Federated Learning of Generative Image Priors for MRI Reconstruction

Gokberk Elmas, Salman U. H. Dar, *Member, IEEE,* Yilmaz Korkmaz, Emir Ceyani, Burak Susam, Muzaffer Ozbey, Salman Avestimehr, *Fellow, IEEE,* and Tolga Çukur, *Senior Member, IEEE*

Abstract— Multi-institutional efforts can facilitate training of deep MRI reconstruction models, albeit privacy risks arise during cross-site sharing of imaging data. Federated learning (FL) has recently been introduced to address privacy concerns by enabling distributed training without transfer of imaging data. Existing FL methods employ conditional reconstruction models to map from undersampled to fully-sampled acquisitions via explicit knowledge of the accelerated imaging operator. Since conditional models generalize poorly across different acceleration rates or sampling densities, imaging operators must be fixed between training and testing, and they are typically matched across sites. To improve patient privacy, performance and flexibility in multi-site collaborations, here we introduce Federated learning of Generative IMage Priors (FedGIMP) for MRI reconstruction. FedGIMP leverages a two-stage approach: cross-site learning of a generative MRI prior, and prior adaptation following injection of the imaging operator. The global MRI prior is learned via an unconditional adversarial model that synthesizes high-quality MR images based on latent variables. A novel mapper subnetwork produces site-specific latents to maintain specificity in the prior. During inference, the prior is first combined with subject-specific imaging operators to enable reconstruction, and it is then adapted to individual cross-sections by minimizing a data-consistency loss. Comprehensive experiments on multi-institutional datasets clearly demonstrate enhanced performance of FedGIMP against both centralized and FL methods based on conditional models.

Index Terms— MRI, accelerated, reconstruction, generative, prior, federated learning, distributed, collaborative.

I. Introduction

MAGNETIC Resonance Imaging (MRI) is a principal radiological modality owing to its non-invasiveness and exceptional soft-tissue contrast. Yet, an inevitable consequence of its low signal-to-noise ratio (SNR) efficiency is prolonged exams that hinder clinical use. Accelerated MRI methods based on undersampled acquisitions improve efficiency by recovering missing data via reconstruction algorithms that incorporate additional prior information [1], [2]. Deep learning models have been adopted for MRI reconstruction, given their strong ability to capture data-driven priors for inverse problems [3], [4], [5], [6], [7], [8], [9], [10], [11]. Deep reconstruction models are typically trained to perform a conditional mapping from undersampled acquisitions to images that are consistent with respective fully-sampled acquisitions [12], [13], [14], [15], [16], [17], [18], [19], [20]. Since these models typically show poor generalization to features scarcely present in the training set, learning of generalizable models involves training on a large and diverse collection of MRI data [21]. Unfortunately, economic and labor costs along with patient privacy concerns can prohibit compilation of comprehensive datasets centralized at a single institution [22].

Aiming at this limitation, federated learning (FL) is a promising framework that facilitates multi-institutional collaborations via decentralized training of learning-based models [23], [24], [25], [26], [27], [28]. An FL server periodically collects locally-trained models from individual sites in order to obtain a shared global model across sites [29], [30]. Following aggregation of local models, the global model is then broadcast back onto individual sites for continual training. This decentralized procedure allows a multi-site model to be collaboratively trained without sharing of local data, thereby mitigating privacy concerns [31]. A multi-site model can improve generalization over single-site models given the native diversity in multi-institutional data, which can substantially benefit sites with relatively limited or uniform training data. However, this comes at the potential expense of lower site-specific performance due to data heterogeneity across sites, lowering sensitivity to site-specific image features particularly for non-adaptive architectures that are vulnerable to domain shifts [27], [32].

Characteristics of acquired data are governed by two extrinsic factors in the context of accelerated MRI. At primary

level, the spatial distribution of tissue magnetization (i.e., the MR image distribution) is determined by the pulse sequence and the scanner [21]. At secondary level, the MR image distribution is further modulated by the accelerated imaging operator, which captures the influences of undersampling in k-space and reception by coil arrays [33]. As such, multi-site MRI data can show substantial heterogeneity in the MR image distribution (e.g., due to different sequences, scanners) or the imaging operator (e.g., due to different acceleration rates, sampling densities). In turn, multi-site models can suffer from performance losses under significant domain shifts across sites, or across the training and test sets [28], [34].

Few recent studies on FL-based MRI reconstruction have considered domain shifts across sites. In [35], adversarial alignment between source and target sites was proposed to improve similarity of latent-space representations in reconstruction models. In [36], reconstruction models were split into a global encoder shared across sites, followed by local decoders trained separately at each site. Despite improved performance against across-site variability, both methods are based on conditional models that assume explicit knowledge of the imaging operator. Conditional models generalize poorly under changes to the imaging operator [37], [38]. In turn, matching acceleration rates and sampling densities must be prescribed between the training and test sets, and similar prescriptions are often utilized among sites [35]. This requirement can restrict flexibility in multi-institutional collaborations and necessitate model retraining under significant changes to the desired imaging operator [34].

Here, we introduce a novel FL method for MRI reconstruction, FedGIMP, to enhance patient privacy, performance and flexibility in multi-site collaborations. Unlike previous FL methods, we propose a two-stage reconstruction approach: cross-site training of a generative MRI prior that captures the MR image distribution (Fig. 1), and prior adaptation following injection of the subject-specific imaging operator (Fig. 2). To improve generalization, we train a global MRI prior operationalized as an unconditional adversarial model that synthesizes high-quality MR images. Training images are derived from coil-combined, fully-sampled acquisitions to decouple the influence of the imaging operator from the prior. The statistics of feature maps across the synthesizer are controlled by latent variables produced by a mapper. To improve specificity, we propose a novel mapper that produces site-specific latents given a site index. During inference, the global MRI prior is combined with subject-specific imaging operators that can vary flexibly across sites and between the training and test sets. Reconstruction is then performed by adapting the MRI prior to enforce consistency between synthesized and acquired k-space data. This adaptation further boosts site specificity by increasing conformity of the multi-site model to the distribution of test data. Code for FedGIMP can be found at: https://github.com/icon-lab/FedGIMP.

Contributions:

- We introduce a novel FL method for MRI reconstruction that decentrally trains a generative MRI prior decoupled from the imaging operator to improve patient privacy and flexibility in multi-site collaborations.

- We operationalize the prior as an unconditional adversarial model with a cross-site-shared generator to capture site-general representations, albeit with site-specific latents to maintain specificity in synthesized MR images.
- We leverage prior adaptation to enhance specificity of the multi-site model and improve reliability against domain shifts across sites and across training and test sets.

II. RELATED WORK

Deep MRI reconstruction is pervasively based on conditional models that directly map undersampled acquisitions to images consistent with fully-sampled acquisitions [4]. These models are trained on large sets of paired input-output data under a specific accelerated imaging operator. Heavy data demand limits applicability since curating large datasets at a single site is challenging [39], [40]. To facilitate curation, unpaired [14], [41], [42], self-supervised [43], [44], [45], [46], [47], [48], or transfer [6], [34] learning strategies were proposed. However, these methods require centralized training following cross-site data transfer that raises patient privacy concerns [22].

FL is a decentralized framework for multi-institutional collaborations that communicates model parameters instead of raw data [31]. FL distributes costs related to the formation and processing of datasets across sites, while mitigating concerns regarding the data privacy [22]. FL methods have readily been demonstrated on imaging tasks such as segmentation and classification [23], [24], [26]. A major consideration is the reliability against domain shifts in multi-site imaging data, collected with varying imaging protocols and scanners. Data harmonization was proposed to remove site-specific variations while emphasizing shared variability across sites [49], [50], [51]. Although harmonization can improve population-level analysis, it can discard patient-level information of diagnostic value. Episodic learning in frequency space was proposed to improve generalization to unseen test domains for segmentation [28]. Adversarial alignment and network splitting methods were also proposed for classification [27], [52], [53]. While promising results were reported, it is nontrivial to directly adopt image analysis models for MRI reconstruction that requires image formation from raw data.

