \frac{\mathbf{g}_j(\mathbf{x}_1)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} - \frac{\mathbf{g}_j(\mathbf{x}_2)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} \right\| \tag{4–26c}$$

$$\leq \left\| \mathbf{g}_j(\mathbf{x}_1) \left(\frac{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2} - \sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} \right) \right\| + \left\| \frac{\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} \right\| \tag{4–26d}$$

$$\leq \left\| \frac{\mathbf{g}_j(\mathbf{x}_1)}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}} \right\| \left\| \frac{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2} - \sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2}} \right\| + \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \tag{4–26e}$$

$$\leq \frac{1}{\varepsilon} \left\| \sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2} - \sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2} \right\| + \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \tag{4–26f}$$

$$\leq \frac{1}{\varepsilon} \frac{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 - \|\mathbf{g}_j(\mathbf{x}_1)\|^2}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2} + \sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}} + \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \tag{4–26g}$$

$$\leq \frac{1}{\varepsilon} \underbrace{\frac{\|\mathbf{g}_j(\mathbf{x}_2)\| + \|\mathbf{g}_j(\mathbf{x}_1)\|}{\sqrt{\|\mathbf{g}_j(\mathbf{x}_2)\|^2 + \varepsilon^2} + \sqrt{\|\mathbf{g}_j(\mathbf{x}_1)\|^2 + \varepsilon^2}}}_{<1} (\|\mathbf{g}_j(\mathbf{x}_2)\| - \|\mathbf{g}_j(\mathbf{x}_1)\|) + \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \tag{4–26h}$$

$$\leq \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_2) - \mathbf{g}_j(\mathbf{x}_1)\| + \frac{1}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \tag{4–26i}$$

$$= \frac{2}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\|. \tag{4–26j}$$

(4–26f) 的推导过程中，我们使用了 $\left\| \frac{g_j(x_1)}{\sqrt{\|g_j(x_1)\|^2+\varepsilon^2}} \right\| < 1$ 和 $\frac{1}{\sqrt{\|g_j(x_1)\|^2+\varepsilon^2}} < \frac{1}{\varepsilon}$。 –

因此，我们有

$$\|\nabla r_\varepsilon(\mathbf{x}_1) - \nabla r_\varepsilon(\mathbf{x}_2)\| \tag{4–27a}$$

$$= \sum_{j=1}^{m} \left\|\nabla \mathbf{g}_j(\mathbf{x}_1)^\top h(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)^\top h(\mathbf{x}_2)\right\| \tag{4–27b}$$

$$= \sum_{j=1}^{m} \left\|\nabla \mathbf{g}_j(\mathbf{x}_1)^\top h(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)^\top h(\mathbf{x}_1) + \nabla \mathbf{g}_j(\mathbf{x}_2)^\top h(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)^\top h(\mathbf{x}_2)\right\| \tag{4–27c}$$

$$\leq \sum_{j=1}^{m} \left\|\left(\nabla \mathbf{g}_j(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)\right)^\top h(\mathbf{x}_1)\right\| + \left\|\nabla \mathbf{g}_j(\mathbf{x}_2)\left(h(\mathbf{x}_1) - h(\mathbf{x}_2)\right)\right\| \tag{4–27d}$$

$$\leq \sum_{j=1}^{m} \|\nabla \mathbf{g}_j(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)\| \, \|h(\mathbf{x}_1)\| + \|\nabla \mathbf{g}_j(\mathbf{x}_2)\| \, \|h(\mathbf{x}_1) - h(\mathbf{x}_2)\|$$

$$\leq \sum_{j=1}^{m} \|\nabla \mathbf{g}_j(\mathbf{x}_1) - \nabla \mathbf{g}_j(\mathbf{x}_2)\| + \|\nabla \mathbf{g}_j(\mathbf{x}_2)\| \frac{2}{\varepsilon} \|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| \text{ by (4–26) and} \tag{4–27e}$$

$$\leq m(L_g \|\mathbf{x}_1 - \mathbf{x}_2\| + M\frac{2}{\varepsilon} \cdot M\|\mathbf{x}_1 - \mathbf{x}_2\|), \tag{4–27f}$$

最后一个不等式的第一项源于 $\nabla \mathbf{g}_j$ 的 L_g-Lipschitz 连续性。第二项则是因为，由中值定理可知，对于某个 $\tilde{\mathbf{x}} \in \mathcal{X}$，$\|\mathbf{g}_j(\mathbf{x}_1) - \mathbf{g}_j(\mathbf{x}_2)\| = \|\nabla \mathbf{g}_j(\tilde{\mathbf{x}})(\mathbf{x}_1 - \mathbf{x}_2)\|$，且 $\|\nabla \mathbf{g}_j(\tilde{\mathbf{x}})\| \leq \sup_{x \in \mathcal{X}} \|\nabla \mathbf{g}_j(\mathbf{x})\| \leq M$。因此，我们得到

$$\|\nabla r_\varepsilon(\mathbf{x}_1) - \nabla r_\varepsilon(\mathbf{x}_2)\| \leq m(L_g + \frac{2M^2}{\varepsilon})\|\mathbf{x}_1 - \mathbf{x}_2\|. \tag{4–28}$$

□

引理 3. 设 ε，η，τ_t，$a > 0$, $0 < \rho < 1$，并选择初始值 $\mathbf{x}_0 \in \mathcal{X}$。假设序列 $\{\mathbf{x}_t\}$ 是通过在固定 $\varepsilon_t = \varepsilon$ 的情况下执行算法 *2* 的第 *3–14* 行生成的，并且存在 $0 < \delta < \varepsilon/(a+L)^2 < 1$，其中 $L_\varepsilon = L_f + m(L_g + M_\varepsilon^2)$ 且 $\phi^* := \min_{x \in \mathcal{X}} \phi(\mathbf{x})$。那么，以下结论成立：

1. $\|\nabla \phi_\varepsilon(\mathbf{x}_t)\| \to 0$ 作为 $t \to \infty$. 2. $\max\{t \in \mathbb{N} \mid \|\nabla \phi_\varepsilon(\mathbf{x}_{t+1})\| \geq \eta\} \leq \max\{2aL_\varepsilon\delta^2\eta^2, a\eta^3\}(\phi_\varepsilon(\mathbf{x}_0) - \phi^* + \varepsilon)$。

证明。 1. 在每次迭代中，我们计算 $\mathbf{u}_{t+1} = \mathbf{z}_{t+1} - \tau_t \sigma(\omega_i) \nabla r_{\varepsilon_t}(\mathbf{z}_{t+1})$。

1.1. 在条件成立的情况下

$$\|\nabla\phi_\varepsilon(\mathbf{x}_t)\| \le a\|\mathbf{u}_{t+1}-\mathbf{x}_t\| \quad \text{and} \quad \phi_\varepsilon(\mathbf{u}_{t+1})-\phi_\varepsilon(\mathbf{x}_t) \le -\frac{1}{a}\|\mathbf{u}_{t+1}-\mathbf{x}_t\|^2 \tag{4–29}$$

对 $a > 0$ 成立，我们令 $\mathbf{x}_{t+1} = \mathbf{u}_{t+1}$，并且有 $\phi_\varepsilon(\mathbf{u}_{t+1}) \le \phi_\varepsilon(\mathbf{x}_t)$。

1.2. 否则，我们计算 $\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t\nabla\phi_\varepsilon(\mathbf{x}_t)$，其中 α_t 通过线搜索确定，直到满足条件

$$\phi_\varepsilon(\mathbf{v}_{t+1})-\phi_\varepsilon(\mathbf{x}_t) \le -\frac{1}{a}\|\mathbf{v}_{t+1}-\mathbf{x}_t\|^2 \tag{4–30}$$

成立，然后令$\mathbf{x}_{t+1} = \mathbf{v}_{t+1}$。根据引理 2，我们知道梯度$\nabla r_\varepsilon(\mathbf{x})$ 是 Lipschitz连续的，其常数为$m(L_g + {}^2M_\varepsilon{}^2)$。此外，我们在 (a1) 中假设 ∇f 是 L_f-Lipschitz连续的。因此，令$L_\varepsilon = L_f + m(L_g + {}^2M_\varepsilon{}^2)$，可得 $\nabla\phi_\varepsilon$ 是 L_ε-Lipschitz连续的，这意味着

$$\phi_\varepsilon(\mathbf{v}_{t+1}) \le \phi_\varepsilon(\mathbf{x}_t) + \langle\nabla\phi_\varepsilon(\mathbf{x}_t), \mathbf{v}_{t+1}-\mathbf{x}_t\rangle + \frac{L_\varepsilon}{2}\|\mathbf{v}_{t+1}-\mathbf{x}_t\|^2. \tag{4–31}$$

此外，根据.....的最优性条件

$$\mathbf{v}_{t+1} = \arg\min_{\mathbf{x}} \langle\nabla f(\mathbf{x}_t), \mathbf{x}-\mathbf{x}_t\rangle + \sigma(\omega_i)\langle\nabla r_\varepsilon(\mathbf{x}_t), \mathbf{x}-\mathbf{x}_t\rangle + \frac{1}{2\alpha_t}\|\mathbf{x}-\mathbf{x}_t\|^2,$$

我们有

$$\langle\nabla\phi_\varepsilon(\mathbf{x}_t), \mathbf{v}_{t+1}-\mathbf{x}_t\rangle + \frac{1}{2\alpha_t}\|\mathbf{v}_{t+1}-\mathbf{x}_t\|^2 \le 0. \tag{4–32}$$

结合（4–31）和（4–32），以及第 8 行中的 $\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t\nabla\phi_\varepsilon(\mathbf{x}_t)$ 算法第 8 行 2 可得

$$\phi_\varepsilon(\mathbf{v}_{t+1})-\phi_\varepsilon(\mathbf{x}_t) \le -\left(\frac{1}{2\alpha_t}-\frac{L_\varepsilon}{2}\right)\|\mathbf{v}_{t+1}-\mathbf{x}_t\|^2. \tag{4–33}$$

因此，只需满足$\alpha_t \le {}^1/_{2+aL_\varepsilon}$ 条件（4–30）即可。由于我们可以找到有限的t使得$\rho^t\alpha_t \le {}^1/_{2+aL_\varepsilon}$，并且通过线搜索可以得到$\phi_\varepsilon(\mathbf{v}_{t+1}) \le \phi_\varepsilon(\mathbf{x}_t)$，因此该过程只需有限次迭代即可完成。

因此，在1.1. 或1.2. 这两种情况下，当我们取$\mathbf{x}_{t+1} = \mathbf{u}_{t+1}$ 或$\mathbf{v}_{t+1}$时，可以得到

$$\phi_\varepsilon(\mathbf{x}_{t+1}) \le \phi_\varepsilon(\mathbf{x}_t), \text{ for all } t \ge 0. \tag{4–34}$$

现在，由情形 1.1.，(4–29) 得到

$$\|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le a^2 \|\mathbf{u}_{t+1} - \mathbf{x}_t\|^2 \le a^3 \left(\phi_\varepsilon(\mathbf{x}_t) - \phi_\varepsilon(\mathbf{u}_{t+1})\right), \tag{4–35a}$$

$$\text{therefore if } \mathbf{x}_{t+1} = \mathbf{u}_{t+1} \text{ we get } \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le a^3 \left(\phi_\varepsilon(\mathbf{x}_t) - \phi_\varepsilon(\mathbf{x}_{t+1})\right). \tag{4–35b}$$

由情形 1.2. 以及 $\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t \nabla\phi_\varepsilon(\mathbf{x}_t)$，可得

$$\phi_\varepsilon(\mathbf{v}_{t+1}) - \phi_\varepsilon(\mathbf{x}_t) \le -\frac{1}{a}\|\mathbf{v}_{t+1} - \mathbf{x}_t\|^2 = -\frac{1}{a}\alpha_t^2 \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \tag{4–36a}$$

$$\Longrightarrow \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le \frac{a}{\alpha_t^2}\big(\phi_\varepsilon(\mathbf{x}_t) - \phi_\varepsilon(\mathbf{v}_{t+1})\big), \tag{4–36b}$$

$$\text{then if } \mathbf{x}_{t+1} = \mathbf{v}_{t+1}, \text{ we have } \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le \frac{a}{\alpha_t^2}\big(\phi_\varepsilon(\mathbf{x}_t) - \phi_\varepsilon(\mathbf{x}_{t+1})\big). \tag{4–36c}$$

由于 $\frac{\delta}{L_\varepsilon} \le \alpha_t \le \frac{1}{2/+aL_\varepsilon}$，可得

$$\|\nabla\phi_\varepsilon(\mathbf{v}_{t+1})\|^2 \le \frac{aL_\varepsilon^2}{\delta^2}\big(\phi_\varepsilon(\mathbf{v}_{t+1}) - \phi_\varepsilon(\mathbf{x}_{t+1})\big). \tag{4–37}$$

结合 (4–35b) 和 (4–37)，并取 $C = \max\{\frac{aL_\varepsilon^2}{\delta^2}, a^3\}$，得到 —

$$\|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le C\big(\phi_\varepsilon(\mathbf{x}_t) - \phi_\varepsilon(\mathbf{x}_{t+1})\big). \tag{4–38}$$

对 $t = 0, \cdots, T$ 对 (4–38) 求和，可得

$$\sum_{t=0}^{T} \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le C\big(\phi_\varepsilon(\mathbf{x}_0) - \phi_\varepsilon(\mathbf{x}_{T+1})\big). \tag{4–39}$$

结合如下事实：对每个 $\mathbf{x} \in \mathcal{X}$，$\phi_\varepsilon(\mathbf{x}) \ge \phi(\mathbf{x}) - \varepsilon \ge \phi^* - \varepsilon$，可得

$$\sum_{t=0}^{T} \|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \le C\left(\phi_\varepsilon(\mathbf{x}_0) - \phi^* + \varepsilon\right). \tag{4–40}$$

右边是一个有限常数，因此令$t \to \infty$，我们知道$\|\nabla\phi_\varepsilon(\mathbf{x}_t)\| \to 0$，这就证明了第一个结论。

2. 记 $\kappa := \max\{t \in \mathbb{N} \mid \|\nabla\phi_\varepsilon(\mathbf{x}_{t+1})\| \geq \eta\}$；那么，我们知道对于所有$t \leq \kappa - 1$, $\|\nabla\phi_\varepsilon(\mathbf{x}_{t+1})\| \geq \eta$。因此，我们有

$$\kappa\eta^2 \leq \sum_{t=0}^{\kappa-1}\|\nabla\phi_\varepsilon(\mathbf{x}_{t+1})\|^2 = \sum_{t=1}^{\kappa}\|\nabla\phi_\varepsilon(\mathbf{x}_t)\|^2 \leq C\left(\phi_\varepsilon(\mathbf{x}_0) - \phi^* + \varepsilon\right). \tag{4–41}$$

这意味着第二个陈述成立。

□

引理 4。 设序列 $\{\mathbf{x}_t\}$ 由算法 *1* 以初始猜测$\mathbf{x}_0$生成。那么，对于任意 $t \geq 0$，都有 $\phi_{\varepsilon_{t+1}}(\mathbf{x}_{t+1}) + \varepsilon_{t+1} \leq \phi_{\varepsilon_t}(\mathbf{x}_t) + \varepsilon_t$。

证明。 为证明这一陈述，我们可以证明

$$\phi_{\varepsilon_{t+1}}(\mathbf{x}_{t+1}) + \varepsilon_{t+1} \leq \phi_{\varepsilon_t}(\mathbf{x}_{t+1}) + \varepsilon_t \leq \phi_{\varepsilon_t}(\mathbf{x}_t) + \varepsilon_t. \tag{4–42}$$