Domain shifts in MRI reconstruction involve heterogeneity in the MR image distribution and in the imaging operator, which can elicit performance losses when heterogeneity is prominent across sites, or across the training-test sets [34]. Few recent FL studies on single-coil MRI reconstruction have considered domain shifts across sites. In [35], cross-site-shared latent-space representations were obtained by adversarially aligning all sites to the targeted test site in each communication round. In [36], a split reconstruction model with a global encoder and unshared decoder was used to maintain site-specific and site-general representations. While demonstrating remarkable results, these recent methods are based on conditional models that are explicitly informed on the imaging operator. This can limit reconstruction performance and necessitate model retraining under notable domain shifts in the imaging operator [37], [38]. As such, previous studies have
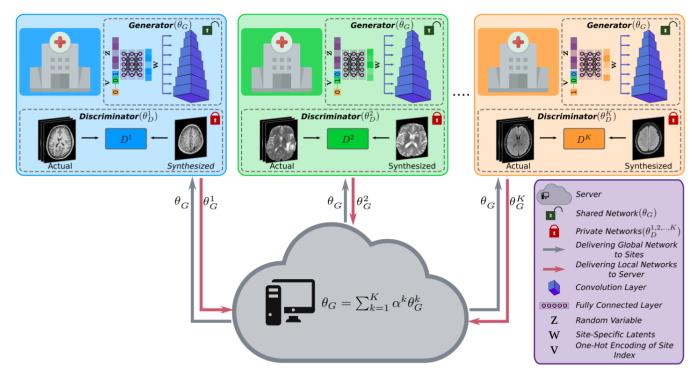
Fig. 1. FedGIMP performs federated learning of a generative image prior for MRI reconstruction. The prior is embodied as an unconditional adversarial model that synthesizes high-quality MR images given site-specific latent variables. Cross-site-shared generators of parameters θ_G are used along with site-specific discriminators of parameters $\theta_D^{1,...,K}$. In each communication round, sites perform local updates to θ_G to minimize a synthesis loss, and the server then aggregates updated models.
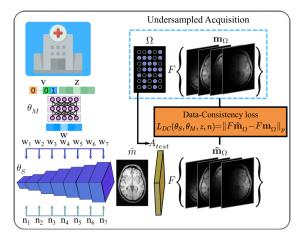

Fig. 2. FedGIMP's MRI prior is embodied as a generator that synthesizes a high-quality, coil-combined MR image ($\hat{m}$). To perform reconstruction, the trained generator is combined with the subject-specific imaging operator at a test site (A_{test}) and adapted to minimize a data-consistency loss on acquired k-space samples (L_{DC}). L_{DC} is expressed by projecting $\hat{m}$ onto individual coils, undersampling multi-coil images in k-space (with the pattern Ω), and comparing synthesized and acquired k-space samples. For each cross-section, inference optimization is conducted over the synthesizer and mapper parameters ($\theta_{S,M}$), normal variables (z) and noise (n).

prescribed matching acceleration rates and sampling densities between the training-test sets, and usually across sites.

Here, we propose an FL method that learns a generative MRI prior, and reconstructs images via prior adaptation after combination with the subject-specific imaging operator. Separation of the MRI prior from the imaging operator has recently been considered for centralized reconstruction models

trained on single-site data [32], [54], [55], yet it has not been studied in the context of FL. To our knowledge, FedGIMP is the first FL method that decouples the MRI prior from the imaging operator to train decentralized reconstruction models on multi-site datasets, and the first FL method that reconstructs multi-coil MRI data. Furthermore, FedGIMP includes several unique design elements compared to previous centralized methods. In [54], a non-adaptive MRI prior was proposed that was kept static during inference. In [32], an adaptive MRI prior was proposed that did not include latents, so no latent optimization was performed during inference. In [55], an adaptive MRI prior was proposed that included mapper-produced latents albeit no site index, and the mapper was removed from the prior to directly optimize latents during inference. Unlike previous methods, FedGIMP introduces a novel site-index in its adaptive MRI prior to maintain site specificity, and it updates the mapper to indirectly optimize latents. These unique aspects enable FedGIMP to offer improved performance for MRI reconstruction on heterogeneous datasets.

III. THEORY

A. Federated Learning of Conditional MRI Models

Accelerated MRI entails reconstruction of an underlying MR image m from undersampled k-space acquisitions y:

$$Am = y, \qquad (1)$$

where A is the imaging operator that includes the effects of coil sensitivities and partial Fourier transformation on acquired k-space. As Eq. 1 is underdetermined, additional prior

information is incorporated to regularize the reconstruction:

$$\widehat{m} = \underset{m}{\operatorname{argmin}} \|y - Am\|_2^2 + H(m), \qquad (2)$$

where $H(m)$ enforces the prior [1]. Deep models have recently become the predominant method for solving Eq. 2 by training of data-driven priors on diverse datasets [4].

In FL, training is performed via communication between multiple sites and a server [31]. The server retains a global model (C_θ with parameters θ), whereas each site retains a local model of matching architecture (C_θ^k for site k, where $k \in \{1, .., K\}$). In each communication round, local models are initialized with the global model broadcast from the server, $\theta_C^k \leftarrow \theta_C$, and updated to minimize a local reconstruction loss:

$$\mathcal{L}_{Rec}^k(\mathcal{D}^k, A_{tr}^k; \theta^k) = \mathbb{E}_{(m_{tr}^k, y_{tr}^k)} \left[\|m_{tr}^k - C_{\theta^k}(A_{tr}^{\dagger k} y_{tr}^k)\|_2 \right], \quad (3)$$

where $\mathbb{E}$ denotes expectation, $\mathcal{D}^k$ are local training data comprising undersampled acquisitions (y_{tr}^k) and reference images obtained from fully-sampled acquisitions (m_{tr}^k), and A_{tr}^k, $A_{tr}^{\dagger k}$ are the imaging operator and its adjoint. C is a conditional model with parameters θ^k that receives zero-filled Fourier reconstruction of y_{tr}^k. At the end of each round, updated local models are aggregated via federated averaging (FedAvg) [29]:

$$\theta = \sum_{k=1}^{K} \alpha^k \theta^k, \qquad (4)$$

where α^k denotes the relative site weights.

The trained global model ($C_{\theta*}$) is then used for inference:

$$\widehat{m}_{k,s} = C_{\theta*}(A_{test}^{\dagger k} y_{test}^{k,s}), \qquad (5)$$

where $\widehat{m}_{k,s}$ is the reconstruction and $y_{test}^{k,s}$ is the undersampled acquisition for the sth subject at the kth site, and $A_{test}^{\dagger k}$ is the adjoint imaging operator at site k. Since conditional models generalize poorly against heterogeneity in the imaging operator, A_{tr}^k and A_{test}^k are typically matched across training-test sets (i.e. across training and test subjects), and across sites.

B. Federated Learning of Generative MRI Priors

To improve flexibility in multi-site collaborations, we propose a novel FL method based on decentralized learning of generative MRI priors. A global MRI prior is trained using an unconditional adversarial model with shared generator and local discriminator networks (Fig. 1). During inference, the prior is combined with subject-specific imaging operators and adapted to the reconstruction task (Fig. 2). The proposed architecture and learning procedures are described below.

1) Unconditional Adversarial Model: FedGIMP employs an unconditional architecture to learn a generative prior for high-quality MR images. Training images are derived from coil-combined, fully-sampled acquisitions to obtain a prior that is agnostic to k-space undersampling and individual coil sensitivities. Here, we propose a style-generative model as inspired by the success of this model family in natural image synthesis [56]. Our proposed model has a generator that synthesizes realistic MR images, wherein the statistics of feature maps at each stage are controlled via latent variables.

These statistics are expected to be different across sites given the native heterogeneity of MR images collected in separate sites. Our model introduces a novel mapper with a site index to produce site-specific latent variables, thereby increasing specificity in synthesized feature maps. Finally, a discriminator differentiates between synthetic and actual MR images.

a) Generator (G): The generator uses mapper (M) and synthesizer (S) subnetworks to transform low-dimensional random variables onto high-dimensional MR images. $\underline{M}$ receives i.i.d. normal variables $z \in \mathbb{R}^{1 \times J}$ concatenated with a one-hot encoding vector for site index $v \in \mathbb{R}^{1 \times K}$. Given an input in $\mathbb{R}^{1 \times (J+K)}$, it computes latent variables $w \in \mathbb{R}^{1 \times J}$:

$$w = M_{\theta_M}(z \oplus v), \qquad (6)$$

where $\oplus$ denotes concatenation. At the ith layer of M with L_M layers, latent variables z_i are mapped onto z_{i+1} as:

$$z_{i+1} = FC_{M,i}(z_i), \qquad (7)$$

where $FC_{M,i}$ is a fully-connected layer with parameters $\theta_{M,i} \in \mathbb{R}^{(J+K) \times J}$ if $i = 1$; $\mathbb{R}^{J \times J}$ if $i \neq 1$. Dimensionality is reduced from $J + K$ to J in the first layer, and retained in remaining layers. Receiving w from M, $\underline{S}$ generates an MR image progressively across L_S layers. In the ith layer, feature maps from the preceding layer ($X_i^0 \in \mathbb{R}^{\frac{h_1}{2} \times \frac{h_2}{2} \times q}$) are two-fold upsampled. The upsampled maps ($X_i^1 \in \mathbb{R}^{h_1 \times h_2 \times q}$) are then processed through a cascade of blocks: $Conv_S^1$ (convolution), NI^1 (noise injection), $AdaIN^1$ (adaptive instance normalization), $Conv_S^2$, NI^2, $AdaIN^2$. The first block extracts local features, $X_i^2 = Conv_S^1(X_i^1)$:

$$X_i^2 = \begin{bmatrix} \sum_c X_i^{1,c} \circledast \theta_{S,i}^{c,1} \\ \vdots \\ \sum_c X_i^{1,c} \circledast \theta_{S,i}^{c,u} \end{bmatrix}, \qquad (8)$$

where $\theta_{S,i}^1 \in \mathbb{R}^{j \times j \times q \times u}$ are convolution kernels, u is the number of output feature channels, c is the channel index, and $\circledast$ denotes convolution. In Eq. 8, X_i^2 is a 3D tensor and matrix rows span the 3rd dimension of X_i^2. Next, the noise-injection block introduces pixel-level intensity modulations in the feature maps. Recent computer vision studies have reported that introducing noise variables across the synthesizer improves details in synthetic natural images [56]. Accordingly, scaled noise variables are added onto feature maps to control low-level structural details, $X_i^3 = NI^1(X_i^2)$:

$$X_i^3 = \begin{bmatrix} \varphi(X_i^{2,1} + \epsilon_i^{1,1} n_i^1) \\ \vdots \\ \varphi(X_i^{2,u} + \epsilon_i^{1,u} n_i^1) \end{bmatrix}, \qquad (9)$$

where $n_i^1 \in \mathbb{R}^{h_1 \times h_2}$ is multiplied with the scalar $\epsilon_i^{1,c}$ and added onto the cth channel of $X_i^2 \in \mathbb{R}^{h_1 \times h_2 \times u}$, φ is an activation function. To control high-level style features, adaptive instance normalization modulates feature maps given intermediate latent variables, $X_i^4 = AdaIN^1(X_i^3, w)$ [57]:

$$X_i^4 = \gamma_i^1(w) \frac{(X_i^3 - \mu(X_i^3))}{\sigma(X_i^3)} + \beta_i^1(w), \qquad (10)$$

Algorithm 1 FedGIMP Training

Input: $\mathcal{D} = \{\mathcal{D}^1, \ldots, \mathcal{D}^K\}$: datasets from K sites.
L: number of communication rounds.
I: number of local epochs.
$\{G^1, \ldots, G^K\}$: local generators of params.
$\{\theta_G^1, \ldots, \theta_G^K\}$.
$\{D^1, \ldots, D^K\}$: local discriminators of params.
$\{\theta_D^1, \ldots, \theta_D^K\}$.
$\{\alpha^1, \ldots, \alpha^K\}$: weighting coefficients of K sites.
G: global generator of params. θ_G.
$Opt()$: optimizer that computes parameter updates.
Output: G trained global model of params. θ_G^*.

1 Initialize parameters
2 **for** $l = 1 : L$ **do**
3   **for** $k = 1 : K$ **do**
4      $\theta_G^k \leftarrow \theta_G$ (receive global generator)
5   **for** $i = 1 : I$ **do**
6   Calculate $\nabla_{\theta_G^k} \mathcal{L}_G^k(\theta_G^k)$ based on Eq. 13
7        $\theta_G^k \leftarrow \theta_G^k - Opt(\nabla_{\theta_G^k} \mathcal{L}_G^k)$
8   Calculate $\nabla_{\theta_D^k} \mathcal{L}_D^k(\mathcal{D}_i^k; \theta_D^k)$ based on Eq. 14
9        $\theta_D^k \leftarrow \theta_D^k - Opt(\nabla_{\theta_D^k} \mathcal{L}_D^k)$
10    $\theta_G \leftarrow \sum_{k=1}^{K} \alpha^k \theta_G^k$ (aggregate local generators)
11 **return** $\theta_G^* = \theta_G$

Algorithm 2 FedGIMP Inference

Input: $y_{test}^{k,s}$: Undersampled data for k^{th} site, s^{th} subject.
$A_{test}^{k,s}$: subject-specific imaging operator for $y_{test}^{k,s}$.
$G = \{M, S\}$: global generator with params.
$\theta_G^* = \{\theta_M^*, \theta_S^*\}$.
v: site-specific one-hot encoding vector.
z: i.i.d normal variable.
n: randomly initialized noise variable.
E: number of iterations for inference.
Output: $\hat{m}_{k,s}$: reconstructed image.

1 $\theta_S^1 \leftarrow \theta_S^*, \theta_M^1 \leftarrow \theta_M^*, z^1 \leftarrow z, n^1 \leftarrow n$ (initialize)
2 **for** $e = 1 : E$ **do**
3   Calculate $\nabla_{\theta_S^e, \theta_M^e, z^e, n^e} \mathcal{L}_{DC}^{k,s}(y_{test}^{k,s}, A_{test}^{k,s}; \theta_S^e, \theta_M^e, z^e, n^e)$ based on Eq. 16
4    $\theta_S^{e+1} \leftarrow \theta_S^e - Opt(\nabla_{\theta_S^e} \mathcal{L}_{DC}^{k,s})$
5    $\theta_M^{e+1} \leftarrow w^e - Opt(\nabla_{\theta_M^e} \mathcal{L}_{DC}^{k,s})$
6    $z^{e+1} \leftarrow z^e - Opt(\nabla_{z^e} \mathcal{L}_{DC}^{k,s})$
7    $n^{e+1} \leftarrow n^e - Opt(\nabla_{n^e} \mathcal{L}_{DC}^{k,s})$
8 $\hat{m}_{k,s} = S_{\theta_S^E}(M_{\theta_M^E}(z^E \oplus v), n^E)$
9 **return** $\hat{m}_{k,s}$

---

where $\gamma_i^1 \in \mathbb{R}^{u \times 1}$ and $\beta_i^1 \in \mathbb{R}^{u \times 1}$ are learnable affine transformations of w that control the scale and bias of each feature channel respectively. μ and σ are mean and variance along the channel dimension of X_i^3. Note that the cascade of convolution, noise injection and adaptive instance normalization is repeated a second time to yield the mappings $X_i^5 = Conv_S^2(X_i^4)$, $X_i^6 = NI^2(X_i^5)$, $X_{i+1}^0 = AdaIN^2(X_i^6, w)$. The parameters of these repeat blocks are distinct from those in the initial blocks, except for tied intermediate latents input to $AdaIN^{1,2}$. Overall, the mapping through the generator is:

$$\hat{x} = G_{\theta_G}(z \oplus v, n) = S_{\theta_S}(M_{\theta_M}(z \oplus v), n), \quad (11)$$

where $\hat{x}$ is the synthesized image, the generator parameters $\theta_G = \{\theta_M, \theta_S\}$ include mapper and synthesizer parameters.

b) Discriminator (D): The discriminator attempts to discriminate between actual or synthetic MR image inputs. D contains $L_D - 1$ layers cascading two-fold downsampling and convolution blocks ($Conv_D$), followed by a final fully-connected layer (FC_D). The overall mapping is:

$$x_D = D_{\theta_D}(x), \quad (12)$$

where x is an actual (x_r) or synthetic image ($\hat{x}$), $x_D \in \mathbb{R}^1$ is the output, and D_{θ_D} is parametrized by θ_D.