第二个不等式可直接由（4–34）得到。现在，我们证明第一个不等式。

对于任意 $\varepsilon > 0$，记

$$r_{\varepsilon,j}(\mathbf{x}) = \sqrt{\|\mathbf{g}_j(\mathbf{x})\|_2^2 + \varepsilon^2} - \varepsilon. \tag{4–43}$$

由于 $\phi_\varepsilon(\mathbf{x}) = f(\mathbf{x}) + \sigma(\omega_i)\sum_{j=1}^{m} r_{\varepsilon,\,j}(\mathbf{x})$，只需证明

$$r_{\varepsilon_{t+1},j}(\mathbf{x}_{t+1}) + \varepsilon_{t+1} \leq r_{\varepsilon_t,j}(\mathbf{x}_{t+1}) + \varepsilon_t \tag{4–44}$$

若 $\varepsilon_{t+1} = \varepsilon_t$，则上述两个量相同，且第一个不等式成立。

现在，假设 $\varepsilon_{t+1} = \gamma\varepsilon_t \leq \varepsilon_t$；则，

$$r_{\varepsilon_{t+1},j}(\mathbf{x}_{t+1}) + \varepsilon_{t+1} = \sqrt{\|\mathbf{g}_j(\mathbf{x})\|_2^2 + \varepsilon_{t+1}^2} \leq \sqrt{\|\mathbf{g}_j(\mathbf{x})\|_2^2 + \varepsilon_t^2} = r_{\varepsilon_t,j}(\mathbf{x}_{t+1}) + \varepsilon_t, \tag{4–45}$$

这意味着（4–42）中的第一个不等式。 □

定理 4.1。 设 $\{\mathbf{x}_t\}$ 是由算法 *2* 在任意初始值 $\mathbf{x}_0$ 下生成的序列，$\epsilon_{\text{tol}} = 0$ 且 $T = \infty$。令 $\{\mathbf{x}_{t_l+1}\}$ 为满足算法 *2* 第 *15* 步约简准则的子序列，即 $\|\nabla\phi_{\varepsilon_{t_l}}(\mathbf{x}_{t_l+1})\| \leq \sigma\varepsilon_{t_l}\gamma$ 对于 $t = t_l$ 以及 $l = 1, 2, \cdots$。

那么 $\{\mathbf{x}_{t_l+1}\}$至少有一个聚点，并且 $\{\mathbf{x}_{t_l+1}\}$的每个聚点都是$\min_{\mathbf{x}} \phi(\mathbf{x}) := f(\mathbf{x}) + \sigma(\omega_i) r(\mathbf{x})$的 *Clarke*驻点。

证明。 根据引理 4 和 $\phi(\mathbf{x}) \leq \phi_\varepsilon(\mathbf{x}) + \varepsilon$ 对所有 $\varepsilon > 0$ 以及 $\mathbf{x} \in \mathcal{X}$，我们知道

$$\phi(\mathbf{x}_t) \leq \phi_{\varepsilon_t}(\mathbf{x}_t) + \varepsilon_t \leq \cdots \leq \phi_{\varepsilon_0}(\mathbf{x}_0) + \varepsilon_0 < \infty. \tag{4–46}$$

由于 ϕ是强制的，我们知道 $\{\mathbf{x}_t\}$是有界的，并且所选取的子序列 $\{\mathbf{x}_{t_l+1}\}$也有界，且至少有一个聚点。

注意，当 $\mathrm{l} \to \infty$时，$\|\nabla \phi_{\varepsilon_{t_l}}(\mathrm{x}_{t_l+1})\| \leq \sigma \varepsilon_{t_l} \gamma = \sigma \varepsilon_0 \gamma^{l+1} \to 0$。令 $\{\mathrm{x}_{p+1}\}$为 $\{\mathrm{x}_{t_l+1}\}$ 的任意收敛子序列，并将 ε_p 记为算法2中生成 x^{p+1}时所使用的相应 ε_t。那么，存在 $\mathrm{x}^* \in \mathcal{X}$，使得当 $\varepsilon_p \to 0$时，$\mathrm{x}_{p+1} \to \mathrm{x}^*$，并且当 $\mathrm{p} \to \infty$时，$\nabla \phi_{\varepsilon_p}(\mathrm{x}^{p+1}) \to 0$。

注意，ϕ在 $\mathbf{x}^*$处的 Clarke 次微分为 $\partial \phi(\mathbf{x}^*) = \partial f(\mathbf{x}^*) + \sigma(\omega_i) \partial r(\mathbf{x}^*)$ ：

$$\partial \phi(\hat{\mathbf{x}}) = \{\nabla f(\hat{\mathbf{x}}) + \sigma(\omega_i) \sum_{j \in I_0} \nabla \mathbf{g}_j(\hat{\mathbf{x}})^\top \mathbf{w}_j + \sigma(\omega_i) \sum_{j \in I_1} \nabla \mathbf{g}_j(\hat{\mathbf{x}})^\top \frac{\mathbf{g}_j(\hat{\mathbf{x}})}{\|\mathbf{g}_j(\hat{\mathbf{x}})\|} \Bigg| \\ \|\Pi(\mathbf{w}_j; \mathcal{C}(\nabla \mathbf{g}_j(\hat{\mathbf{x}})))\| \leq 1, \ \forall j \in I_0\}, \tag{4–47}$$

其中 $I_0 = \{j \in [m] \mid \|\mathbf{g}_i(\hat{\mathbf{x}})\| = 0\}$ 和 $I_1 = [m] \setminus I_0$。

若 $j \in I_0$,我们有 $\|\mathbf{g}_j(\mathbf{x})\| = 0 \iff \mathbf{g}_j(\mathbf{x}) = 0$: 那么,

$$\partial r_\varepsilon(\mathbf{x}) = \sum_{j \in I_0} \nabla \mathbf{g}_j(\mathbf{x})^\top \frac{\mathbf{g}_j(\mathbf{x})}{\left(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2\right)^{\frac{1}{2}}} + \sum_{j \in I_1} \nabla \mathbf{g}_j(\mathbf{x})^\top \frac{\mathbf{g}_j(\mathbf{x})}{\left(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2\right)^{\frac{1}{2}}} \tag{4–48a}$$

$$= \mathbf{0} + \sum_{j \in I_1} \nabla \mathbf{g}_j(\mathbf{x})^\top \frac{\mathbf{g}_j(\mathbf{x})}{\left(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon^2\right)^{\frac{1}{2}}} \tag{4–48b}$$

因此，我们得到

$$\nabla \phi_{\varepsilon_p}(\mathbf{x}_{p+1}) = \nabla f(\mathbf{x}_{p+1}) + \sigma(\omega_i) \sum_{j \in I_1} \nabla \mathbf{g}_j(\mathbf{x})^\top \frac{\mathbf{g}_j(\mathbf{x})}{\left(\|\mathbf{g}_j(\mathbf{x})\|^2 + \varepsilon_p^2\right)^{\frac{1}{2}}}. \tag{4–49}$$

比较（4–47）和（4–49）可知，（4–49）右侧的第一项收敛于（4–47）的对应项，这是因为 $\mathbf{x}_{p+1} \to \hat{\mathbf{x}}$ 以及 $\nabla \mathbf{f}$ 的连续性。结合 $\mathbf{g}_i$ 、 $\nabla \mathbf{g}_i$ 的连续性，当 $\varepsilon_p \to 0$ 和 $\|\mathbf{g}_j(\mathbf{x})\| > 0$ 时，（4–47）的最后一项收敛于（4–49）的最后一项。此外，显然，$\mathbf{0}$ 是（4–49）中第二项的一种特殊情况。因此，我们知道

$$\operatorname{dist}(\nabla \phi_{\varepsilon_p}(\mathbf{x}_{p+1}), \partial \phi(\hat{\mathbf{x}})) \to 0,$$

作为 $p \to \infty$。由于 $\nabla \phi_{\varepsilon_p}(\mathbf{x}_{p+1}) \to 0$ 和 $\partial \phi(\hat{\mathbf{x}})$ 是闭集，我们得出 $0 \in \partial \phi(\hat{\mathbf{x}})$。 □

第5章 用于多模态 MRI 联合重建与合成的基于优化的模型

生成同一解剖结构的多对比度/多模态 MRI 可丰富诊断信息，但在实际应用中受限于过长的数据采集时间。本文提出了一种新颖的深度学习模型，利用多个源模态的不完整 k 空间数据作为输入，实现多模态 MRI 的联合重建与合成。模型的输出包括源模态的重建图像，以及目标模态中的高质量合成图像。所提出的模型被表述为一个变分问题，其中利用了多个可学习的模态特定特征提取器和一个多模态合成模块。我们提出了一种可学习优化算法来求解该模型，由此构建出一个多阶段网络，其参数可使用多模态 MRI 数据进行训练。此外，我们采用双层优化框架进行稳健的参数训练。我们通过大量数值实验验证了所提方法的有效性。

5.1 引言

磁共振成像（MRI）是一项重要的医学成像技术，能够在相同解剖结构下提供丰富多样的图像对比度，从而增加诊断信息。与单一模态相比，多模态磁共振图像可为临床应用和研究提供更多诊断信息[224, 87, 39, 182, 86]。通过改变采集参数——重复时间（TR）和回波时间（TE）——可生成多种不同对比度的图像。这些图像的解剖结构相似，但能突出显示不同的软组织。例如，采集 T1加权图像时使用较短的 TR 和 TE，高脂肪含量区域呈亮色，脑脊液（CSF）充盈的腔隙呈暗色，因此能提供更多解剖信息。T1 脑部图像可区分灰质和白质组织。T2加权图像需要较长的 TR 和 TE。一般而言，T2加权图像的对比度与 T1加权图像相反，且通常能提供更多

用于界定水肿的病理信息。T2图像可区分液体与皮质组织。液体衰减反转恢复（FLAIR）需要非常长的 TR 和 TE，其对比度与 T2加权图像相似，但会抑制脑脊液，使自由水呈暗色，并最大限度地降低灰质与白质之间的对比度。磁共振成像的一项主要局限是扫描期间数据采集时间相对较长，这会使患者感到不适或引入运动伪影，从而降低诊断可及性。缩短扫描时间的一种主要方法是重建欠采样的 k空间采集数据，另一种方法是根据全采样数据合成缺失的对比度磁共振图像 [204, 42]。本文将磁共振图像的不同对比度称为不同模态。我们提出联合重建来自多个可用模态（即源模态）的欠采样 k空间磁共振数据，并根据源模态合成缺失模态（即目标模态）的图像。

压缩感知磁共振成像（CS-MRI）重建是加速磁共振采集的主要方法，它通过求解变分模型所构建的逆问题来实现。传统的压缩感知磁共振成像采用手工设计的正则化项（例如全变分）为待重建图像引入先验信息 [18, 83, 117, 50,144, 200]。近几十年来，基于深度学习的模型利用大型数据集，进一步探索相较于传统方法提升重建性能的潜力，并已成功应用于临床领域。大多数基于深度学习的重建方法采用端到端深度网络，例如基于GAN的方法：DAGAN [199], RefineGAN [146]；级联网络方法：E2E-VN [168],Cascade-Net [160]；以及采用深度残差学习的方法 [99, 100, 40], 等。为克服端到端网络中黑箱模型的不足，人们开发了受可学习优化算法（LOA）启发的若干方法，这些方法具有更易解释的网络架构。基于LOA的重建方法将迭代优化算法展开为多阶段网络，在该网络中，正则化参数和图像变换可通过优化算法进行有效学习

以提升网络性能 [132, 111, 17, 69, 172, 218, 220, 79]。例如：ADMM-Net [172] 通过展开交替方向乘子法（ADMM）算法而提出。ISTA-Net+ [218]将深度神经网络嵌入迭代收缩阈值算法（ISTA），以解决压缩感知磁共振成像重建问题。PD-Net[35] 展开原始-对偶混合梯度算法，其中近端算子被参数化为两个通过训练数据学习得到的网络模块。

近年来，采用各种深度学习框架进行磁共振图像合成日益受到关注，大致可分为单模态合成 [41, 165, 195, 201,212] 和多模态合成 [25, 42, 161, 225, 214]。单模态合成是一种一对一的方法，旨在根据相应的单一源模态估计目标模态的图像。一种常见的基于深度学习的策略是采用对抗方法，例如 pGAN [41], ，通过最小化对抗损失函数来捕获可靠的高频纹理信息。对于配对数据转换，基于图像块的方法 [92, 175] 和基于图谱的方法 [130, 153, 22] 取得了巨大成功。例如， [154] 提出了一种图像块匹配方法：首先将多个带有 T1加权成像和 T2加权成像的图谱配准到目标 T1加权成像，然后根据目标 T1加权成像的图谱图像块，组合图谱 T2加权成像图像块来合成 T2加权成像。[91] 估计受试者的脉冲序列参数，并合成一幅额外的图谱图像；随后通过在图像块上进行随机森林回归学习，学习额外图谱图像与目标图谱图像之间的强度变换；最后将回归应用于给定的受试者图像，以获得目标模态的合成图像。对于非配对数据转换，近年来，图像到图像转换在多模态医学成像中备受关注，其中包括基于生成对抗网络的模型 [30, 195] ，其生成器被设计为确定性映射；以及基于流的方法 [21] ，其生成器是可逆映射，可通过将数据从源域映射到目标域实现循环一致转换，并确保数据点返回源域。多模态合成是一种多对一或多对多类型的方法。大多数多模态合成采用基于生成对抗网络的网络。主流的

上述合成方法大多是采用编码器-解码器架构的端到端网络，具体而言，这些架构包含在对抗学习的生成器网络中。例如，MM-GAN [161] 将所有可用模态沿通道维度与代表缺失模态的零图像拼接，并结合课程学习为 GAN 填补缺失的输入。多模态磁共振（MM）[114],MMGradAdv [25]和 Hi-Net [225]利用多模态数据之间的相关性，并采用稳健的特征融合方法形成统一的潜在表示。

为了利用源模态的部分扫描 k 空间数据合成目标模态，而非使用最先进的多模态合成方法中采用的完整扫描数据，本文提出了一种基于 LOA 的磁共振成像联合重建与合成深度神经网络。该网络的输入是多个部分采样的源模态，输出包括重建目标模态和合成目标模态。我们的贡献总结如下：

(1) 我们提出了一种新颖的 LOA，用于多模态磁共振成像的联合重建与合成，并提供理论分析保证；

(2) 通过双层优化算法训练网络参数，以缓解过拟合问题；

(3) 大量实验结果证明了所提方法的有效性和优越性。

5.2 所提出的方法

5.2.1 模型

本节将详细介绍所提出的用于联合磁共振成像重建与合成的算法及其对应网络。给定源模态（例如：T1加权成像和T2加权成像）的部分 k 空间数据 $\{\mathbf{f}_1,\mathbf{f}_2\}$，我们的任务是重建相应的图像 $\{\mathbf{x}_1,\mathbf{x}_2\}$，并合成$\mathbf{x}_3$（例如：液体衰减反转恢复），而无需提供任何 k 空间信息。我们的模型通过求解以下优化问题，联合重建并合成这些模态

$$\min_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \Psi_{\Theta,\gamma}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3) := \sum_{i=1}^{2} \frac{1}{2}\|P_i F\,\mathbf{x}_i - \mathbf{f}_i\|_2^2 + \sum_{i=1}^{3} \|h_{w_i}(\mathbf{x}_i)\|_{2,1} + \tag{5–1a}$$

$$\frac{\gamma}{2}\|g_\theta([h_{w_1}(\mathbf{x}_1), h_{w_2}(\mathbf{x}_2)]) - \mathbf{x}_3\|_2^2, \tag{5–1b}$$

其中，第一项是源模态的数据保真度，用于确保重建图像 $\{\mathbf{x}_1,\mathbf{x}_2\}$ 与感知的部分 k 空间数据 $\{\mathbf{f}_1,\mathbf{f}_2\}$保持一致。$h_{w_i}$ 表示模态特定特征提取算子，它将输入$\mathbf{x}_i \in \mathbb{C}^n$映射为高维特征张量$h_{w_i}(\mathbf{x}_i) \in \mathbb{C}^{m\times d}$，其中$m$为特征图的空间维度，$d$ 为其通道数。第二项是所有模态 $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ 的先验项，用于在学习变换下增强稀疏性，其定义如下

$$\|h_{w_i}(\mathbf{x}_i)\|_{2,1} = \sum_{j=1}^{m} \|h_{w_i,j}(\mathbf{x}_i)\|, \tag{5–2}$$

其中$h_{w_i}(*) = (h_{w_i,1}(*), \ldots, h_{w_i,m}(*))$，每个$h_{w_i,j}(*)$都参数化为用于$j = 1, \cdots, m$的卷积神经网络（CNN），而$w_i$是相应的可学习网络参数。这里，每个$h_{w_i,j} \in \mathbb{R}^d$都可以视为空间位置$j$处的特征向量。

在（5–1）的第三项中，$[\cdot,\cdot]$ 表示参数的拼接，而 $g_\theta: \mathbb{C}^{m\times 2d} \to \mathbb{C}^n$ 是多模态合成模块，它以 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 的拼接特征为输入，并合成缺失模态 $\mathbf{x}_3$。为此，我们添加第三项，尝试最小化合成结果 $g_\theta([h_{w_1}(\mathbf{x}_1), h_{w_2}(\mathbf{x}_2)])$与目标模态 $\mathbf{x}_3$之间的差异。

此外，Θ 在（5–1）中收集了函数卷积层中的参数Ψ，即 $\Theta = \{w_1, w_2, w_3, \theta\}$。系数 λ 被参数化为一个超参数，通过最小化验证集上的重建损失进行调优。权重 γ 是一个超参数，在平衡模型（5–1）的重建部分（（5–1）中的前两项）和图像合成部分（（5–1）中的最后一项）方面发挥着关键作用，因此会显著影响最终的图像重建和合成质量。为解决

这一重要问题，我们提出使用双层超参数调优框架，通过最小化验证集和训练数据集上的重建损失来学习 γ。第5.2.4节将介绍这种超参数训练的详细过程。在接下来的小节中，我们将详细介绍上述模块的结构。

5.2.1.1 模态特定特征提取算子

我们将模态特定特征提取算子 h_{w_i} 设计为一个不含偏置、带有非线性激活函数 σ 的普通l层卷积神经网络，如下所示：

$$h_{w_i}(\mathbf{x}) = \mathbf{c}_{w_i,l} * \sigma \cdots \ \sigma(\mathbf{c}_{w_i,3} * \sigma(\mathbf{c}_{w_i,2} * \sigma(\mathbf{c}_{w_i,1} * \mathbf{x}))), \tag{5–3}$$

其中，$\{\mathbf{c}_{w_i,q}\}_{q=1}^{l}$ 表示由d 个具有相同空间卷积核尺寸（$\kappa \times \kappa$）的卷积核组成的卷积权重，$*$ 表示卷积操作。在此，我们使用平滑修正线性单元 [33]作为激活函数 σ。除平滑的 σ之外，式（5–3）中h_{w_i} 的每个卷积操作都是线性算子，这使得h_{w_i} 可微；根据链式法则，可以很容易地得到 ∇h_{w_i}，其中每个$\mathbf{c}_{w_i,q}^{\top}$ 均可实现为转置卷积运算 [49]。

5.2.1.2 多模态合成模块

多模态合成模块g_θ： $\mathbb{C}^{m\times 2d} \to \mathbb{C}^n$ 是一个多模态合成模块，以拼接特征 [$h_{w_1}(\mathbf{x}_1)$、$h_{w_2}(\mathbf{x}_2)$] 以及 $\{\mathbf{x}_1,\mathbf{x}_2\}$ 作为输入，并合成缺失模态$\mathbf{x}_3$。与h_{w_i}类似，g_θ 也被参数化为由 σ分隔的卷积算子堆叠：

$$g_\theta(\mathbf{x}) = \mathbf{c}_{\theta,l} * \sigma \cdots \ \sigma(\mathbf{c}_{\theta,3} * \sigma(\mathbf{c}_{\theta,2} * \sigma(\mathbf{c}_{\theta,1} * \mathbf{x}))). \tag{5–4}$$

图 5-2 展示了算子 h_{w_i} 和 g_θ的示意图。

5.2.2 高效可学习优化算法

本节提出一种新颖且高效的可学习优化算法（LOA），用于求解非凸非光滑最小化问题（5–1）。随后，我们设计

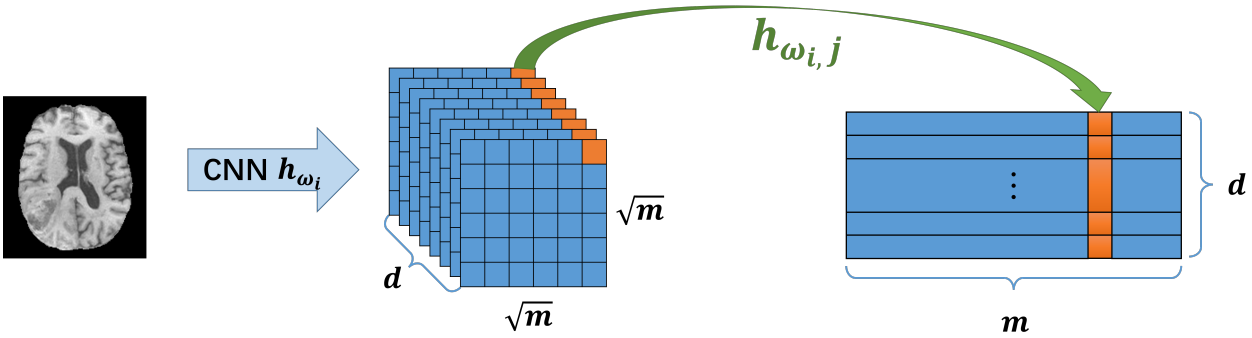

图5-1。空间位置j处的特征张量 h_{w_i}和特征向量 $h_{w_i,j}$。

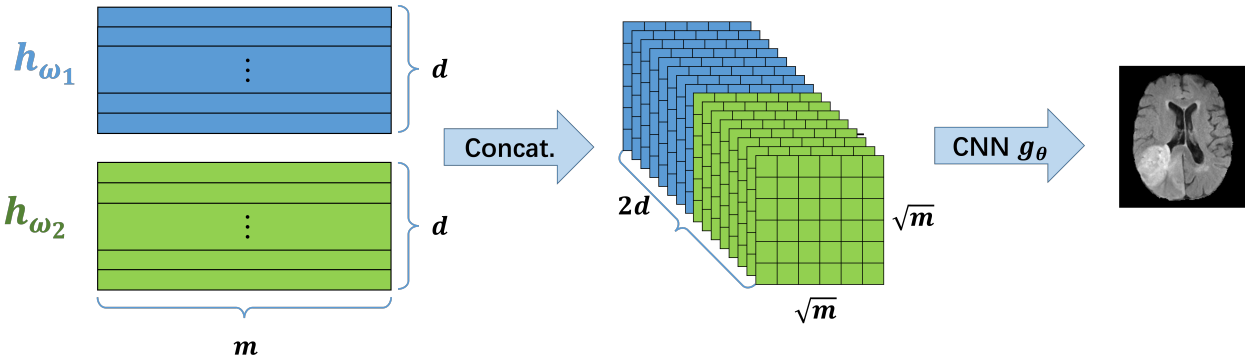

图5-2。合成映射 g_θ 将拼接特征 $[h_{w_1}, h_{w_2}]$ 映射到图像 $\mathbf{x}_3$。

一个架构严格遵循该算法的深度神经网络（DNN），并且可以从数据中学习该网络的参数。这样，DNN便继承了LOA的所有收敛性质。

由于式（5–2）中的 $l_{2,1}$-范数是非光滑的，本文考虑 $l_{2,1}$-范数的一个平滑版本，称为 $l_{\varepsilon_2,1}$-范数，其定义为 $\|h_{w_i}(x_i)\|_{\varepsilon_2,1} = \sum^{m}_{j=1} \sqrt{\|h_{w_i,j}(x_i)\|^2 + \varepsilon^2} - \varepsilon$。通过这~~一修改，得到~~的新函数 Ψ_ε 是光滑的，从而为我们提供了另一种使用梯度下降算法求解最小化问题（5–1）的方法。当 ε 趋于 0 时，替代函数趋近于式（5–1）中的原始非光滑正则化项。所提出的算法的一个关键特征是引入了一种平滑级别降低机制，该机制会在迭代过程中自动减小 ε，具体说明如下。

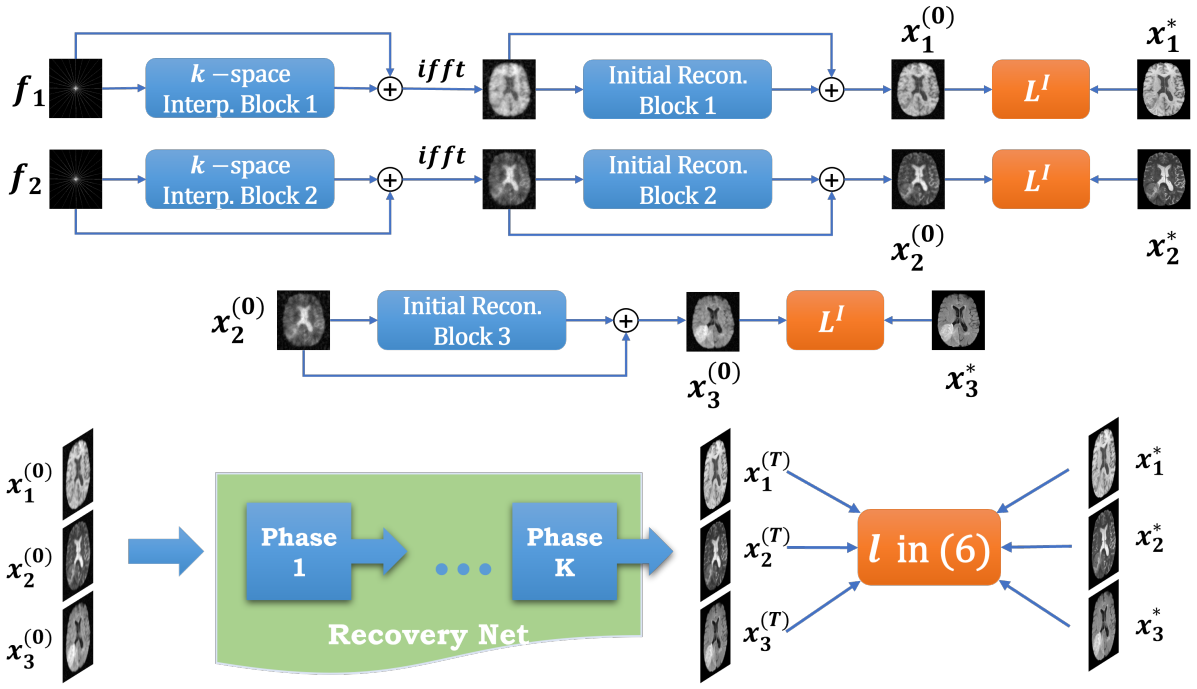

图5-3。所提出的网络架构。我们将L^l 定义为公式（5–10）中定义的恒等损失函数。底部橙色框中的损失函数（记为 ℓ）定义见（5–6）。

不过，对于每个固定的 $\epsilon > 0$，我们都能得到一个光滑替代函数 $\Psi^{\varepsilon}_{\Theta,\gamma}$，并可通过梯度下降步更新对式（5–1）解的近似。为说明这一点，为简化记号，令 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$，然后使用初始值为 $\mathbf{X}^{(0)}$ 的初始值，通过算法4求解 $\min_{\mathbf{X}} \Psi_{\Theta,\gamma}(\mathbf{X})$。初始值 $\mathbf{X}^{(0)}$ 由预训练初始网络得到，具体说明见第5.3.1节。在算法4的第3–8行中，当平滑参数 $\varepsilon_t > 0$ 固定时，我们通过线搜索确定步长，并据此计算梯度下降更新。在第9行，算法根据约简准则更新平滑参数 ε_t。ε_t 的降低确保指定的子序列（满足 ε_t 约简准则的子序列）必有一个聚点，且该聚点是优化问题（5–1）的 Clarke驻点 [33]，如定理5.1所述。

定理5.1。 假设 $\{\mathbf{x}_t\}$ 是通过算法4 生成的序列，其初始值为$\mathbf{X}^{(0)}$, $\epsilon_{\text{tol}} = 0$ 和 $T = \infty$。令 $\{\mathbf{X}^{(t_l+1)}\}$ 为满足算法中约简条件的子序列

（算法 *4*第*9*步中的判据）。则 $\{\mathbf{X}^{(t_l+1)}\}$ 至少有一个聚点，并且 $\{\mathbf{X}^{(t_l+1)}\}$ 的每个聚点都是 $\min_{\mathbf{X}} \Psi(\mathbf{X})$的*Clarke*驻点。

我们假设$\Psi_{\Theta,\gamma}$ 是强制的，且$\Psi^*_{\Theta,\gamma} = \min_{\mathbf{X}} \Psi_{\Theta,\gamma}(\mathbf{X}) > -\infty$。对于任意集合$\mathcal{S} \subset \mathbb{R}^n$, 我们记 $\mathrm{dist}(\mathbf{y}, \mathcal{S}) := \inf\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in \mathcal{S}\}$。

引理5。 $\Psi^{\varepsilon}_{\Theta,\gamma}(\mathbf{X})$ 的梯度是*Lipschitz*连续的。

证明。 $\Psi^{\varepsilon}_{\Theta,\gamma}(\mathbf{X})$是式（1）中$\Psi_{\Theta,\gamma}(\mathbf{X})$的平滑替代函数，其中第二个求和项里的 $\|h_{w_i}(\mathbf{x}_i)\|_{2,1}$ 替换为 $\|h_{w_i}(\mathbf{x}_i)\|_{\varepsilon_{2,1}}$。由于$h_{w_i}$和$g_\theta$都以平滑激活函数参数化为标准卷积神经网络，因此很容易验证$\Psi^{\varepsilon}_{\Theta,\gamma}(\mathbf{X})$的第一项和最后一项是Lipschitz连续的。第二个求和项的Lipschitz连续性由 [17]中的引理A2证明。

□

引理6。 假设序列 $\{\mathbf{x}_t\}$ 由执行算法 *1* 的第 *3* 行生成，且固定 $\varepsilon_t = \varepsilon$ ，则

1. $\|\nabla \Psi^{\varepsilon}_{\Theta,\ \gamma}(\mathbf{x}_t)\| \to 0$ 当 $t \to \infty$ 时。*2.* 经过有限次迭代后，即可满足步骤 *4* 中用于减少 ε的条件。

引理7。 假设序列 $\{\mathbf{x}_t\}$ 由算法 *1* 在初始值 $\mathbf{X}^{(0)}$下生成，则对于任意 $t \geq 0$，有 $\Psi^{\varepsilon_{t+1}}_{\Theta,\ \gamma}(\mathbf{x}_{t+1}) + m\varepsilon_{t+1} \leq \Psi^{\varepsilon_t}_{\Theta,\ \gamma}(\mathbf{x}_t) + m\varepsilon_t$。