2) Training of the Global MRI Prior: To learn a global MRI prior, FedGIMP performs federated training of the unconditional adversarial model (Alg. 1). The training lasts a total of L communication rounds between the sites and the server. A shared generator (G) and K local generator copies (G^k) are maintained. Meanwhile, K local discriminators (D^k) are not exchanged to limit communication load and augment data

privacy. In the first round, G and D^k are randomly initialized. At the start of each round, local generators are initialized with the global generator broadcast from the server, $\theta_G^k \leftarrow \theta_G$. In each round, I local epochs are performed to update local models. The local generator updates are calculated to minimize a non-saturated logistic adversarial loss (L_G^k):

$$L_G^k(\theta_G^k) = -\mathbb{E}_{p(z)}\left[\log(f(D^k(G_{\theta_G^k}^k(z \oplus v, n)))\right], \quad (13)$$

where $\mathbb{E}_{p(.)}$ is expectation with respect to probability distribution p. The local discriminator updates are calculated to also minimize a non-saturated logistic adversarial loss (L_D^k), along with a gradient penalty according to the learned distribution of actual MR images $p(x_r^k)$ [56]:

$$
\begin{aligned}
L_D^k(\mathcal{D}^k; \theta_D^k) = &-\mathbb{E}_{p(z)}\left[\log(1 - f(D_{\theta_D^k}^k(G(z \oplus v, n))))\right] \\
&- \mathbb{E}_{p(x_r^k)}\left[\log(f(D_{\theta_D^k}^k(x_r^k)))\right] \\
&+ \frac{\delta}{2}\mathbb{E}_{p(x_r^k)}\left[\left\|\nabla D_{\theta_D^k}^k(x_r^k)\right\|^2\right],
\end{aligned}
\quad (14)
$$

where $\mathcal{D}^k$ are training data from site k, i.e. coil-combined MR images derived from fully-sampled acquisitions (x_r^k). After I iterations of updates according to Eqs. 13 and 14, local generators are sent to the server for aggregation [29]:

$$\theta_G = \sum_{k=1}^{K} \alpha^k \theta_G^k \quad (15)$$

Here, α^k is set to $\frac{N^k}{N}$, where N is the total number of training samples across all sites and N^k is the number of training samples at the kth site.

3) Inference at a Test Site: FedGIMP trains an unconditional prior that randomly generates high-quality, coil-combined, synthetic MR images. Synthetic images do not carry information on coil sensitivities or k-space sampling patterns, and they lack anatomical correspondence to the actual subject. Thus, a dedicated inference procedure is required for reconstruction. Synthetic images define a constraint reflecting the learned image distribution, whereas the subject's undersampled acquisitions define an anatomical constraint in accordance with the imaging operator. The two constraints do not necessarily have an intersection (e.g., a prior trained on brain images used to reconstruct knee images). For reconstruction, FedGIMP adapts its prior to refine the associated constraint and intersect it with the anatomical constraint. This adaptation enforces the prior to generate an image whose undersampled k-space data matches acquired data.

During inference, FedGIMP first combines the trained MRI prior with the subject-specific imaging operator at a test site ($A_{test}^{k,s}$). To do this, the prior can be included in the optimization problem for various model-based reconstructions [8], [9], [13], [58]. Here, we follow a straightforward approach where the prior is adapted to minimize a data-consistency loss between synthesized and acquired k-space data (Alg. 2):

$$L_{DC}^{k,s}(y_{test}^{k,s}, A_{test}^{k,s}; \theta_S, \theta_M, z, n) = \left\| A_{test}^{k,s} S_{\theta_S}(w, n) - y_{test}^{k,s} \right\|_2$$
$$+ \eta \left\| \nabla S_{\theta_S}(w, n) \right\|, \quad (16)$$

where $w = M_{\theta_M}(z \oplus v)$, and a gradient penalty is included with weight η to prevent noise amplification [59]. To compute Eq. 16, FedGIMP estimates synthetic k-space data by projecting the synthetic image $S_{\theta_S}(w, n)$ through the imaging operator $A_{test}^{k,s} = \Omega^{k,s} \mathcal{F} B^{k,s}$, which embodies multiplication with coil sensitivities ($B^{k,s}$), Fourier transformation ($\mathcal{F}$), and undersampling with the prescribed sampling pattern ($\Omega^{k,s}$) [55]. Anatomical correspondence between synthetic and actual data is then achieved by minimizing $L_{DC}^{k,s}$. For each cross-section, the optimization is performed over θ_S, θ_M, z and n. Given the trained global generator $G_{\theta_G^*}$, synthesizer and mapper parameters are initialized as $\theta_S^1 \leftarrow \theta_S^*$ and $\theta_M^1 \leftarrow \theta_M^*$, whereas instance-specific variables are randomly initialized as $z^1 \leftarrow z$ and $n^1 \leftarrow n$. Following a total of E iterations, the adapted prior is used to compute the final reconstruction:

$$\hat{m}_{k,s} = S_{\theta_S^E}(M_{\theta_M^E}(z^E \oplus v), n^E) \quad (17)$$

Thus, the synthetic image produced by the generator at the end of the optimization yields the final reconstructed image. Note that the values of z^E and n^E are not random but optimized.

IV. METHODS

A. Architectural Details

The unconditional adversarial model in FedGIMP used a mapper with 8 FC layers receiving a standard normal vector and a one-hot binary encoding vector of site index as inputs, while outputting 32 intermediate latent variables. A synthesizer with 8 layers was used, where each layer contained a bilinear upsampling block for 2-fold increase of feature map resolution, followed by two serial cascades of Conv, NI and AdaIN blocks. The first layer received a 4×4 map of learnable constant values, initialized with ones, as input. The learnable noise variable was randomly initialized from a standard normal distribution. A discriminator with 8 layers was used, each containing a bilinear downsampling block for 2-fold reduction in resolution, and a convolution block with 3×3 kernel size. Leaky ReLU activation functions were used. Two separate channels at the generator's output and the discriminator's input were used to represent real and imaginary parts of complex MR images. During training, images were zero-padded to match the resolution of the output layer if needed. During inference, synthesized images were centrally cropped based on the size of the acquisition matrix prior to calculation of data-consistency loss. The synthesizer and discriminator were trained non-progressively, with all layers intact.

B. Competing Methods

FedGIMP was demonstrated against a traditional reconstruction (LORAKS), non-federated models (GAN_cond/GIMP), and federated conditional models (FL-MRCM, FedGAN, LG-Fed, FedMRI). Hyperparameter selection was performed based on validation performance. For each method, a single set of learning rate, number of epochs (non-federated), number of communication rounds and epochs (federated), and loss term weights were selected that yielded near-optimal performance across tasks. All models were trained via the Adam optimizer. A learning rate of 2×10^{-3} and $\delta = 10$ was selected for unconditional models; a learning rate of 2×10^{-4} was selected for conditional models. Decay rate parameters for the optimizer were adopted from previous studies as $\beta_1 = 0.0$, $\beta_2 = 0.99$ for unconditional [56], and $\beta_1 = 0.5$, $\beta_2 = 0.99$ for conditional models [18]. Training lasted 100 epochs for non-federated models, $L = 100$ rounds and $I = 1$ epochs for federated models. During inference, unconditional models enforced anatomical consistency to a test subject by minimizing a data-consistency loss. No notable improvement was observed in conditional models subjected to inference optimization based on data-consistency loss, so they used a strict data-consistency projection following a forward pass. Strict data consistency was enforced on all reconstructions prior to reporting. To ensure adherent implementation of competing methods, we utilized libraries from the public repositories shared by the originally proposing authors. Accordingly, we implemented LORAKS in Matlab,[1] conditional models in PyTorch.[2],[3],[4] Unconditional models were implemented in TensorFlow to adopt libraries from a style-generative architecture for natural image synthesis.[5] Models were executed on a system with four Nvidia RTX 3090s.

FedGIMP: The proposed model was trained to synthesize coil-combined MR images. The model output magnitude images for single-coil experiments, and complex images with real and imaginary components as separate channels

---

[1] https://mr.usc.edu/download/LORAKS2
[2] https://github.com/icon-lab/ProvoGAN
[3] https://github.com/guopengf/FL-MRCM
[4] https://github.com/chunmeifeng/FedMRI
[5] https://github.com/NVlabs/stylegan

for multi-coil experiments. Inference was performed via the Adam optimizer, with optimized parameters transferred across consecutive cross-sections in a volume to improve convergence [55]. A learning rate of 10^{-2}, $\eta=10^{-4}$ and $E=1200$ iterations were selected via cross validation. Reconstruction performance improves for higher E, yet the benefits become marginal beyond a certain level. Thus, we exercised early stopping based on the L-curve criterion to maintain a favorable compromise between performance and inference time [55].

LORAKS: A traditional autocalibrated low-rank reconstruction was performed [60]. The k-space neighborhood radius and the rank of LORAKS matrix were selected as (2,6) for single-coil reconstruction and (2,30) for multi-coil reconstruction.

GAN$_{cond}$: A non-federated conditional model was trained to map zero-filled (ZF) reconstruction of undersampled acquisitions to reference images of fully-sampled acquisitions [18]. Weights of (pixel-wise, adversarial, perceptual) losses were set as (100, 1, 100). Centralized training was performed following dataset aggregation across sites, albeit an unshared discriminator was used to process data coming from different sites as in FedGIMP. GAN$_{cond}$ serves as a privacy-violating benchmark for conditional reconstruction.

GIMP: A non-federated unconditional model was trained based on the architecture and loss functions in FedGIMP. Centralized training was performed, other training and inference procedures were identical to FedGIMP. GIMP serves as a privacy-violating benchmark for unconditional reconstruction.