证明。 首先，我们很容易验证 $\|h_{w_i}(\mathbf{x}_i)\|_{\varepsilon_{2,\ 1}} \leq \|h_{w_i}(\mathbf{x}_i)\|_{2,\ 1} \leq \|h_{w_i}(\mathbf{x}_i)\|_{\varepsilon_{2,\ 1}} + m\varepsilon$。 引理6 和 7 的证明见 [33, 17]。

□

5.2.3 收敛性分析

定理1的证明概述如下。

证明。由不等式 $\|h_{w_i}(\mathbf{x}_i)\|_{2,1} \leq \|h_{w_i}(\mathbf{x}_i)\|_{\varepsilon_{2,1}} + m\varepsilon$，可得对于任意 $\varepsilon > 0$，$\Psi_{\Theta,\gamma}(\mathbf{X}) \leq \Psi^{\varepsilon}_{\Theta,\gamma}(\mathbf{X}) + m\varepsilon$。结合引理7，可得 $\Psi_{\Theta,\gamma}(\mathbf{x}_t) \leq \Psi^{\varepsilon_t}_{\Theta,\gamma}(\mathbf{x}_t) + m\varepsilon_t \leq$

$\cdots \leq \Psi^{\varepsilon_0}_{\Theta,\gamma}(\mathbf{X}(0)) + m\varepsilon_0 < \infty$。由于我们假设 $\Psi_{\Theta,\gamma}$ 是强制的，因此可得 $\{\mathrm{xt}\}$ 有界。因此，所选子序列 $\{\mathbf{X}(\mathrm{tl}+1)\}$ 也有界，且至少存在一个聚点。引理6表明，从 $\mathrm{tl}-1+1$ 到 $\mathrm{tl}+1$ 的迭代次数有限。由于 $\mathbf{X}(\mathrm{tl}+1)$ 满足算法4第4步中的约简准则，因此我们有

$\|\nabla\Psi^{\varepsilon_{t_l}}_{\Theta,\gamma}(\mathbf{X}^{(t_l+1)})\| \le \sigma\varepsilon_{t_l}\eta = \sigma\varepsilon_0\eta^{l+1} \to 0$ 当 $l \to \infty$ 时。因此，$\mathbf{X}^{(t_l+1)}$ 至少存在一个收敛子序列，记为 $\{\mathbf{X}^{(k+1)}\}$，满足 $\mathbf{X}^{(k+1)} \to \hat{\mathbf{X}}$ 当 $\varepsilon_k \to 0$ 时；并且 $\nabla\Psi^{\varepsilon_k}_{\Theta,\gamma}(\mathbf{X}^{(k+1)}) \to 0$ 当 $k \to \infty$ 时，其中 ε_k 是与$\mathbf{X}^{(k+1)}$相关联的相应 ε_{t_l}。

如前所述，为简化记号，我们将 $\mathrm{X} = \{\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3\}$。每个 x_i 的 Clark次微分几乎与 [17] 完全相同，只是在式（1）中增加了一个光滑项，因此对每个 x_i 的分析将与 [17]相同。文献 [17]已证明，当 $\mathrm{k} \to \infty$时，$\mathrm{dist}(\nabla\Psi^{\varepsilon_k}_{\Theta,\gamma}(\mathrm{x}^{(k+1)}\mathrm{i}), \partial^C\Psi_{\Theta,\gamma}(\hat{}\mathrm{x}_i)) \to 0$，其中 ∂^C表示 Clark次微分。由于这对每个 x_i都成立，因此当 $\mathrm{k} \to \infty$时，我们得到 $\mathrm{dist}(\nabla\Psi^{\varepsilon_k}_{\Theta,\gamma}(\mathrm{X}^{(k+1)}), \partial^C\Psi_{\Theta,\gamma}(\hat{}\mathrm{X})) \to 0$。由于我们已经证明 $\nabla\Psi^{\varepsilon_k}_{\Theta,\gamma}(\mathrm{X}^{(k+1)}) \to 0$，且 $\partial^C\Psi_{\Theta,\gamma}(\hat{}\mathrm{X})$ 是闭集，因此我们得出 $0 \in \partial^C\Psi_{\Theta,\gamma}(\hat{}\mathrm{X})$。

□

算法4： 具有可证明收敛性的算法展开方法

1: **输入：** 初始值 $\mathbf{X}^0$, $0 < \rho, \eta < 1$ 和 ε_0，$a, \sigma > 0$。最大总阶段数 T或容差 $\epsilon_{\mathrm{tol}} > 0$。
2: **对于** $t = 0, 1, 2, \ldots, T-1$ **执行**
3: $\mathbf{v}_{t+1} = \mathbf{x}_t - \alpha_t\nabla\Psi_{\varepsilon_t}(\mathbf{x}_t)$,
4: **如果** $\Psi_{\varepsilon_t}(\mathbf{v}_{t+1}) - \Psi_{\varepsilon_t}(\mathbf{x}_t) \le -^1{}_a\|\mathbf{v}_{t+1} - \mathbf{x}_t\|^2$成立，**则**
5: 设置 $\mathbf{x}_{t+1} = \mathbf{v}_{t+1}$,
6: **else**
7: 更新 $\alpha_t \leftarrow \rho\alpha_t$，然后**转到**3,
8: **结束如果**
9: **如果** $\|\nabla\Psi_{\varepsilon_t}(\mathbf{x}_{t+1})\| < \sigma\eta\varepsilon_t$，设置 $\varepsilon_{t+1} = \eta\varepsilon_t$；**否则**，设置 $\varepsilon_{t+1} = \varepsilon_t$。
10: **如果** $\sigma\varepsilon_t < \epsilon_{\mathrm{tol}}$，终止。
11: </style}结束循环
12: **输出：** $\mathbf{X}^{(T)}$.

该诱导梯度下降网络严格遵循算法4，网络的每个阶段都对应算法的一次迭代。我们将$\mathbf{X}^{(T)}$ 作为多阶段网络的输出，即$\mathbf{X}^{(T)} = \phi_{\hat{}\Theta}(\mathbf{X}^{(0)}, \mathbf{f}_1, \mathbf{f}_2)$，其中 ϕ表示算法诱导网络，其参数汇总为 $\hat{} \Theta = \{\Theta, \lambda, \alpha_0, \alpha_1, \ldots\alpha_{T-1}\}$。

5.2.4 用于网络训练的双层优化算法

假设我们随机抽取 $\mathcal{M}_{tr}$ 个数据对 $\{\mathcal{D}_i^{tr}\}^{\mathcal{M}_{tr}}{}_{i=1}$ 用于训练，并抽取 $\mathcal{M}_{val}$ 个数据对 $\{\mathcal{D}_i^{val}\}^{\mathcal{M}_{val}}{}_{=1}{}_i$用于验证，其中每个 $\mathcal{D}_i^{tr}$（或 $\mathcal{D}_i^{val}$）由 $\{\mathbf{f}_1^{(i)}$、$\mathbf{f}_2^{(i)}$、$\mathbf{X}^{*(i)}\}$组成；$\mathbf{f}_1^{(i)}$、$\mathbf{f}_2^{(i)}$ 表示给定的源模态部分 k 空间数据，而$\mathbf{X}^{*(i)}$ $=\{\mathbf{x}_1^{*(i)}$、$\mathbf{x}_2^{*(i)}$、$\mathbf{x}_3^{*(i)}\}$ 表示$\mathbf{x}_1^{(i)}$、$\mathbf{x}_2^{(i)}$、$\mathbf{x}_3^{(i)}$ 对应的参考图像。

为了为$\mathbf{x}_3$ （5–1）中的约束项选择一个最优系数，我们提出了一种新颖的学习框架，将网络训练构建为双层优化问题，以学习 Θ 和 λ （5–1）中的

$$\min_{\lambda} \quad \sum_{i=1}^{\mathcal{M}_{val}} \ell(\Theta(\lambda), \lambda; \mathcal{D}_i^{val}) \tag{5–5a}$$

$$\text{s.t.} \qquad \Theta(\lambda) = \arg\min_{\Theta} \sum_{i=1}^{\mathcal{M}_{tr}} \ell(\Theta, \lambda; \mathcal{D}_i^{tr}), \tag{5–5b}$$

其中

$$\ell(\Theta, \lambda; \mathcal{D}_i) := \sum_{j=1}^{3} \frac{1}{2} \|\mathbf{x}_j^T(\Theta, \lambda; \mathcal{D}_i) - \mathbf{x}_j^{*(i)}\|_2^2 + (1 - SSIM(\mathbf{x}_j^T(\Theta, \lambda; \mathcal{D}_i), \mathbf{x}_j^{*(i)})) \tag{5–6a}$$

$$+ \frac{\mu}{2} \|g_\theta([h_{w_1}(\mathbf{x}_1^{*(i)}), h_{w_2}(\mathbf{x}_2^{*(i)})]) - \mathbf{x}_3^{*(i)}\|_2^2, \tag{5–6b}$$

并且，$\mathbf{x}^{(}$ˆT) j (·) 表示第 j 个模态的 ˆT 阶段网络的输出。(5–6) 的第一项旨在最小化网络输出与真实值之间的差异。第二项是结构相似性指数 [193]。第三项用于约束 g_θ 合成 $\mathbf{x}_3$。在 (5–5) 中，下层优化在训练数据集上使用固定系数 λ 学习卷积层的网络参数 Θ，上层则在验证集上调整系数 λ，使系数λ也能泛化到验证集。在介绍双层训练算法之前，我们简化并重新定义 $\mathcal{L}(\Theta, \lambda; \mathcal{D}) := \sum^{\mathcal{M}} \mathbf{i}_{=1} \ell(\Theta, \lambda; \mathcal{D}_i)$，然后将 (5–5) 重新表述为

$$\min_{\lambda} \mathcal{L}(\Theta(\lambda), \lambda; \mathcal{D}^{val}) \qquad \text{s.t.} \quad \Theta(\lambda) = \arg\min_{\Theta} \mathcal{L}(\Theta, \lambda; \mathcal{D}^{tr}). \tag{5–7}$$

由于上述双层优化问题 (5–7) 难以求解，因此我们首先考虑按照 [125]将其放宽为单层约束优化问题。具体而言，我们使用下层部分的一阶必要条件作为约束项，如下所示

$$\min_{\lambda} \mathcal{L}(\Theta(\lambda), \lambda; \mathcal{D}^{val}) \qquad \text{s.t.} \quad \nabla_{\Theta}\mathcal{L}(\Theta, \lambda; \mathcal{D}^{tr}) = 0. \tag{5–8}$$

上述约束优化问题可以通过将约束替换为惩罚项进一步放宽，从而得到以下组合优化问题

$$\min_{\Theta,\lambda} \left\{ \widetilde{\mathcal{L}}(\Theta, \lambda; \mathcal{D}^{tr}, \mathcal{D}^{val}) := \mathcal{L}(\Theta, \lambda; \mathcal{D}^{val}) + \frac{\lambda}{2} \|\nabla_{\Theta}\mathcal{L}(\Theta, \lambda; \mathcal{D}^{tr})\|^2 \right\}. \tag{5–9}$$

由于训练和测试数据集的体积很大，无法使用全批量方法求解（5–9）。这里我们使用算法5中总结的小批量随机交替方向惩罚法来训练参数。

算法5： 小批量交替方向惩罚算法

Input $\mathcal{D}^{tr}$, $\mathcal{D}^{val}$, $\delta_{tol} > 0$.
Initialize Θ, λ, δ, $\gamma > 0$ and $\nu_{\delta} \in (0, 1)$, $\nu_{\gamma} > 1$.
while $\delta > \delta_{tol}$ **do**
  Sample training and validation batch $\mathcal{B}^{tr} \subset \mathcal{D}^{tr}, \mathcal{B}^{val} \subset \mathcal{D}^{val}$.
  while $\|\nabla_{\Theta}\widetilde{\mathcal{L}}(\Theta, \lambda; \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 + \|\nabla_{\lambda}\widetilde{\mathcal{L}}(\Theta, \lambda; \mathcal{B}^{tr}, \mathcal{B}^{val})\|^2 > \delta$ **do**
    for $k = 1, 2, \ldots, K$ (inner loop) **do**
      $\Theta \leftarrow \Theta - \rho_{\Theta}^{k} \nabla_{\Theta}\widetilde{\mathcal{L}}(\Theta, \lambda; \mathcal{B}^{tr}, \mathcal{B}^{val})$
    end for
    $\lambda \leftarrow \lambda - \rho_{\lambda} \nabla_{\lambda}\widetilde{\mathcal{L}}(\Theta, \lambda; \mathcal{B}^{tr}, \mathcal{B}^{val})$
  end while
  update $\delta \leftarrow \nu_{\delta}\delta$, $\gamma \leftarrow \nu_{\gamma}\gamma$
end while
output: Θ, λ

5.3 实验

5.3.1 初始化

初始重建 $\{\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}, \mathbf{x}_3^{(0)}\}$ 通过图5-3 所示的初始化网络获得，该网络由三个子网络组成，分别用于$\mathbf{x}_1^{(0)}$、$\mathbf{x}_2^{(0)}$、$\mathbf{x}_3^{(0)}$，然后输入

多阶段 LOA 驱动的恢复网络中，如第5.2.2节所示。初始化网络的每个模块均采用残差结构设计。

5.3.1.1 初始化网络

用于$\mathbf{x}_1^{(0)}$ 和$\mathbf{x}_2^{(0)}$ 的初始化模块结构相同，由一个k空间插值模块、一个逆傅里叶变换算子以及一个图像域残差块依次组成。k空间插值模块用于插值补全缺失的k空间数据，由4个卷积层组成，各层之间使用RELU。随后，我们将插值后的k空间数据输入逆傅里叶变换算子，得到一幅初始图像。接着，经过一个由4个卷积层组成的残差块，得到精细化的初始图像，并将其作为多阶段恢复网络的输入。

由于$\mathbf{x}_3^{(0)}$ 的k空间数据缺失，因此这里使用初始图像$\mathbf{x}_1^{(0)}$ 或$\mathbf{x}_2^{(0)}$ 作为输入，以初始化$\mathbf{x}_3^{(0)}$。我们通过一个残差初始化网络获得初始$\mathbf{x}_3^{(0)}$ 图像，该网络包含4个卷积层，层间使用ReLU。

5.3.1.2 初始化网络训练

为了训练上述初始化网络中的参数，我们使用下文定义的损失函数L^I，最小化初始化网络的输出与真实图像之间的差异

$$L^I(\mathbf{x}_j^{(0)}, \mathbf{x}_j^*) = \|\mathbf{x}_j^{(0)} - \mathbf{x}_j^*\|_1, \quad j = 1, 2, 3. \tag{5–10}$$

我们首先训练用于$\mathbf{x}_1^{(0)}$ 和$\mathbf{x}_2^{(0)}$ 的初始化网络。然后，我们将适用于可用模态$\mathbf{x}_1^{(0)}$ 或$\mathbf{x}_2^{(0)}$ 的初始化网络的输出作为输入，训练用于$\mathbf{x}_3^{(0)}$ 的初始化网络。完成全部三个初始化网络的训练后，在训练多阶段恢复网络时，我们将固定这些网络。

5.3.2 实验设置

我们使用的数据集来自 BRATS 2018 [128]挑战赛，包含对 285 名患有高级别或低级别胶质瘤的患者进行扫描的数据，涉及四种模态：T1加权成像、T2加权成像、液体衰减反转恢复和 T1CE。每位患者的数据扫描体积大小为 $240 \times 240 \times 155$。我们的

实验在高级别胶质瘤（HGG）数据集上进行。我们从 6 名患者中随机选取中心的 10 个切片作为测试数据集，裁剪后的尺寸为160 × 180，共得到 60 张测试图像；其余 HGG 数据集则划分为训练数据（1020 张图像）和验证数据（1020 张图像）。我们将其与三种最先进的多模态磁共振合成方法进行比较：多模态磁共振（MM）[25],MM-GAN [161], MMGradAdv [114]和 Hi-Net [225]。我们按照论文中给出的超参数实现了这些方法，并使用我们自己的训练数据集训练这些方法，再在我们的测试数据集上进行测试。MM、MMGradAdv 和 Hi-Net 的输入图像是源模态的真实图像，而我们的输入是使用射频掩模以40%采样率扫描得到的部分 k 空间数据。我们将所有数据集的强度值归一化至 $[0, 1]$。

5.3.3 超参数选择

我们的所有实验均在 Windows 工作站上的 Nvidia GTX-1080Ti GPU 上实现，可训练参数采用 Xavier 初始化 [60]，并使用初始学习率为 0.001 的 Adam优化器 [93] 进行训练。我们在 (5–1) 中设置 $\lambda = 1$，并在损失函数 ℓ中设置 $\mu = 0.1$。在实验中，我们使用所有复数卷积算子 [191]，其中 $\mathbf{h}_{w_i}$ 采用 4 个卷积核大小为 $3 \times 3 \times 64$ 的卷积，$\mathbf{g}_\theta$采用 6 个卷积核大小为 $3 \times 3 \times 128$ 的卷积。对于算法4，综合考虑算法收敛和计算效率，经试验后参数设置如下：$\alpha_0 = 0.01$、$\eta_0 = 0.01$、$\varepsilon_0 = 0.001$、$\mathrm{a} = 10^5$、$\sigma = 10^3$、$\rho = 0.9$ 和 $\gamma = 0.9$。我们还设置了终止容差 $\epsilon_{\text{tol}} = 1 \times 10^{-3}$。采用此容差以及第 10 行定义的终止条件，算法将在 11 个阶段后停止。若获接收，我们将提供公开共享代码。

对于算法4，综合考虑算法收敛和计算效率，经过试验后，我们采用以下参数：$\alpha_0 = 0.01$，$\eta_0 = 0.01$，ε_0 =0.001，$a = 10^5$，$\sigma = 10^3$，$\rho = 0.9$。我们还将终止容差设为 $\epsilon_{\text{tol}} = 1 \times 10^{-3}$,

结合第5行定义的终止条件，算法将在经过11个阶段后停止。

同样，在算法 5中，我们选择的参数如下： $\nu_\delta = 0.95$， $\delta = 1 \times 10^{-3}$，$\lambda = 10^{-4}$，$\nu_\lambda = 1.001$ 和 $\rho_\gamma = 0.9$。考虑到图形处理器内存和数据大小，我们将批量大小设为 2 。我们设置 $\delta_{tol} = 4$。35×10^{-6} 这使得算法在约 1000 个训练轮次后停止。

5.3.4 性能评估

我们将所提方法与多模态MRI合成方法进行了比较：MM [25],MM-GAN [161], MMGradAdv [114] 和 Hi-Net [225], ，并选取四种合成方向，包括：T1 + T2 → 液体衰减反转恢复，T1 + 液体衰减反转恢复 → T2，T2 + 液体衰减反转恢复 → T1和T1 +T 2 → T1CE。表 5-1 报告了所提方法与最先进方法的定量结果。所提模型的输入是源模态的部分 k 空间数据。 f 表示相应模态的部分 k 空间数据。显然，在所有性能指标方面，我们所提出的网络均优于MM及基于GAN的方法（MM-GAN、MMGradAdv、Hi-Net）。与基线Hi-Net相比，我们的方法平均峰值信噪比提高约0.67分贝，结构相似性指数提高约0.01，归一化均方误差降低约0.01。我们将源模态的部分 k 空间数据作为输入，同时重建源模态图像并合成目标模态图像。我们的模型首先学习算法4 的良好初始输入，然后通过迭代该算法获得更新后的重建图像和合成图像。该算法具有理论收敛保证，因此性能更佳。

我们还进行了纯合成实验（T1 + T2 → 液体衰减反转恢复），输入全扫描源数据，并且模型（5–1）仅针对 $\mathbf{x}_3$ 进行最小化，且不包含数据保真项。结果见表 5-1 ，峰值信噪比值比基线方法 Hi-Net 高约 1.16 分贝。

部分重建结果列于表5-2中，该表展示了仅重建源模态（仅最小化$\mathbf{x}_1$ 和 $\mathbf{x}_2$ 在模型（5–1）

以及联合重建源模态 $\mathbf{x}_1$, $\mathbf{x}_2$ 和合成目标模态$\mathbf{x}_3$之间的比较。表5-2显示，与不合成液体衰减反转恢复、仅重建T1加权成像和T2加权成像相比，联合重建与合成使峰值信噪比提高了0.46分贝。我们认为，这是因为合成算子g_θ 还利用数据 $\mathbf{x}_3$ 来辅助调整$\mathbf{x}_1$ 和 $\mathbf{x}_2$, 的特征图，从而提高了后续图像的重建质量。

我们的结果表明，仅对源模态进行重建而不进行合成，其性能不如使用我们的模型联合重建源模态和合成目标模态（5–1）。这表明所提方法在磁共振成像重建和合成方面均有效。

图 5-4 展示了不同源模态和目标模态图像上的合成磁共振成像结果。与其他参考方法相比，所提出的方法生成的合成图像保留了更多细节，组织边界也更加清晰锐利（红色放大窗口和绿色箭头所示）。

表 5-1. 最先进的多模态合成方法与所提方法的合成结果定性比较。

	方法	PSNR	SSIM	NMSE
T1 + T2 → 液体衰减反转恢复	多模态 [25]	22.8905 ± 1.4794	0.6671 ± 0.0586	0.0693 ± 0.0494
	MM-GAN [161]	23.3469 ± 1.0276	0.7084 ± 0.0370	0.0620 ± 0.0426
	MMGradAdv [114]	24.0275 ± 1.3959	0.7586 ± 0.0326	0.0583 ± 0.0380
	Hi-Net [225]	25.0299 ± 1.3845	0.8499 ± 0.0300	0.0254 ± 0.0097
	所提出的方法	26.1851 ± 1.3357	0.8677 ± 0.0307	0.0205 ± 0.0087
$\mathbf{f}_{T1}$ +$\mathbf{f}_{T2}$ → 液体衰减反转恢复	所提出的方法	25.7355 ± 1.2475	0.8597 ± 0.0315	0.0215 ± 0.0085
T1 + 液体衰减反转恢复 → T2加权成像	多模态 [25]	23.8916 ± 1.6094	0.6895 ± 0.0511	0.0494 ± 0.0185
	MM-GAN [161]	24.1474 ± 0.8991	0.7217 ± 0.0432	0.0431 ± 0.0114
	MMGradAdv [114]	25.0563 ± 1.4895	0.7597 ± 0.0486	0.0406 ± 0.0165
	Hi-Net [225]	25.9490 ± 1.4958	0.8552 ± 0.0410	0.0229 ± 0.0070
$\mathbf{f}_{T1}$ +$\mathbf{f}_{FLAIR}$ → T2加权成像	所提出的方法	26.6585 ± 1.6178	0.8610 ± 0.0438	0.0207 ± 0.0072
T2 + 液体衰减反转恢复 → T1加权成像	多模态 [25]	23.5347 ± 2.1764	0.7825 ± 0.0470	0.0301 ± 0.0149
	MM-GAN [161]	23.6321 ± 2.3085	0.7908 ± 0.0421	0.0293 ± 0.0119
	MMGradAdv [114]	24.7342 ± 2.2255	0.8065 ± 0.0423	0.0252 ± 0.0118
	Hi-Net [225]	25.6405 ± 1.5887	0.8729 ± 0.0349	0.0130 ± 0.0097
$\mathbf{f}_{T2}$ +$\mathbf{f}_{FLAIR}$ → T1加权成像	所提出的方法	26.3063 ± 1.7955	0.9085 ± 0.0311	0.0112 ± 0.0113
T1 + T2 → T1CE	多模态 [25]	23.3678 ± 1.5612	0.7272 ± 0.0574	0.0312 ± 0.0138
	MM-GAN [161]	23.6795 ± 0.9729	0.7577 ± 0.0637	0.0302 ± 0.0133
	MMGradAdv [114]	24.2311 ± 1.8961	0.7887 ± 0.0519	0.0273 ± 0.0136
	Hi-Net [225]	25.2149 ± 1.2048	0.8650 ± 0.0328	0.0180 ± 0.0134
$\mathbf{f}_{T1}$ +$\mathbf{f}_{T2}$ → T1CE	所提出的方法	25.9103 ± 1.2077	0.8726 ± 0.0340	0.0167 ± 0.0133

表5-2. 所提出的方法采用联合重建，以及仅对T1加权成像和T2加权成像进行合成与重建时，重建结果的定量比较。

模态	是否涉及液体衰减反转恢复?	峰值信噪比	SSIM	NMSE
T1	No	37.0040 ± 0.7411	0.9605 ± 0.0047	0.0008 ± 0.0002
	Yes	37.4967 ± 0.8361	0.9628 ± 0.0074	0.0007 ± 0.0002
T2	No	37.2401 ± 1.2252	0.9678 ± 0.0028	0.0027 ± 0.0010
	Yes	37.6675 ± 1.3427	0.9663 ± 0.0043	0.0024 ± 0.0009

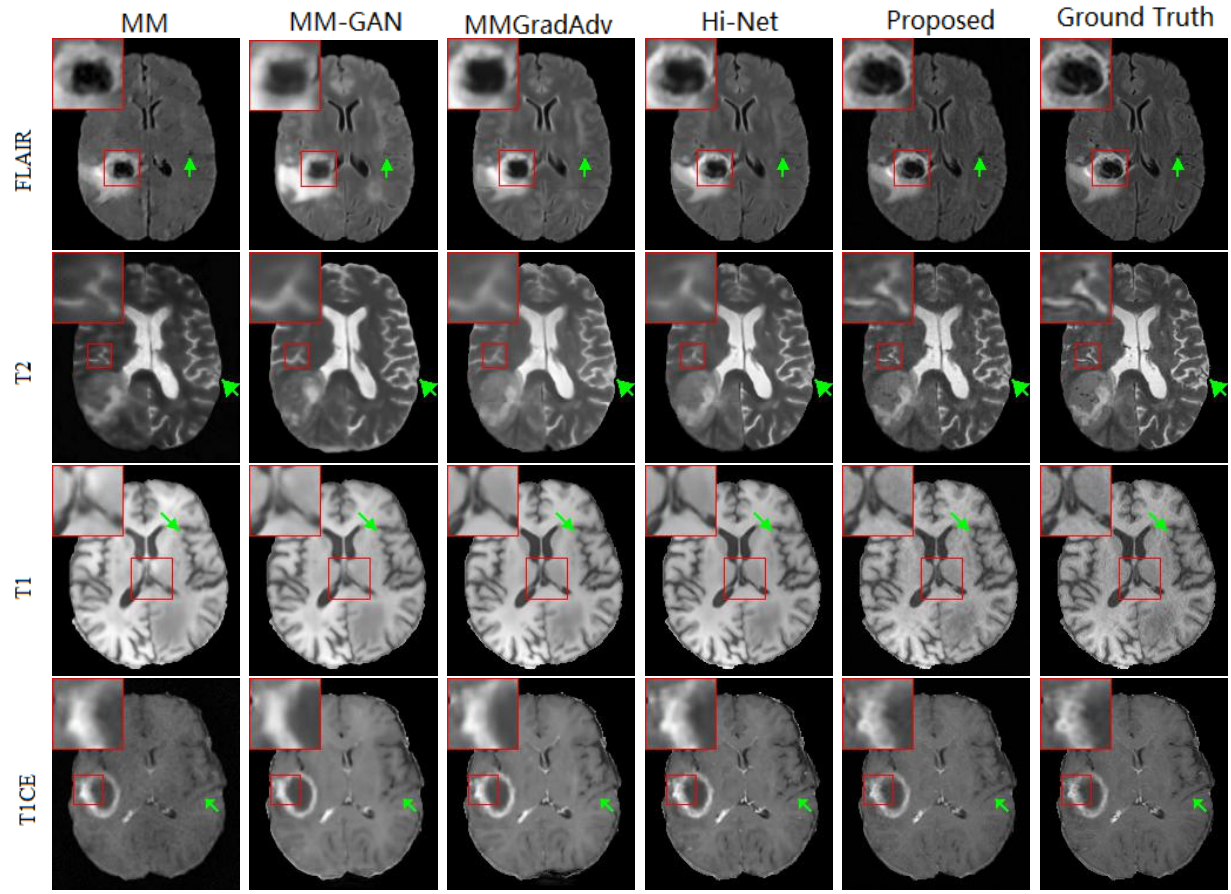

图 5-4. 最先进的多模态合成方法与所提出的方法之间的定性比较。从第一行到最后一行：T1 + T2 →液体衰减反转恢复、T1 + 液体衰减反转恢复 → T2加权成像、T2 + 液体衰减反转恢复 → T1加权成像，以及 T1 + T2 → T1CE。

5.4 结论

我们提出了一种新颖的多模态磁共振成像联合重建与合成模型。该模型通过迭代一种具有收敛保证的可学习优化算法，同时从部分扫描的k空间MR数据中重建源模态图像，并在不使用任何k空间信息的情况下合成目标模态图像。