FL-MRCM: A federated conditional model was trained with adversarial alignment of latent representations across sites [35]. For fair comparison among FL methods, architecture and loss functions were adopted from GAN$_{cond}$. Latent representations from the residual backbone were passed through a convolution layer and provided to a domain-alignment network. Weights of (pixel-wise, adversarial, perceptual, reconstruction, domain-alignment) losses were set as (100, 1, 100, 0.5, 0.5).

FedGAN: A federated conditional model was trained with a shared encoder and decoder across sites [61]. The architecture and loss functions followed GAN$_{cond}$. Weights of (pixel-wise, adversarial, perceptual) losses were set as (100, 1, 100).

LG-Fed: A federated conditional model was trained with site-specific encoders and a shared decoder [62]. The architecture and losses followed GAN$_{cond}$. Weights of (pixel-wise, adversarial, perceptual) losses were set as (100, 1, 100).

FedMRI: A federated conditional model was trained with a shared encoder and site-specific decoders [36]. The architecture and losses followed GAN$_{cond}$, albeit an added contrastive loss was included [36]. Weights of (pixel-wise, adversarial, perceptual, contrastive) losses were set as (100, 1, 100, 10).

C. Experiments

1) Datasets: FL experiments were conducted using an in-house dataset acquired at Bilkent University along with three public datasets (IXI http://brain-development.org/ixi-dataset/), fastMRI [33], BRATS [63]). We describe the in-house imaging protocol below (see related references for public datasets). Acquisitions were retrospectively undersampled using variable- (VD) and uniform-density (UD) patterns at acceleration rates R = $(3\times, 6\times)$ [1]. There was no subject overlap between training, validation, and test sets. Each FL site corresponded to a distinct dataset, and all examined FL setups had 3 sites. The training set in each site aggregated MR images across multiple different contrasts. Training samples were randomly drawn from this aggregated set, without any special procedures for handling different contrasts. While modeling each contrast separately could enhance performance, we reasoned that training on mixed contrasts is a more realistic scenario given the ubiquity of multi-contrast protocols in clinical practice [34].

In-House: T$_1$-, T$_2$- and PD-weighted scans were performed in 10 subjects on a 3T Siemens Tim Trio scanner located at Bilkent University using a 32-channel coil. An MP-RAGE sequence was used for T$_1$-weighted scans with TE/TI/TR = 3.87/1100/2000 ms, 20° flip angle; and an FSE sequence was used for T$_2$-/PD-weighted scans with TE$_{PD}$/TE$_{T2}$/TR = 12/118/1000 ms, 90° flip angle. All scans were performed with 192×256x176 mm^3 field-of-view and 1×1x2 mm^3 voxel size. Ethics approval was obtained from the local ethics committee, and all participants gave written-informed consent.

2) Single-Coil Reconstruction: Experiments were performed on IXI, fastMRI, and BRATS. Coil-combined magnitude images in each dataset were treated as single-coil data, so their analyses omitted the channel dimension of coil arrays. T$_1$-, T$_2$-, PD-weighted acquisitions in IXI and BRATS, T$_1$c-, T$_2$-, FLAIR-weighted acquisitions in fastMRI were considered. For each dataset, multi-contrast acquisitions of (40,5,10) subjects were reserved as (training, validation, test) sets, with 21 cross-sections per contrast randomly selected in each subject. As such, the training, validation and test sets within each dataset (i.e., each site) contained (840,105,210) cross-sections.

3) Multi-Coil Reconstruction: Experiments were performed on fastMRI brain, fastMRI knee, and In-House brain datasets, which all contained multi-coil k-space data. For fastMRI brain, T$_1$-, T$_2$-, FLAIR-weighted acquisitions from (36,6,9) subjects were reserved, with 8 cross-sections per contrast randomly selected in each subject. For fastMRI knee, PD-, PDFS-weighted acquisitions from (48,8,12) subjects were reserved, with 9 cross-sections per contrast. For In-House, T$_1$-, T$_2$-, PD-weighted acquisitions from (6,1,3) subjects were reserved, with 48 cross-sections per contrast in the training and validation sets, and with 24 cross-sections per contrast in the test set. As such, the training, validation and test sets within each dataset contained (864,144,216) cross-sections. Coil compression was performed onto 5 virtual coils [64]. Conditional models mapped multi-coil, complex ZF reconstructions of undersampled data to reference images derived from fully-sampled data. GIMP and FedGIMP synthesized complex coil-combined MR images during training [58]. The imaging operator was only injected during inference to enforce data consistency based on multi-coil k-space data.

TABLE I
SINGLE-COIL RECONSTRUCTION PERFORMANCE WITH THE IMAGING OPERATOR MATCHED ACROSS SITES, AND ACROSS THE TRAINING-TEST SETS. BOLDFACE INDICATES THE TOP-PERFORMING FL METHOD

			IXI		fastMRI		BRATS	
			PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-fed	LORAKS	3x	28.6±2.5	69.9±6.5	29.1±4.6	70.4±11.3	35.1±4.0	95.1±0.5
		6x	26.0±1.9	61.9±6.6	26.0±3.9	62.0±12.2	31.7±3.6	92.3±0.9
	GAN_{cond}	3x	38.6±1.2	94.6±1.0	39.2±2.3	95.1±1.3	44.1±1.3	98.5±0.3
		6x	35.1±1.1	91.2±1.5	35.8±2.2	91.9±2.0	40.3±1.3	97.1±0.5
	GIMP	3x	42.2±1.4	98.2±0.5	40.6±2.2	97.1±1.0	50.6±1.9	99.7±0.1
		6x	37.6±1.3	96.4±1.0	36.6±2.3	94.5±1.6	45.4±1.7	99.2±0.2
Fed	FL-MRCM	3x	38.1±1.2	94.2±1.1	38.8±2.3	94.6±1.4	43.4±1.3	98.3±0.3
		6x	34.7±1.2	90.8±1.6	35.3±2.3	90.9±2.2	39.6±1.3	96.8±0.6
	FedGAN	3x	38.5±1.2	94.5±1.0	38.9±2.3	94.9±1.4	44.1±1.2	98.4±0.3
		6x	35.0±1.1	91.2±1.5	35.6±2.2	91.7±2.0	40.1±1.3	97.0±0.6
	LG-Fed	3x	38.4±1.2	94.3±1.0	38.9±2.3	94.9±1.4	44.0±1.3	98.4±0.3
		6x	34.8±1.0	90.6±1.4	35.6±2.3	91.6±2.0	40.1±1.3	97.1±0.6
	FedMRI	3x	38.6±1.2	94.4±1.0	39.0±2.3	94.9±1.3	44.2±1.3	98.5±0.3
		6x	34.9±1.1	90.5±1.5	35.5±2.2	91.4±2.0	40.3±1.4	97.2±0.5
	FedGIMP	3x	**42.9±1.5**	**98.5±0.5**	**41.1±2.2**	**97.4±0.8**	**50.9±1.9**	**99.7±0.1**
		6x	**37.8±1.4**	**96.5±1.0**	**36.9±2.2**	**94.8±1.4**	**45.4±1.7**	**99.2±0.2**

TABLE II
SINGLE-COIL RECONSTRUCTION PERFORMANCE WITH THE IMAGING OPERATOR MISMATCHED ACROSS THE TRAINING-TEST SETS, $A \to B$ DENOTES THE DOMAIN SHIFT IN R (UPPER PANEL) / SAMPLING DENSITIES (LOWER PANEL)