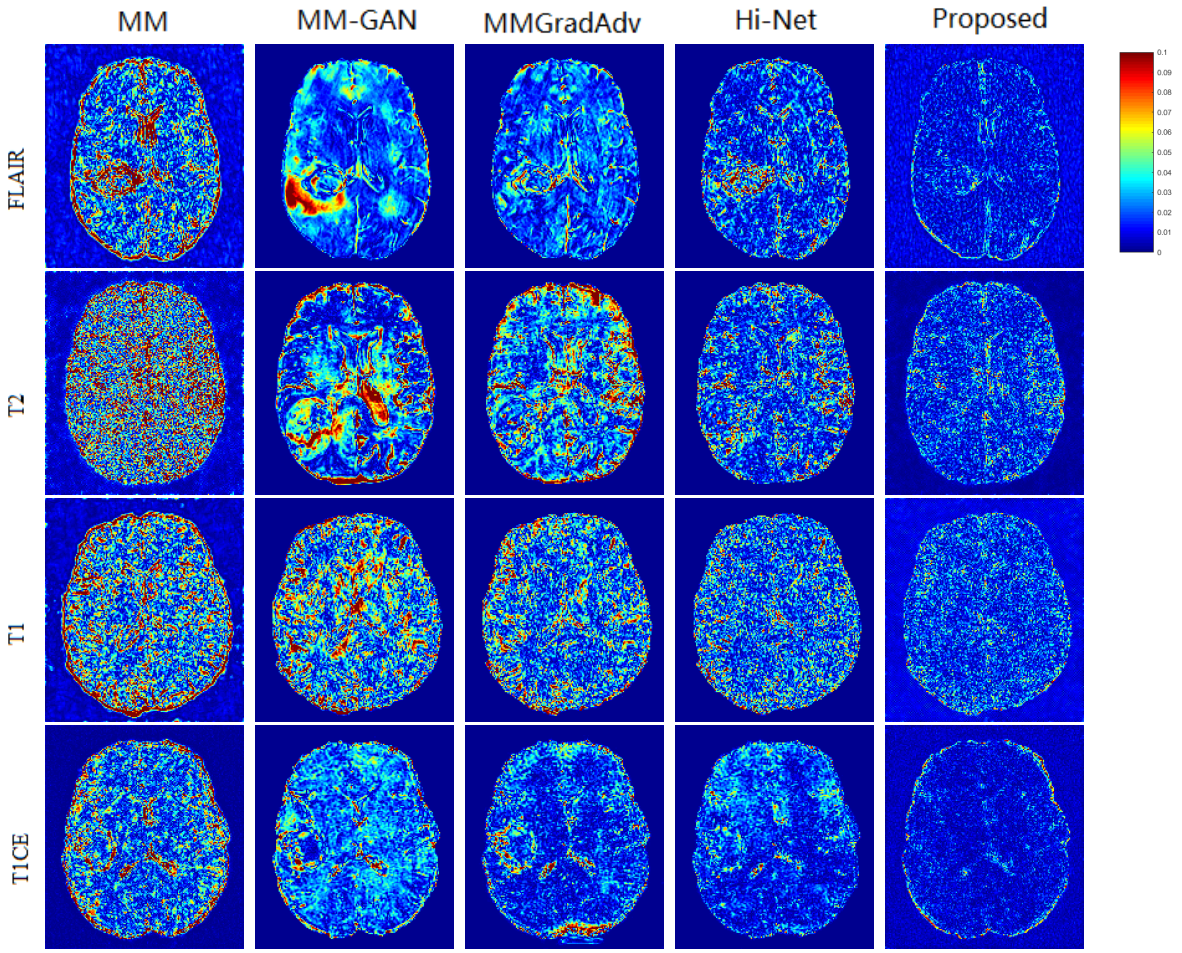

图5-5. 合成图像与其对应真实值之间的逐点误差图。从第一行到最后一行：T1 + T2 → 液体衰减反转恢复、T1 + 液体衰减反转恢复→ T2加权成像、T2 + 液体衰减反转恢复 → T1加权成像以及 T1 + T2 → T1CE。

此外，我们采用结合训练集和验证集的双层优化训练算法，进一步提升性能。在多种不同模态的脑部数据上进行的大量实验验证了所提出的模型的卓越性能。

参考文献

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z.,Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J

., Man´e, D., Monga, R., Moore,S.,Murray,D.,Olah, C., Schuster,M., Shlens, J.,Steiner, B.,Sutskever, I., Talwar,K.,Tucker,P.,Vanhoucke, V., Vasudevan, V.,Vi´egas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.：TensorFlow：面向异构系统的大规模机器学习（2015年），https://www.tensorflow.org/，软件可从 tensorflow.org 获取[2] Abadi, M. 等：TensorFlow：面向大规模机器学习的系统。载于：第 12 届{USENIX} 操作系统设计与实现研讨会（{OSDI}16），第 265–283 页（2016）.[3] Adler, J., Oktem, O.：学习型原始-对偶重建。IEEE医学影像汇刊 37(6)，1322–1332（2018年）[4] Aggarwal, H.K., Mani, M.P., Jacob, M.：MoDL：用于逆问题的基于模型的深度学习架构。IEEE医学影像汇刊 38(2)，394–405（2018年）[5] Aggarwal, H.K., Mani, M.P., Jacob, M.：MoDL：用于逆问题的基于模型的深度学习架构。IEEE医学影像汇刊 38(2)，394–405（2019年2月）[6] Ahishakiye, E., Van Gijzen, M.B., Tumwiine, J., Wario, R., Obungoloch, J.：医学图像重建中的深度学习综述。Intelligent Medicine（2021年

)

[7] Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., De Freitas, N.：通过梯度下降来学习梯度下降。在：神经信息处理系统进展。第 3981–3989 页（2016）[8] Antoniou, A., Edwards, H., Storkey, A.：如何训练你的 MAML。arXiv 预印本 arXiv:1810.09502（2018）[9] Balaji, Y., Sankaranarayanan, S., Chellappa, R.：MetaReg：使用元正则化实现领域泛化。神经信息处理系统进展

ng

系统 31，998–1008（2018年）

[10] Beck, A., Teboulle, M.：一种快速的迭代收缩阈值算法，用于线性 r
逆问题。《SIAM成像科学杂志》2(1)，183–202（2009）

[11] Ben-David, S., Blitzer, J., Crammer, K., Pereira, F., 等：领域自适应的表示分析。神经信息处理系统进展 19，137（2007年）

[12] Bengio, Y., Simard, P., Frasconi, P.：通过梯度下降学习长期依赖关系十分困难。IEEE 神经网络汇刊 5(2)，157–166（1994年）。https://doi.org/10.1109/72.279181 [13] Bernstein, M.A., Fain, S.B., Riederer, S.J.：加窗和零填充重建对 MRI 数据空间分辨率与采集策略的影响。《磁共振成像杂志》：国际磁共振医学学会官方期刊 14(3)，270–280（2001年）[14] Bian, W., Chen, Y., Ye, X.：无需线圈灵敏度的深度并行 MRI 重建网络。载于：医学图像重建机器学习国际研讨会。第 17–26 页。施普林格（2020年）[15] Bian, W., Chen, Y., Ye, X.：无需线圈灵敏度的深度并行 MRI 重建网络。载于：Deeba, F., Johnson, P., W¨urfl, T., Ye, J.C.（编）：医学图像重建中的机器学习。第 17–26 页。施普林格国际出版，Cham（2020年）[16] Bian, W., Chen, Y., Ye, X.：无需线圈灵敏度的联合通道并行 MRI 重建最优控制框架。《磁共振成像》89，1–11（2022年）[17] Bian, W., Chen, Y., Ye, X., Zhang, Q.：用于多样化数据集 MRI 重建的基于优化的元学习模型。《成像杂志》7(11)（2021年）[18] Block, K.T., Uecker, M., Frahm, J.：多线圈欠采样径向 MRI。采用全变分约束的迭代图像重建。《医学磁共振》：国际磁共振医学学会官方期刊 57(6)，1086–1098（2007年）[19] Blomgren, P., Chan, T.：彩色全变分：用于向量值图像复原的全变分方法。IEEE图像处理汇刊 7(3)，304–309（1998年）。https://doi.org/10.1109/83.661180[20] Boyd, S., Parikh, N., Chu, E.：基于交替方向乘子法的分布式优化与统计学习。Now 出版社（2011年）[21] Bui, T.D., Nguyen, M., Le, N., Luu, K.：用于非配对多对比度 MRI 图像到图像转换的基于流的变形引导。载于：医学图像计算与计算机辅助干预国际会议。第 728–737 页。施普林格（2020年）[22] Burgos, N., Cardoso, M.J., Thielemans, K., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Ahmed, R., Mahoney, C.J., Schott, J.M., 等：混合 PET-MR 扫描仪的衰减校正合成：在脑部研究中的应用。IEEE医学影像汇刊 33(12)，2332–2341（2014年）

[23] Chambolle, A., Pock, T.：一种用于凸问题的一阶原始-对偶算法及其在成像中的应用。《数学成像与视觉杂志》40(1), 120–145（2011年）[24] Chandra, S.S., Bran Lorenzana, M., Liu, X., Liu, S., Bollmann, S., Crozier, S.：深度学习在磁共振图像重建中的应用。医学影像与放射肿瘤学期刊（2021年）[25] Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A.：通过模态不变潜在表示进行多模态磁共振合成。IEEE图像处理汇刊37(3), 803–814（2017年）[26] Chen, C., Huang, J.：利用压缩感知MRI中的小波结构。《磁共振成像杂志》32(10), 1377–1389（2014年）[27] Chen, E.Z., Chen, T., Sun, S.：利用神经常微分方程进行学习优化，以重建MRI图像。载于：医学图像计算与计算机辅助干预国际会议。第83–93页。施普林格（2020年）[28] Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.：神经常微分方程。arXiv预印本 arXiv:1806.07366（2018年）[29] Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C., Huang, J.B.：深入探究少样本分类。载于：国际学习表征会议（2019年）[30] Chen, X., Lian, C., Wang, L., Deng, H., Fung, S.H., Nie, D., Thung, K.H., Yap, P.T., Gateno, J., Xia, J.J., 等：用于颅颌面骨性结构MRI分割的单样本生成对抗学习。IEEE图像处理汇刊39(3), 787–796（2019年）[31] Chen, Y., 等：用于并行磁共振成像的基于模型的卷积去混叠网络学习。载于：医学图像计算与计算机辅助干预——MICCAI 2019。第30–38页。施普林格国际出版，Cham（2019年）[32] Chen, Y., Liu, H., Ye, X., Zhang, Q.：用于非光滑非凸图像重建的可学习下降算法。《SIAM成像科学杂志》14(4), 1532–1564（2021年）。https://doi.org/10.1137/20M1353368[33] Chen, Y., Liu, H., Ye, X., Zhang, Q.：用于非光滑非凸图像重建的可学习下降算法。《SIAM成像科学杂志》14(4), 1532–1564（2021年）[34] Chen, Y., Ye, X., Zhang, Q.：用于图像重建的基于变分模型的深度神经网络。《计算机视觉与成像中的数学模型和算法手册：数学成像与视觉》，第1–29页（2021年）[35] Cheng, J., 等：模型学习：用于快速磁共振成像的原始-对偶网络。载于：医学图像计算与计算机辅助干预国际会议。第21–29页。施普林格（2019年）

[36] Chernousko, F.L., Lyubushin, A.A.：求解最优控制问题的逐次逼近法。《最优控制应用与方法》3(2), 101–114（1982年）[37] Cole, E.K., 等：用于MRI重建的深度复值卷积神经网络分析。arXiv:2004.01738（2020年）, https://arxiv.org/abs/2004.01738[38] Constantinides, C.D., Atalar, E., McVeigh, E.R.：核磁共振相控阵幅度图像中的信噪比测量。《医学磁共振》38(5), 852–857（1997年）。https://doi.org/10.1002/mrm.1910380524[39] Cordier, N., Delingette, H., Lˆe, M., Ayache, N.：扩展模态传播：病理病例的图像合成。IEEE图像处理汇刊35(12), 2598–2608（2016年）[40] Dai, Y., Zhuang, P.：通过多尺度空洞残差卷积网络实现压缩感知MRI。《磁共振成像杂志》63, 93–104（2019年）[41] Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cˌukur, T.：利用条件生成对抗网络进行多对比度MRI图像合成。IEEE图像处理汇刊38(10), 2375–2388（2019年）[42] Dar, S.U., Yurt, M., Shahdloo, M., Ildız, M.E., Tınaz, B., Cˌukur, T.：通过生成对抗网络进行先验引导的加速多对比度MRI图像重建。IEEE 信号处理精选主题期刊14(6), 1072–1087（2020年）[43] Day, O., Khoshgoftaar, T.M.：异构迁移学习综述。《大数据期刊》4(1), 1–42（2017年）[44] Dedmari, M.A., 等：用于磁共振图像重建的复值全卷积神经网络。载于：医学图像重建机器学习国际研讨会。第30–38页。施普林格（2018年）[45] Deshmane, A., 等：并行磁共振成像。《磁共振成像杂志》36(1), 55–72（2012年）[46] Dong, W., Shi, G., Li, X., Ma, Y., Huang, F.：通过非局部低秩正则化实现压缩感知。IEEE图像处理汇刊23(8), 3618–3632（2014年）[47] Donoho, D.L.：压缩感知。IEEE信息论汇刊52(4), 1289–1306（2006年）[48] Duan, J., 等：VS-Net：用于加速并行MRI重建的变量分裂网络。载于：医学图像计算与计算机辅助干预——MICCAI 2019。第713–722页。施普林格国际出版，Cham（2019年）

[49] Dumoulin, V., Visin, F.：深度学习卷积运算指南。arXi v

预印本 arXiv:1603.07285（2016）[50] Eksioglu, E.M.：使用非局部块匹配模型进行磁共振成像重建的解耦算法：BM3D-MRI。《数学成像与视觉杂志》56(3)，430–440（2016）[51] Eo, T., 等：KIKI-net：用于重建欠采样磁共振图像的跨域卷积神经网络。《医学磁共振》80(5)，2188–2201（2018年）。https://doi.org/10.1002/mrm.27201[52] Esser, E., Zhang, X., Chan, T.F.：面向成像科学中凸优化的一类一阶原始-对偶算法的通用框架。SIAM Journal o

n

成像科学 3(4)，1015–1046（2010）[53] Fan, J., Han, F., Liu, H.：大数据分析的挑战。《国家科学评论》1(2)，293–314（2014年）[54] Finn, C., Abbeel, P., Levine, S.：用于快速适应深度网络的模型无关元学习。载于：国际机器学习会议。第1126–113

5.

PMLR（2017年）[55] Finn, C., Rajeswaran, A., Kakade, S., Levine, S.：在线元学习。载于：国际机器学习会议。第1920–1930页。PMLR（2019年）[56] Finn, C., Xu, K., Levine, S.：概率模型无关元学习。arXiv预印本 arXiv:1806.02817（2018年）[57] Franceschi, L., Donini, M., Frasconi, P., Pontil, M.：基于梯度的超参数优化的正向与反向方法。载于：国际机器学习会议。第1165–1173页。PMLR（2017年）[58] Franceschi, L., Frasconi, P., Salzo, S., Grazzi, R., Pontil, M.：用于超参数优化和元学习的双层规划。载于：国际机器学习会议。

n

机器学习。第1568–1577页。PMLR（2018年）[59] Glorot, X., Bengio, Y.：理解深度前馈神经网络训练的难点。载于：Teh, Y.W., Titterington, M.（编）《第十三届国际人工智能与统计学会议论文集》。《机器学习研究论文集》，第9卷，第249–256页。PMLR，奇亚拉贡

a

撒丁岛，意大利（2010年5月13日至15日）[60] Glorot, X., Bengio, Y.：理解深度前馈神经网络训练的困难。在：第十三届人工智能与统计学国际会议论文集。第249–256页（2010年）[61] Glorot, X., Bengio, Y.：理解深度前馈神经网络训练的困难。在：人工智能与统计学国际会议论文集。人工智能与统计学会（2010年）

[62] Goldstein, T., Osher, S.：l1 正则化问题的分裂 Bregman 方法。《SIAM成像科学杂志》**2**(2)，323–343（2009年）[63] Grant, E., Finn, C., Levine, S., Darrell, T., Griffiths, T.：将基于梯度的元学习重新表述为层次贝叶斯。arXiv 预印本 arXiv:1801.08930（2018年）[64] Graves, A., Wayne, G., Danihelka, I.：神经图灵机。arXiv 预印本 arXiv:1410.5401（2014年）[65] Griswold, M.A. 等：广义自动校准部分并行采集（GRAPPA）。《医学磁共振》：国际磁共振医学学会官方期刊**47**(6)，1202–1210（2002年）[66] Guerquin-Kern, M., Haberlin, M., Pruessmann, K.P., Unser, M.：一种快速的基于小波的磁共振成像重建方法。《IEEE医学影像汇刊》**30**(9)，1649–1660（2011年）[67] Haldar, J.P., Hernando, D., Liang, Z.P.：采用随机编码的压缩感知 MRI。《IEEE医学影像汇刊》**30**(4)，893–903（2010年）[68] Halkin, H.：由非线性差分方程描述的系统的庞特里亚金型极大值原理。《SIAM控制杂志》**4**(1)，90–111（1966年）。https://doi.org/10.1137/0304009[69] Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.：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年）[70] Hammernik, K., Knoll, F., Sodickson, D.K., Pock, T.：L2 还是非 L2：损失函数设计对深度学习 MRI 重建的影响。在：ISMRM 第 25 届年会。第0687页（2017年）[71] Hammernik, K., Schlemper, J., Qin, C., Duan, J., Summers, R.M., Rueckert, D.：采用灵敏度加权线圈组合对快速并行 MRI 重建的迭代深度神经网络进行系统评估。《医学磁共振》（2021年）[72] Hammernik, K. 等：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年）[73] He, K., Zhang, X., Ren, S., Sun, J.：用于图像识别的深度残差学习。在：2016 IEEE计算机视觉与模式识别会议（CVPR）。第770–778页（2016年）。https://doi.org/10.1109/CVPR.2016.90[74] He, K., Zhang, X., Ren, S., Sun, J.：用于图像识别的深度残差学习。在：IEEE计算机视觉与模式识别会议论文集。第770–778页（2016年）

[75] Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., 等：Flexisp：灵活的相机图像处理框架。ACM图形学汇刊（ToG）33(6)，1–13（2014年）[76] Hochreiter, S.：循环神经网络学习过程中的梯度消失问题及解决方案。《国际不确定性、模糊性与知识型系统杂志》06(02)，107–116（1998年）。https://doi.org/10.1142/S0218488598000094[77] Hore, A., Ziou, D.：图像质量指标：峰值信噪比与结构相似性指数之比较。载于：2010年第20届国际模式识别会议。第2366–2369页。IEEE（2010年）[78] Hospedales, T.M., Antoniou, A., Micaelli, P., Storkey, A.J.：神经网络中的元学习：综述。IEEE 模式分析与机器智能汇刊（2021年）[79] Hosseini, S.A.H., Yaman, B., Moeller, S., Hong, M., Aķcakaya, M.：用于加速磁共振成像的稠密循环神经网络：优化算法的历史感知展开。IEEE 信号处理精选主题期刊14(6)，1280–1291（2020年）[80] Huang, F.：使用神经网络进行医学成像（2020年12月24日），美国专利申请 16/904,981[81] Huang, F., Chen, M.：磁共振成像方法和设备（2020年12月1日），美国专利 10,852,376[82] Huang, F., Han, D., Mei, L.：利用深度中性网络进行磁共振成像（2020年7月9日），美国专利申请 16/735,874[83] Huang, J., Chen, C., Axel, L.：快速多对比度磁共振成像重建。磁共振成像32(10)，1344–1352（2014年）[84] Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., Zhang, X.P.：用于压缩感知MRI的贝叶斯非参数字典学习。IEEE图像处理汇刊23(12)，5007–5019（2014年）[85] Huisman, M., van Rijn, J.N., Plaat, A.：深度元学习综述。《人工智能评论》第1–59页（2021年）[86] Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A.：Synseg-net：无需目标模态真实标注的合成分割。IEEE医学影像汇刊38(4)，1016–1025（2018年）[87] Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B.：合成磁共振成像对比度对跨模态分析有用吗？载于：医学图像计算与计算机辅助干预国际会议。第631–638页。施普林格（2013年）

[88] Islam, S.R., Maity, S.P., Ray, A.K.：通过深度学习实现压缩感知正则化的无校准并行磁共振成像。生物医学信号处理与控制**66**，102399（2021年）[89] Jenni, S., Favaro, P.：深度双层学习。载于：欧洲计算机视觉会议（ECCV）论文集。第618–633页（2018年）[90] Jiang, J., 等：用于无监督多域MRI腹部器官分割的统一跨模态特征解耦器。载于：医学图像计算与计算机辅助干预 – MICCAI 2020。第347–358页。施普林格国际出版，Cham（2020年）[91] Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L.：通过邻域集成上的对比学习进行磁共振图像合成。医学图像分析**24(1)**，63–76（2015年）[92] Jog, A., Roy, S., Carass, A., Prince, J.L.：通过图块回归进行磁共振图像合成。载于：2013年IEEE第十届生物医学成像国际研讨会。第350–353页。IEEE（2013年）[93] Kingma, D.P., Ba, J.：Adam：一种随机优化方法。arXiv预印本 arXiv:1412.6980（2014年）[94] Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D.K., Akcakaya, M.：用于并行磁共振成像重建的深度学习方法：当前方法、趋势与问题综述。IEEE信号处理杂志**37(1)**，128–140（2020年）[95] Koch, G., Zemel, R., Salakhutdinov, R., 等：用于单样本图像识别的孪生神经网络。载于：ICML深度学习研讨会。第2卷。里尔（2015年）[96] Kwon, K., Kim, D., Park, H.：一种使用多层感知机的并行磁共振成像方法。《医学物理学》**44(12)**，6209–6224（2017年）[97] Larkman, D.J., Nunes, R.G.：并行磁共振成像。医学与生物学物理**52(7)**，R15（2007年）[98] LeCun, Y., Bengio, Y., Hinton, G.：深度学习。《自然》**521(7553)**，436–444（2015年）[99] Lee, D., Yoo, J., Tak, S., Ye, J.C.：利用幅度与相位网络进行加速磁共振成像的深度残差学习。IEEE生物医学工程汇刊**65(9)**，1985–1995（2018年）[100] Lee, D., Yoo, J., Ye, J.C.：用于压缩感知MRI的深度残差学习。载于：2017年IEEE第十四届生物医学成像国际研讨会（ISBI 2017）。第15–18页。IEEE（2017年）

[101] Lee, Y., Choi, S.：基于梯度的元学习：学习逐层度量与子空间。载于：国际机器学习会议。第2927–2936页。PMLR（2018年）[102] Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.：学习实现泛化：用于领域泛化的元学习。载于：第三十二届AAAI人工智能会议（2018年）[103] Li, E., Khalvati, F., Shafiee, M.J., Haider, M.A., Wong, A.：使用跨域随机全连接条件随机场进行压缩感知磁共振成像的稀疏重建。BMC医学影像 16(1)，1–12（2016年）[104] Li, K., Malik, J.：学习优化。arXiv预印本 arXiv:1606.01885（2016年）[105] Li, Q.：动力系统与机器学习。北京大学暑期学校（2020年）[106] Li, Q., Chen, L., Tai, C., 等：基于极大值原理的深度学习算法。arXiv预印本 arXiv:1710.09513（2017年）[107] Li, Q., Hao, S.：深度学习的最优控制方法及其在离散权重神经网络中的应用。载于：Dy, J., Krause, A.（编），第35届国际机器学习会议论文集。《机器学习研究论文集》，第80卷，第2985–2994页。PMLR（2018年7月10–15日）[108] Li, Q., Lin, T., Shen, Z.：通过动力系统实现深度学习：一种近似视角。arXiv预印本 arXiv:1912.10382（2019年）[109] Li, Y., Yang, Y., Zhou, W., Hospedales, T.：用于异质领域泛化的特征评论网络。载于：国际机器学习会议。第3915–3924页。PMLR（2019年）[110] Li, Z., Zhou, F., Chen, F., Li, H.：Meta-SGD：面向小样本学习的快速学习方法。arXiv预印本 arXiv:1707.09835（2017年）[111] Liang, D., Cheng, J., Ke, Z., Ying, L.：深度磁共振图像重建：逆问题与神经网络的交汇。IEEE信号处理杂志37(1)，141–151（2020年）[112] Lions, P.L., Mercier, B.：两个非线性算子之和的分裂算法。SIAM数值分析杂志16(6)，964–979（1979年）[113] Liu, R., Zhang, Y., Cheng, S., Luo, Z., Fan, X.：一种集成原理模块的压缩感知磁共振成像深度框架：展开视角、收敛行为与实用建模。IEEE医学影像汇刊39(12)，4150–4163（2020年）

[114] Liu, X., Yu, A., Wei, X., Pan, Z., Tang, J.：利用梯度先验和对抗学习进行多模态磁共振图像合成。IEEE 信号处理精选主题期刊14(6)，1176–1188（2020年）[115] Lu, T., Zhang, X., Huang, Y., Guo, D., Huang, F., Xu, Q., Hu, Y., Ou-Yang, L., Lin, J., Yan, Z., 等：用于并行磁共振成像重建的PFISTA-SENSE-ResNet。期刊

of

《医学磁共振》318，106790（2020年）[116] Lundervold, A.S., Lundervold, A.: 医学成像中深度学习的概述

g

聚焦于磁共振成像。《医学物理学杂志》29(2)，102–127（2019年）[117] Lustig, M., Donoho, D., Pauly, J.M.: 稀疏磁共振成像：压缩感知在快速磁共振成像中的应用。《医学磁共振：国际磁共振医学学会官方期刊》

al

国际磁共振医学学会 58(6)，1182–1195（2007年）[118] Lustig, M., Pauly, J. M.: SPIRiT：从任意k空间进行迭代自洽并行成像重建。《医学磁共振》64(2)，457–471（2010年）[119] Lustig, M., Pauly, J.M.: SPIRiT：从任意k空间进行迭代自洽并行成像重建。《医学磁共振》64(2)，457–471（2010年）。 https://doi.org/10.1002/mrm.22428[120] Lv, J., Wang, C., Yang, G.: PIC-GAN：一种用于加速多通道磁共振成像重建的并行成像耦合生成对抗网络。《诊断学》11(1)，61（2021年）[121] Lyra-Leite, D.M., da Costa, J.P.C.L., de Carvalho, J.L.A.: 使用基于奇异值分解的低秩近似改进磁共振成像重建和去噪。载于：2012年工程应用研讨会。第1–6页。IEEE（2012年）[122] Lysaker, M., Lundervold, A., Tai, X.C.: 使用四阶偏微分方程去除噪声，并应用于时空医学磁共振图像。《IEEE图像处理汇刊》12(12)，1579–1590（2003年）[123] Mardani, M., et al.: 用于压缩感知磁共振成像的深度生成对抗神经网络。《IEEE医学影像汇刊》38(1)，167–179（2019年1月）[124] McCann, M.T., Jin, K.H., Unser, M.: 成像逆问题中的卷积神经网络：综述。《IEEE信号处理杂志》34(6)，85–95（2017年）[125] Mehra, A., Hamm, J.: 无求逆深度双层优化的惩罚方法

.

arXiv预印本 arXiv:1911.03432（2019年）[126] Meinhardt, T. 等：学习近端算子：使用去噪网络对逆成像问题进行正则化。载于：IEEE国际计算机视觉会议论文集。第1781–1790页（2017年）

[127] Meng, N. 等：一种用于联合图像与灵敏度估计的先验学习网络 n
用于并行磁共振成像。载于：医学图像计算与计算机辅助干预——MICCAI 2019。第732–740页。施普林格国际出版，Cha m
(2019)[128] Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., 等：多模态脑肿瘤图像分割基准（BRATS）。IEEE医学影像汇刊34(10), 1993–2024 (2014)[129] Micaelli, P., Storkey, A.：长时域上的非贪心梯度超参数优化。arXiv预印本 arXiv:2007.07869 (2020)[130] Miller, M.I., Christensen, G.E., Amit, Y., Grenander, U.：可变形神经解剖结构数学教材。《美国国家科学院院刊》90(24), 11944–11948 (1993)[131] Mishra, N., Rohaninejad, M., Chen, X., Abbeel, P.：一种简单的神经注意元学习器。arXiv预印本 arXiv:1707.03141 (2017)[132] Monga, V., Li, Y., Eldar, Y.C.：算法展开：面向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)[133] Munkhdalai, T., Yu, H.：元网络。载于：国际机器学习会议。第2554–2563页。PMLR (2017)[134] Nichol, A., Achiam, J., Schulman, J.：关于一阶元学习算法。arXi
v
预印本 arXiv:1803.02999（2018年）[135] Nichol, A., Schulman, J.：Reptile：一种可扩展的元学习算法。arXiv 预印本 t
arXiv:1803.02999 2(3), 4（2018年）[136] Nitski, O. 等：CDF-Net：用于加速磁共振成像重建的跨域融合网络。载于：医学图像计算与计算机辅助干预 –
MICCAI 2020。第421–430页。施普林格国际出版，Cham（2020年）[137] Nyquist, H.：电报传输理论中的若干专题。美国电气工程师学会汇刊 47(2), 617–644（1928年）[138] Parikh, N., Boyd, S.：近端算法。优化领域的基础与趋势 n
1(3), 127–239 (2014)[139] Pedregosa，F.：使用近似梯度进行超参数优化。载于：国际机器学习会议。第737–746页。PMLR（2016年）[140] Pezzotti，N.等：一种用于欠采样膝部MRI重建的自适应智能算法：应用于2019年fastMRI挑战赛。arXiv:2004.07339（2020年）

[141] Poli, A., Cirillo, M.：关于使用归一化均方误差评估扩散模型性能。大气环境：A辑，综合主题**27**，2427–2434（1993年10月）。https://doi.org/10.1016/0960-1686(93)90410-Z[142] Pruessmann, K.P.等：SENSE：用于快速磁共振成像的灵敏度编码。《医学磁共振》：国际磁共振医学学会官方期刊**42(5)**，952–962（1999年）[143] Qiao, S., Liu, C., Shen, W., Yuille, A.L.：通过预测激活值中的参数进行少样本图像识别。载于：IEEE计算机视觉与模式识别会议论文集。第7229–7238页（2018年）[144] Qu, X., Hou, Y., Lam, F., Guo, D., Zhong, J., Chen, Z.：使用基于图像块的非局部算子，从欠采样测量值重建磁共振图像。医学图像分析**18(6)**，843–856（2014年）[145] Quan, T.M., Nguyen-Duc, T., Jeong, W.K.：使用带循环损失的生成对抗网络进行压缩感知MRI重建。IEEE医学影像汇刊**37**，1488–1497（2018年）[146] Quan, T.M., Nguyen-Duc, T., Jeong, W.K.：使用带循环损失的生成对抗网络进行压缩感知MRI重建。IEEE医学影像汇刊**37(6)**，1488–1497（2018年）[147] Quan, Y., Chen, Y., Shao, Y., Teng, H., Xu, Y., Ji, H.：使用复值深度卷积神经网络进行图像去噪。模式识别**111**，107639（2021年）[148] Qui˜nonero-Candela, J., Sugiyama, M., Lawrence, N.D., Schwaighofer, A.：机器学习中的数据集偏移。MIT出版社（2009年）[149] Rajeswaran, A., Finn, C., Kakade, S.M., Levine, S.：使用隐式梯度进行元学习。神经信息处理系统进展**32** （2019年）[150] Ravi, S., Larochelle, H.：将优化视为小样本学习的模型。国际学习表征会议（2016年）[151] Ravishankar, S., Bresler, Y.：通过字典学习，从高度欠采样的k空间数据重建磁共振图像。IEEE医学影像汇刊**30(5)**，1028–1041（2010年）[152] Rebuffi, S.A., Bilen, H., Vedaldi, A.：使用残差适配器学习多个视觉域。载于：神经信息处理系统进展（2017年）[153] Roy, S., Carass, A., Prince, J.L.：基于示例的磁共振图像对比度合成。IEEE医学影像汇刊**32(12)**，2348–2363（2013年）[154] Roy, S., Chou, Y.Y., Jog, A., Butman, J.A., Pham, D.L.：全头部磁共振图像的基于图像块的合成：应用于EPI畸变校正。载于：医学影像仿真与合成国际研讨会。第146–156页。施普林格（2016年）