			IXI		fastMRI		BRATS	
Acceleration Rate (6x→3x or 3x→6x)								
			PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-fed	LORAKS	3x	28.6±2.5	69.9±6.5	29.1±4.6	70.4±11.3	35.1±4.0	95.1±0.5
		6x	26.0±1.9	61.9±6.6	26.0±3.9	62.0±12.2	31.7±3.6	92.3±0.9
	GAN_{cond}	3x	37.8±1.1	92.9±1.1	38.0±1.9	93.7±1.4	42.2±1.3	97.8±0.4
		6x	33.5±1.2	89.6±1.7	34.3±2.4	90.2±2.4	39.3±1.4	96.9±0.6
	GIMP	3x	42.2±1.4	98.2±0.5	40.6±2.2	97.1±1.0	50.6±1.9	99.7±0.1
		6x	37.6±1.3	96.4±1.0	36.6±2.3	94.5±1.6	45.4±1.7	99.2±0.2
Fed	FL-MRCM	3x	37.4±1.1	92.6±1.2	37.3±2.0	92.8±1.7	40.7±1.4	96.9±0.7
		6x	33.1±1.3	88.5±2.1	33.9±2.4	89.1±2.6	38.6±1.3	96.4±0.7
	FedGAN	3x	37.5±1.1	92.7±1.2	38.1±1.9	93.8±1.4	41.9±1.2	97.6±0.5
		6x	33.3±1.2	89.2±1.7	34.1±2.4	89.8±2.5	39.1±1.3	96.8±0.6
	LG-Fed	3x	37.6±1.0	92.9±1.0	37.8±1.9	93.4±1.6	42.2±1.3	97.8±0.5
		6x	33.9±1.2	89.9±1.5	34.3±2.3	90.2±2.3	39.6±1.3	97.0±0.6
	FedMRI	3x	37.6±1.0	92.8±1.1	37.8±1.8	93.5±1.5	42.7±1.3	98.0±0.4
		6x	34.1±1.2	90.2±1.5	34.3±2.4	89.9±2.4	39.6±1.4	97.1±0.6
	FedGIMP	3x	**42.9±1.5**	**98.5±0.5**	**41.1±2.2**	**97.4±0.8**	**50.9±1.9**	**99.7±0.1**
		6x	**37.8±1.4**	**96.5±1.0**	**36.9±2.2**	**94.8±1.4**	**45.4±1.7**	**99.2±0.2**
Sampling Density (VD→UD)								
			PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-fed	LORAKS	3x	25.5±1.4	80.9±3.7	26.1±2.9	79.9±6.5	31.5±2.9	94.5±0.6
		6x	21.7±0.8	67.4±4.5	21.2±1.7	63.9±7.1	24.7±2.4	85.6±1.4
	GAN_{cond}	3x	30.3±1.0	89.5±1.5	31.0±2.8	88.7±3.2	37.0±1.4	96.1±0.7
		6x	24.2±0.7	77.0±2.9	25.1±1.9	76.5±4.5	30.4±1.3	90.0±1.3
	GIMP	3x	35.3±1.6	95.7±1.1	35.1±2.3	93.8±1.6	47.0±2.1	99.5±0.2
		6x	26.8±0.9	84.0±2.4	28.1±2.6	82.3±3.9	33.4±1.4	93.6±1.0
Fed	FL-MRCM	3x	30.5±1.0	89.8±1.3	30.7±2.6	87.5±3.1	36.1±1.5	95.1±0.9
		6x	23.9±0.6	74.8±2.5	24.5±1.6	73.4±3.9	29.4±1.3	86.8±1.7
	FedGAN	3x	30.2±1.0	89.3±1.5	31.0±2.7	89.0±3.1	37.1±1.4	96.2±0.7
		6x	24.1±0.7	76.6±3.0	25.0±1.9	76.4±4.6	30.2±1.3	89.8±1.3
	LG-Fed	3x	30.6±1.1	90.0±1.4	31.0±2.7	88.7±3.1	36.9±1.5	96.1±0.8
		6x	24.2±0.7	77.4±2.9	25.1±1.9	76.5±4.7	30.2±1.3	89.9±1.4
	FedMRI	3x	30.7±1.1	90.4±1.3	31.1±2.7	89.0±2.9	37.2±1.5	96.3±0.7
		6x	24.3±0.7	77.8±2.6	25.0±1.8	76.1±4.4	30.4±1.3	90.3±1.3
	FedGIMP	3x	**35.1±1.4**	**95.4±1.0**	**35.3±2.4**	**94.2±1.7**	**47.4±2.2**	**99.6±0.2**
		6x	**26.6±1.0**	**83.9±2.5**	**28.0±2.5**	**82.3±4.0**	**33.4±1.4**	**93.4±1.0**

D. Quantitative Assessments

Reconstructed images were compared against Fourier reconstruction of fully-sampled acquisitions as reference. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were measured, following normalization of images to [0 1]. Tables list mean and standard deviation of performance metrics across test subjects. Statistical significance of performance differences was assessed via Wilcoxon sign-rank tests.

V. RESULTS

A. Single-Coil Reconstruction

1) Domain Shifts in the MR Image Distribution: Multi-site datasets of multi-contrast MRI data contain intrinsic domain shifts in the image distribution within and across sites [35]. To assess the influence of these intrinsic shifts, we examined performance when the imaging operator was matched across sites and across the training-test sets (Table I). FedGIMP outperforms all competing methods ($p<0.05$), except on BRATS, R=6× where the privacy-violating benchmark GIMP performs similarly. On average across sites and R, FedGIMP yields 3.7dB PSNR, 3.0% SSIM improvement over the second-best FL method, demonstrating the efficacy of FedGIMP against heterogeneity in the MR image distribution.

2) Domain Shifts in the Imaging Operator: We then examined reconstruction performance under additional domain shifts due to the imaging operator. First, we considered homogeneous imaging operators across sites, while either acceleration rates (e.g., training at 3×, testing at 6×) or sampling densities (training with VD, testing with UD patterns) were mismatched across the training and test tests (Table II). FedGIMP is the top performer among all methods ($p<0.05$), except for GIMP that performs similarly. Compared to the second-best FL method, FedGIMP offers 4.8dB PSNR, 4.0% SSIM improvement under mismatched R, and 4.5dB PSNR, 4.8% SSIM improvement under mismatched sampling density. Representative reconstructions under mismatched sampling density are shown in Fig. 3. Federated conditional models and LORAKS yield prominent aliasing artifacts and blurring, whereas FedGIMP achieves high visual acuity and improved artifact suppression due to its enhanced generalization performance.

Next, we considered heterogeneous imaging operators across sites for improved flexibility in collaborations. We separately examined performance when the heterogeneous operators were matched or mismatched between the training-test sets (Table III). When the training-test operators match, FedGIMP achieves the highest performance among competing methods ($p<0.05$), except for GIMP that performs similarly, and in BRATS where LG-Fed performs slightly better and GAN_{cond}, FedGAN, FedMRI perform similarly. When the training-test operators mismatch, FedGIMP achieves the highest performance among competing methods ($p<0.05$), except GIMP that performs similarly. Compared to the second-best FL method, FedGIMP offers 1.3dB PSNR and 2.7% SSIM improvement under matched operators, and 5.1dB PSNR, 3.0% SSIM improvement under mismatched operators.

B. Multi-Coil Reconstruction

1) Domain Shifts in the MR Image Distribution: We first evaluated reconstruction performance when the imaging operator
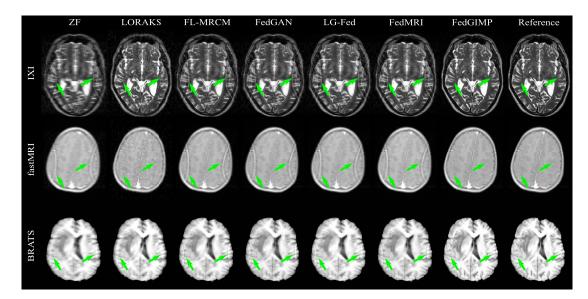
Fig. 3. Representative reconstructions of a T_2-weighted acquisition in IXI, a T_1c-weighted acquisition in fastMRI, and a T_1-weighted acquisition in BRATS datasets at R=3× based on uniform-density patterns. Results are shown for ZF, LORAKS, FL-MRCM, FedGAN, LG-Fed, FedMRI, and FedGIMP along with the reference images. Conditional models were trained on variable-density patterns at R=3×.
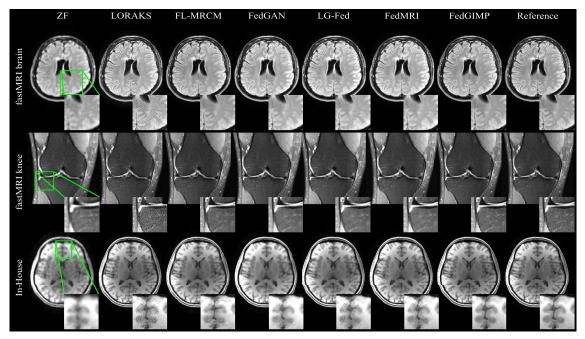

Fig. 4. Representative reconstructions of a FLAIR acquisition of the brain in fastMRI, a fat suppressed PD-weighted acquisition of the knee in fastMRI, and a T_1-weighted acquisition of the brain in In-House datasets at R=6×. Results are shown for ZF, LORAKS, FL-MRCM, FedGAN, LG-Fed, FedMRI and FedGIMP along with the reference images. Conditional models were trained at R=3×.

was matched across sites and training-test sets to assess intrinsic domain shifts due to the image distribution. Differing from single-coil experiments, multi-coil FL experiments were conducted on datasets containing two separate anatomies, i.e. the knee and the brain (Table IV). FedGIMP outperforms all competing methods (p<0.05), except GIMP that performs similarly. On average, FedGIMP achieves 6.3dB PSNR, 4.2% SSIM improvement over the second-best FL method. These results corroborate the efficacy of FedGIMP in multi-coil settings, and its efficiency in coping with heterogeneity in the image distribution due to diverse anatomy.