[155] Rudin, L.I., Osher, S., Fatemi, E.：基于非线性全变分的噪声去除算法。《Physica D：非线性现象》60(1-4), 259–268 (1992)[156] Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R.：通过潜在嵌入优化进行元学习。载于：国际学习表征会议 (2018)[157] Sandino, C.M., 等：压缩感知：从研究走向临床实践——利用深度神经网络缩短磁共振成像扫描时间。IEEE信号处理杂志 37(1), 117–127 (2020)[158] Scardapane, S., 等：采用非参数激活函数的复值神经网络。IEEE计算智能新兴主题汇刊 4(2), 140–150 (2020)[159] Schlemper, J., 等：用于动态磁共振图像重建的深度卷积神经网络级联。IEEE医学影像汇刊 37(2), 491–503 (2018)[160] Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.：用于动态磁共振图像重建的深度卷积神经网络级联。IEEE医学影像汇刊 37(2), 491–503 (2017)[161] Sharma, A., Hamarneh, G.：使用多模态生成对抗网络合成缺失的磁共振成像脉冲序列。IEEE医学影像汇刊 39(4), 1170–1183 (2019)[162] Singha, A., Thakur, R.S., Patel, T.：深度学习在医学图像分析中的应用。《用于信息检索的生物医学数据挖掘：方法、技术与应用》，第293–350页 (2021)[163] Snell, J., Swersky, K., Zemel, R.S.：用于小样本学习的原型网络。arXiv预印本 arXiv:1703.05175 (2017)[164] Sodickson, D.K., Manning, W.J.：空间谐波同时采集（SMASH）：利用射频线圈阵列实现快速成像。《医学磁共振》38(4), 591–603 (1997)[165] Sohail, M., Riaz, M.N., Wu, J., Long, C., Li, S.：使用生成对抗网络进行非配对多对比度磁共振图像合成。载于：医学影像仿真与合成国际研讨会，第22–31页。施普林格 (2019)[166] Souza, R., Frayne, R.：用于磁共振图像重建的混合频域/图像域深度网络。载于：2019年第32届SIBGRAPI图形、模式与图像会议（SIBGRAPI），第257–264页。IEEE (2019)[167] Souza, R., Lebel, R.M., Frayne, R.：用于磁共振图像重建的混合双域卷积神经网络级联。载于：医学影像深度学习国际会议，第437–446页。PMLR (2019)

[168] Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C.L., Yakubova, N., Knoll, F., Johnson, P.：用于加速磁共振成像重建的端到端变分网络。载于：医学图像计算与计算机辅助干预国际会议，第64–73页。施普林格 (2020)[169] Sriram, A., 等：用于加速磁共振成像重建的端到端变分网络。载于：医学图像计算与计算机辅助干预——MICCAI 2020，第64–73页。施普林格国际出版，Cham (2020)[170] Sriram, A., 等：GrappaNet：结合并行成像与深度学习进行多线圈MRI重建。载于：IEEE/CVF计算机视觉与模式识别会议论文集，第14315–14322页 (2020)[171] Sudarshan, V.P., 等：正电子发射断层成像 + 并行磁共振成像的联合重建：贝叶斯耦合字典MRF框架。载于：医学图像计算与计算机辅助干预——MICCAI 2019，第39–47页。施普林格国际出版，Cham (2019)[172] Sun, J., Li, H., Xu, Z., 等：用于压缩感知磁共振成像的Deep ADMM-Net。《神经信息处理系统进展》**29** (2016)[173] Tavaf, N., Torfi, A., Ugurbil, K., Van de Moortele, P.F.：用于并行磁共振成像重建的GRAPPA-GANs。arXiv预印本 arXiv:2101.03135 (2021)[174] Thrun, S., Pratt, L.：学会学习：导论与概述。载于：《学会学习》，第3–17页。施普林格 (1998)[175] Torrado-Carvajal, A., Herraiz, J.L., Alcain, E., Montemayor, A.S., Garcia-Canamaque, L., Hernandez-Tamames, J.A., Rozenholc, Y., Malpica, N.：用于脑部研究中PET/MR衰减校正的T1加权MR图像快速基于图像块的伪CT合成。《核医学杂志》**57**(1),136–143(2016)[176] Trabelsi, C., 等：深度复数网络。arXiv:1705.09792 (2017)，https://arxiv.org/abs/1705.09792[177] Triantafillou, E., Zhu, T., Dumoulin, V., Lamblin, P., Evci, U., Xu, K., Goroshin, R., Gelada, C., Swersky, K.J., Manzagol, P.A., Larochelle, H.：Meta-Dataset：用于从少量示例中学习的多个数据集构成的数据集。载于：国际学习表征会议 (2020)[178] Trzasko, J., Manduca, A., Borisch, E.：动态磁共振成像序列重建中的局部与全局低秩促进。载于：国际磁共振医学研讨会论文集，第19卷，第4371页 (2011)[179] Uecker, M., 等：用于压缩感知和并行成像的软件工具箱与编程库。载于：ISMRM数据采样与图像重建研讨会，第41页。Citeseer (2013)

[180] Uecker, M., et al.：Espirit——一种用于自校准并行磁共振成像的特征值方法：SENSE与GRAPPA的交汇。《医学磁共振》**71**(3), 990–1001 (2014年)[181] Valkonen, T.：一种用于非线性算子的原始-对偶混合梯度方法及其在磁共振成像中的应用。逆问题 **30**(5), 055012 (2014年)[182] Van Tulder, G., de Bruijne, M.：为什么合成数据能改善多序列分类？载于：医学图像计算与计算机辅助干预国际会议。pp. 531–538。施普林格 (2015年)[183] Vasudeva, B., Deora, P., Bhattacharya, S., Pradhan, P.M.：Co-VeGAN：用于压缩感知磁共振图像重建的复值生成对抗网络 (2020年)[184] Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.：用于单样本学习的匹配网络。神经信息处理系统进展 **29**, 3630–3638 (2016年)[185] Virtue, P., Yu, S.X., Lustig, M.：优于真实：用于磁共振指纹成像的复值神经网络。载于：2017年IEEE国际图像处理会议 (ICIP)。pp. 3953–3957 (2017年)[186] Vuorio, R., Sun, S.H., Hu, H., Lim, J.J.：通过任务感知调制实现多模态模型无关元学习。arXiv预印本 arXiv:1910.13616 (2019年)[187] Walsh, D.O., Gmitro, A.F., Marcellin, M.W.：相控阵磁共振图像的自适应重建。《医学磁共振》 **43**(5), 682–690 (2000). https://doi.org/10.1002/(SICI)1522-2594(200005)43:5¡682::AID-MRM10¿3.0.CO;2-G [188] Walsh, D., Gmitro, A., Marcellin, M.：相控阵磁共振图像的自适应重建。《医学磁共振》 **43**(5), 682–690 (2000年)[189] Wang, S., Su, Z., Ying, L., Peng, X., Zhu, S., Liang, F., Feng, D., Liang, D.：通过深度学习加速磁共振成像。载于：2016年第13届IEEE生物医学成像国际研讨会 (ISBI)。pp. 514–517 (2016年4月)[190] Wang, S., Tan, S., Gao, Y., Liu, Q., Ying, L., Xiao, T., Liu, Y., Liu, X., Zheng, H., Liang, D.：学习联合稀疏编码用于无校准并行磁共振成像。IEEE医学影像汇刊 **37**(1), 251–261 (2017年)[191] Wang, S., et al.：DeepcomplexMRI：利用具有复数卷积的深度残差网络实现快速并行磁共振成像。磁共振成像 **68**, 136 – 147 (2020年)[192] Wang, Z., et al.：IKWI-net：一种用于欠采样磁共振图像重建的跨域卷积神经网络。磁共振成像 **73**, 1–10 (2020年)

[193] Wang, Z., et al.：图像质量评估：从误差可见性到结构相似性。IEEE图像处理汇刊 13(4), 600–612 (2004年)[194] Weinan, E.：关于通过动力系统进行机器学习的建议。数学与统计学通讯 5(1), 1–11 (2017年)[195] Welander, P., Karlsson, S., Eklund, A.：用于多对比度磁共振图像图像到图像转换的生成对抗网络——CycleGAN与UNIT的比较。arXiv预印本 arXiv:1806.07777 (2018年)[196] Wichrowska, O., Maheswaranathan, N., Hoffman, M.W., Colmenarejo, S.G., Denil, M., Freitas, N., Sohl-Dickstein, J.：可扩展且可泛化的学习型优化器。载于：国际机器学习会议。pp. 3751–3760。PMLR (2017年)[197] Wood, M.L., Mark Henkelman, R.：磁共振成像中的截断伪影。《医学磁共振》 2(6), 517–526 (1985年)[198] Wu, C., Tai, X.C.：用于ROF、向量总变分和高阶模型的增广拉格朗日方法、对偶方法与分裂布雷格曼迭代。《SIAM成像科学杂志》 3(3), 300–339 (2010年)[199] Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., et al.：DAGAN：用于快速压缩感知磁共振成像重建的深度去混叠生成对抗网络。IEEE医学影像汇刊 37(6), 1310–1321 (2017年)[200] Yang, J., Zhang, Y., Yin, W.：一种用于从部分傅里叶数据进行TVL1-L2信号重建的快速交替方向法。IEEE 信号处理精选主题期刊 4(2), 288–297 (2010年)[201] Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.C., Xu, Y., et al.：用于跨模态图像配准和分割的磁共振图像到图像转换。arXiv预印本 arXiv:1801.06940 (2018年)[202] Yang, Y., et al.：ADMM-CSNet：一种用于图像压缩感知的深度学习方法。IEEE 模式分析与机器智能汇刊 42(3), 521–538 (2020 年 3 月)[203] Yang, Y., Sun, J., Li, H., Xu, Z.：ADMM-CSNet：一种用于图像压缩感知的深度学习方法。IEEE 模式分析与机器智能汇刊42(3), 521–538 (2018年)[204] Yang, Y., Wang, N., Yang, H., Sun, J., Xu, Z.：由跨对比度磁共振图像引导、用于超快速压缩感知磁共振成像的模型驱动的深度注意力网络。载于：医学图像计算与计算机辅助干预国际会议。pp. 188–198。施普林格 (2020年)

[205] Yang, Y. 等：用于压缩感知磁共振成像的 Deep ADMM-Net。载于：Lee, D.D.、Sugiyama, M.、Luxburg, U.V.、Guyon, I.、Garnett, R.（编），《神经信息处理系统进展》第29卷，第10–18页。Curran Associates, Inc.（2016年）[206] Yang, Z.、Jacob, M.：逆问题的非局部正则化：统一的变分框架。《IEEE图像处理汇刊》**22**（8），3192–3203（2012年）[207] Yao, H.、Huang, L.K.、Zhang, L.、Wei, Y.、Tian, L.、Zou, J.、Huang, J. 等：通过任务增强改善元学习的泛化能力。载于：《国际机器学习会议》，第11887–11897页。PMLR（2021年）[208] Yao, H.、Wei, Y.、Huang, J.、Li, Z.：层次结构化元学习。载于：《国际机器学习会议》，第7045–7054页。PMLR（2019年）[209] Yao, H.、Wu, X.、Tao, Z.、Li, Y.、Ding, B.、Li, R.、Li, Z.：自动关系元学习。arXiv预印本 arXiv:2001.00745（2020年）[210] Yin, M.、Tucker, G.、Zhou, M.、Levine, S.、Finn, C.：无记忆元学习。载于：《国际学习表征会议》（2020年）[211] Yoon, J.、Kim, T.、Dia, O.、Kim, S.、Bengio, Y.、Ahn, S.：贝叶斯模型无关元学习。载于：第32届神经信息处理系统国际会议论文集，第7343–7353页（2018年）[212] Yu, B.、Zhou, L.、Wang, L.、Fripp, J.、Bourgeat, P.：基于三维条件生成对抗网络的跨模态磁共振图像合成，用于脑肿瘤分割。载于：2018年IEEE第15届生物医学成像国际研讨会（ISBI 2018），第626–630页。IEEE（2018年）[213] Yu, T.、Quillen, D.、He, Z.、Julian, R.、Hausman, K.、Finn, C.、Levine, S.：Meta-World：多任务与元强化学习的基准测试和评估。载于：《机器人学习会议》，第1094–1100页。PMLR（2020年）[214] Yurt, M.、Dar, S.U.、Erdem, A.、Erdem, E.、Oguz, K.K.、Cukur, T.：MustGAN：用于磁共振图像合成的多流生成对抗网络。《医学图像分析》**70**，101944（2021年）[215] Zbontar, J. 等：fastMRI：用于加速磁共振成像的开放数据集与基准测试。arXiv:1811.08839（2018年），https://arxiv.org/abs/1811.08839[216] Zhan, Z.、Cai, J.F.、Guo, D.、Liu, Y.、Chen, Z.、Qu, X.：用于磁共振成像重建的几何方向快速多类别字典学习。《IEEE生物医学工程汇刊》**63**（9），1850–1861（2015年）[217] Zhang, C. 等：APIR-Net：使用神经网络的自校准并行成像重建。载于：《医学图像重建中的机器学习》，第36–46页。施普林格国际出版，Cham（2019年）

[218] Zhang, J.、Ghanem, B.：ISTA-Net：受可解释优化启发的深度网络 k
用于图像压缩感知。载于：IEEE 计算机会议论文集。 er
视觉与模式识别。第1828–1837页（2018年）[219] Zhang, K., Zuo, W., Gu, S., Zhang, L.：学习用于图像的深度CNN去噪器先验。 e
复原。载于：IEEE计算机视觉与模式识别会议论文集。 n
识别。第 3929–3938 页（2017年）[220] Zhang, X., Lian, Q., Yang, Y., Su, Y.：一种受全变分启发的压缩感知MRI深度展开网络。《数字信号处理》107，102856（2020年）[221] Zhang, X., Burger, M., Bresson, X., Osher, S.：Bregman化非局部正则化 n
用于反卷积和稀疏重建。《SIAM成像科学杂志》3(3),
253–276 (2010)[222] Zhao, H., 等：用于神经网络图像复原的损失函数。《IEEE计算成像汇刊》3(1), 47–57 (2017年)[223] Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., van Ginneken, B., Madabhushi, A., Prince, J.L., Rueckert, D., Summers, R.M.：医学成像中的深度学习综述：图像特征、技术趋势、突出进展的案例研究及未来展望。未知期刊（2020年）[224] Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.：Hi-Net：用于多模态MR图像合成的混合融合网络。《IEEE医学影像汇刊》39(9), 2772–2781 (2020年)[225] Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.：Hi-Net：用于多模态MR图像合成的混合融合网络。《IEEE医学影像汇刊》39(9), 2772–2781 (2020年)[226] Zhou, Z., Han, F., Ghodrati, V., Gao, Y., Yin, W., Yang, Y., Hu, P.：并行成像与卷积神经网络相结合的快速MR图像重建：在低延迟加速实时成像中的应用。《医学物理学》46(8), 3399–3413 (2019年)。 https://doi.org/10.1002/mp.13628, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13628[227] Zhou, Z., 等：并行成像与卷积神经网络相结合的快速MR图像重建：在低延迟加速实时成
g.
《医学物理学》46(8), 3399–3413 (2019)[228] Zhu, B. 等：通过域变换流形学习进行图像重建。《自 re
555, 487–492 (2018年). https://doi.org/10.1038/nature25988[229] Zhu, M., Chan, T.：一种用于全变差图像复原的高效原始-对偶混合梯度算法。UCLA CAM 报告 34, 8–34 (2008)

[230] Zhu, Y. 等：通过共享潜在高斯混合模型实现跨域医学图像转换。载于：医学图像计算与计算机辅助干预 – MICCAI 2020。第 379–389 页。施普林格国际出版，Cham（2020年）

个人简介

Wanyu Bian 于2017年毕业于密苏里大学哥伦比亚分校，获得数学和统计学学位。她于2022年获得佛罗里达大学数学系博士学位。