2) Domain Shifts in the Imaging Operator: We also evaluated reconstruction performance under additional domain shifts due to mismatched imaging operators with different acceleration rates across the training-test sets (Table V). FedGIMP is the top performer among competing methods (p<0.05), except for GIMP that performs similarly. FedGIMP offers 6.9dB PSNR, 4.5% SSIM improvement over the second-best FL method. Representative reconstructions are shown in Fig. 4. LORAKS suffers from noise amplification, and conditional models yield notable blur and artifacts. In contrast, FedGIMP achieves high visual acuity while mitigating noise amplification.

TABLE III
SINGLE-COIL RECONSTRUCTION PERFORMANCE WITH HETEROGENEOUS IMAGING OPERATORS ACROSS SITES. THE OPERATORS WERE EITHER MATCHED (UPPER PANEL) OR MISMATCHED (LOWER PANEL) ACROSS THE TRAINING-TEST SETS

		IXI (9x, VD)		fastMRI (3x, VD)		BRATS (6x, UD)	
		PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
		Matched across training-test sets					
Non-fed	LORAKS	25.0±1.6	58.8±6.5	29.1±4.6	70.4±11.3	24.7±2.4	85.6±1.4
	GAN_cond	33.4±1.1	88.8±1.7	39.0±2.2	94.9±1.3	33.0±1.2	93.2±1.0
	GIMP	35.5±1.3	95.0±1.3	40.6±2.2	97.1±1.0	33.4±1.4	93.6±1.0
Fed	FL-MRCM	32.9±1.0	88.5±1.5	37.6±2.3	93.1±1.7	30.5±1.3	89.9±1.4
	FedGAN	33.0±1.0	88.4±1.5	38.0±1.9	93.5±1.5	32.5±1.2	92.9±0.9
	LG-Fed	33.4±1.0	88.8±1.6	38.9±2.3	94.9±1.4	**33.9±1.3**	**94.2±0.9**
	FedMRI	33.4±1.0	88.7±1.5	38.9±2.2	94.8±1.3	33.0±1.2	93.2±1.0
	FedGIMP	**35.6±1.4**	**95.2±1.4**	**41.1±2.2**	**97.4±0.8**	33.4±1.4	93.4±1.0
		Mismatched across training-test sets					
		IXI (9x, VD)→(3x, UD)		fastMRI (3x, VD)→(6x, UD)		BRATS (6x, UD)→(3x, VD)	
		PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-fed	LORAKS	25.5±1.4	80.9±3.7	21.2±1.7	63.9±7.1	35.1±4.0	95.1±0.5
	GAN_cond	31.3±1.1	91.2±1.3	26.1±2.3	79.4±4.8	42.1±1.0	97.7±0.3
	GIMP	35.3±1.6	95.7±1.1	28.1±2.6	82.3±3.9	50.6±1.9	99.7±0.1
Fed	FL-MRCM	31.9±0.9	91.6±1.1	25.8±2.0	77.7±4.3	40.8±1.2	97.0±0.6
	FedGAN	32.3±1.0	92.5±1.1	25.9±2.1	79.3±4.5	40.6±1.3	96.7±0.7
	LG-Fed	31.5±1.0	91.1±1.2	24.9±1.9	75.8±4.8	38.0±1.0	94.6±0.8
	FedMRI	31.7±1.1	91.4±1.2	24.9±1.8	75.8±4.5	37.8±1.1	94.2±1.0
	FedGIMP	**35.1±1.4**	**95.4±1.0**	**28.0±2.5**	**82.3±4.0**	**50.9±1.9**	**99.7±0.1**

TABLE IV
MULTI-COIL RECONSTRUCTION PERFORMANCE WITH THE IMAGING OPERATOR MATCHED ACROSS SITES, AND THE TRAINING-TEST SETS

			fastMRI brain		fastMRI knee		In-House brain	
			PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-fed	LORAKS	3x	36.1±0.9	89.6±1.1	33.0±1.4	79.0±2.3	51.1±1.0	99.3±0.1
		6x	30.8±1.6	80.3±2.6	30.9±1.5	74.1±2.7	39.3±1.4	95.7±0.9
	GAN_cond	3x	40.1±0.9	95.6±0.6	36.1±1.2	89.9±1.6	43.5±1.0	98.7±0.2
		6x	36.3±0.9	91.9±0.9	33.2±1.1	83.1±2.4	38.8±0.9	96.9±0.5
	GIMP	3x	47.5±0.9	98.8±0.2	43.6±1.2	97.5±0.4	52.4±1.1	99.6±0.1
		6x	40.8±0.9	95.7±0.5	38.0±1.3	91.6±1.3	43.9±1.0	98.5±0.2
Fed	FL-MRCM	3x	39.4±0.9	95.8±0.5	36.9±1.3	90.6±1.3	40.4±0.9	97.5±0.4
		6x	35.9±0.8	92.2±0.8	33.8±1.2	82.1±2.4	36.0±0.8	95.0±0.7
	FedGAN	3x	39.8±0.9	95.6±0.5	37.0±1.2	90.0±1.5	41.7±1.1	98.1±0.3
		6x	35.9±0.8	92.1±0.9	34.0±1.2	82.7±2.5	36.7±0.9	95.7±0.7
	LG-Fed	3x	39.8±0.8	95.7±0.5	36.6±1.2	89.5±1.6	42.8±0.9	98.5±0.2
		6x	35.9±0.8	92.4±0.8	33.7±1.2	82.2±2.5	38.4±0.9	96.6±0.5
	FedMRI	3x	39.9±0.9	95.8±0.5	36.8±1.2	89.7±1.6	43.0±0.9	98.6±0.2
		6x	36.1±0.9	92.4±0.8	33.7±1.2	81.7±2.6	38.0±0.9	96.6±0.5
	FedGIMP	3x	**47.2±0.9**	**98.7±0.2**	**43.3±1.2**	**97.4±0.4**	**52.8±1.2**	**99.6±0.1**
		6x	**40.7±0.9**	**95.7±0.5**	**37.9±1.3**	**91.5±1.4**	**44.0±1.0**	**98.5±0.2**

TABLE V
MULTI-COIL RECONSTRUCTION PERFORMANCE WITH THE IMAGING OPERATOR MISMATCHED ACROSS TRAINING-TEST SETS

			fastMRI brain		fastMRI knee		In-House brain	
			PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Non-Fed	LORAKS	3x	36.1±0.9	89.6±1.1	33.0±1.4	79.0±2.3	51.1±1.0	99.3±0.1
		6x	30.8±1.6	80.3±2.6	30.9±1.5	74.1±2.7	39.3±1.4	95.7±0.9
	GAN_cond	3x	39.0±0.8	95.1±0.5	35.8±1.1	89.5±1.5	41.5±0.9	98.4±0.2
		6x	36.4±1.0	92.2±1.0	33.4±1.2	83.3±2.4	37.4±1.0	95.8±0.6
	GIMP	3x	47.5±0.9	98.8±0.2	43.6±1.2	97.5±0.4	52.2±1.1	99.7±0.1
		6x	40.8±0.9	95.7±0.5	38.0±1.3	91.6±1.3	43.6±1.0	98.4±0.2
Fed	FL-MRCM	3x	37.7±0.8	94.8±0.8	35.8±1.2	88.5±2.3	37.9±0.8	96.0±0.7
		6x	36.3±1.0	93.0±0.7	34.2±1.3	83.9±1.7	35.9±0.9	94.7±0.6
	FedGAN	3x	38.4±0.8	95.0±0.5	36.1±1.2	88.7±1.7	39.7±0.9	97.7±0.4
		6x	36.0±1.1	92.6±0.9	34.2±1.2	84.0±2.3	36.2±1.1	95.3±0.7
	LG-Fed	3x	38.6±0.7	95.1±0.5	35.8±1.2	88.2±1.7	41.3±0.8	98.1±0.3
		6x	36.2±1.0	92.8±0.9	34.0±1.2	83.5±2.4	37.4±1.1	96.1±0.6
	FedMRI	3x	38.9±0.8	95.1±0.5	35.9±1.2	88.0±1.8	40.8±0.8	98.2±0.3
		6x	36.2±1.0	92.8±0.9	34.1±1.2	83.3±2.4	37.6±1.0	95.9±0.6
	FedGIMP	3x	**47.2±0.9**	**98.7±0.2**	**43.3±1.2**	**97.4±0.4**	**52.8±1.2**	**99.6±0.1**
		6x	**40.7±0.9**	**95.7±0.5**	**37.9±1.3**	**91.5±1.4**	**44.0±1.0**	**98.5±0.2**

heterogeneity in each site. Importantly, inter-site heterogeneity is significantly higher than intra-site heterogeneity for all sites ($p<0.05$). The increase from intra-site to inter-site FID score is 9.2 for single-coil (a 94.1% increase) and 5.0 for multi-coil (a 105.0% increase) settings. We then examined the influence of this apparent heterogeneity on single-site models trained on local data from individual sites. Specifically, we assessed the loss incurred in single-site models due to domain shifts in the image distribution between training and test sites. To do this, we measured the within-site and across-site reconstruction performances of GAN_cond and GIMP while the imaging operator was matched across sites and training-test sets. We find that across-site performance is lower than within-site performance for both models ($p<0.05$). On average, GAN_cond suffers from 1.9dB PSNR, 2.0% SSIM loss in single-coil, and 2.5dB PSNR, 1.8% SSIM loss in multi-coil reconstruction. In contrast, GIMP shows relatively lower 1.5dB PSNR, 0.6% SSIM loss in single-coil, and 0.8dB PSNR, 0.2% SSIM loss in multi-coil reconstruction. Taken together, these findings indicate that our results cannot be attributed to a lack of data heterogeneity across sites. Furthermore, the adaptive GIMP model shows higher reliability against data heterogeneity compared to the non-adaptive GAN_cond model.

D. Ablation Studies

Several lines of ablation studies were conducted to demonstrate the individual design elements in FedGIMP. To assess the contribution of using an adaptive prior, we built a static-prior variant that optimized model inputs but kept model weights fixed as in [54]. To assess the federated training of the prior, we built an untrained variant where inference adaptation was performed on a randomly initialized generator. To assess the mapper, we built a mapper-ablated variant that only used a synthesizer as in [32]. To assess the site index, we built a site-index-ablated variant that used a mapper producing site-general latent variables. To assess the mapper adaptation to indirectly optimize latents, we built a static-mapper variant that directly optimized latents as in [55]. Table VI

C. Heterogeneity of MRI Data

FL of a multi-site model on homogeneous datasets would correspond to training on more samples from a common distribution, and model performance would solely reflect the influence of prolonged training. To rule out this nuisance explanation, we first measured the intra-site and inter-site heterogeneity of cross-sectional images in the training set. For intra-site heterogeneity, the image distributions were compared among individual subjects within a given site. For inter-site heterogeneity, the image distributions were compared among individual subjects from separate sites. Frechet inception distance (FID) was used to measure dissimilarity of distributions. As shown in Fig. 5, we find varying albeit notable levels of
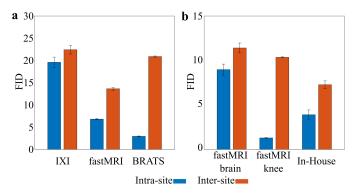
Fig. 5. Heterogeneity of MRI datasets were examined by measuring FID between respective image distributions, reported as mean±se (error bars) across subjects. Intra-site (blue) and inter-site (orange) heterogeneity is shown for each dataset examined for (a) single-coil, (b) multi-coil reconstruction.

demonstrates that FedGIMP outperforms all ablated variants consistently across sites. These results indicate the importance of prior adaptation, learning of a high-quality prior, inclusion of the mapper, inclusion of the site index, and indirect optimization of latents for improving MRI reconstruction performance. In particular, utilizing a site-index in the mapper achieves 1.1dB PSNR and 0.5% SSIM improvement, and in theory the benefits of the site index should increase further for growing number of sites and data heterogeneity.

Next, we evaluated the influence of the site index on practicality when a site joins or leaves training abruptly. Note that an additional site can join training by designating spare digits in the site-index vector, and an existing site can easily abstain from training given that its digit in the site-index vector is not reassigned to a different site. Thus, we examined FedGIMP's performance under scenarios where an initially held-out site joined the training halfway, and where an initially included site left the training halfway. In both scenarios, FedGIMP was compared against the site-index-ablated variant, and against the original model trained with all sites intact. Performance metrics are listed in Table VII. FedGIMP with joining or abstaining sites outperforms the ablated variant by 1.2dB PSNR, 0.6% SSIM on average; and it offers similar performance to the original FedGIMP trained with all sites intact. These results suggest that the site index does not limit practicality of model training, and it rather introduces a degree of reliability against joining or abstaining sites.

Lastly, we examined the utility of local discriminators in FedGIMP. A primary benefit of unshared discriminators is theoretical improvements in privacy against inference attacks [65]. Local discriminators also reduce communication load by 46.7% for FedGIMP, given that the discriminator and generator have comparable complexity (23.0 versus 26.3 million parameters). Beyond these theoretical benefits, we assessed the influence of local discriminators on model performance. First, we inspected the training losses for the generator (unreported). The loss curves showed smooth behavior across training epochs, with no apparent signs of mode collapse [56]. Second, we compared FedGIMP against a variant with shared discriminators across sites. Metrics in Table VII indicate on par performance with both models. These results

TABLE VI
RECONSTRUCTION PERFORMANCE OF FEDGIMP AND ABLATED VARIANTS AT R=3×, WITH MATCHED IMAGING OPERATORS ACROSS SITES

	IXI		fastMRI		BRATS	
	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
Static Prior	29.5±0.7	71.5±2.4	30.5±1.9	74.9±4.8	34.5±1.2	88.0±1.9
Untrained	34.7±1.0	88.6±1.9	35.4±2.0	90.3±2.4	38.7±1.4	95.4±1.1
No Mapper	40.1±1.2	97.0±0.7	38.8±2.0	95.7±1.2	47.2±1.8	99.2±0.3
No Site-Index	41.0±1.3	97.7±0.7	40.2±2.1	96.8±1.0	50.5±1.9	99.6±0.2
Static Mapper	42.3±1.4	98.2±0.5	40.7±2.2	97.1±0.9	50.6±1.9	99.6±0.1
FedGIMP	**42.9±1.5**	**98.5±0.5**	**41.1±2.2**	**97.4±0.8**	**50.9±1.9**	**99.7±0.1**

TABLE VII
RECONSTRUCTION PERFORMANCE AT R=3× WITH MATCHED IMAGING OPERATORS. RESULTS LISTED FOR FEDGIMP WITH ALL SITES, A JOINING SITE (FEDGIMP-JOIN), AND AN ABSTAINING SITE (FEDGIMP-ABS); FOR THE SITE-INDEX-ABLATED VARIANT WITH A JOINING SITE (NSI-JOIN) AND AN ABSTAINING SITE (NSI-ABS); AND FOR A VARIANT WITH SHARED DISCRIMINATORS. THE JOINING AND ABSTAINING SITES WERE TAKEN AS BRATS

	IXI		fastMRI		BRATS	
	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑	PSNR ↑	SSIM↑
FedGIMP	42.9±1.5	98.5±0.5	41.1±2.2	97.4±0.8	50.9±1.9	99.7±0.1
FedGIMP-join	42.7±1.4	98.4±0.5	41.0±2.2	97.3±0.9	50.6±1.9	99.7±0.1
FedGIMP-abs	43.0±1.6	98.5±0.5	41.0±2.3	97.3±0.9	50.9±2.0	99.7±0.1
NSI-join	40.9±1.2	97.5±0.6	39.9±2.2	96.6±1.1	49.8±1.9	99.6±0.2
NSI-abs	40.4±1.2	97.2±0.7	40.1±2.1	96.8±1.0	50.7±1.9	99.7±0.1
Shared D	42.9±1.5	98.4±0.5	41.2±2.2	97.4±0.8	51.1±1.9	99.7±0.1

TABLE VIII
AVERAGE INFERENCE TIME IN SECS PER CROSS-SECTION FOR SINGLE-COIL AND MULTI-COIL RECONSTRUCTIONS

	LORAKS	Non-adaptive	Adaptive
Single-coil	2.40	0.034	20.30
Multi-coil	6.99	0.080	27.89

suggest that there are no apparent performance drawbacks to using local discriminators.

E. Inference Times

Inference times of competing methods are listed in Table VIII. Matching architectures yield near identical inference time, so results are reported for LORAKS, non-adaptive models (GAN$_{cond}$, FL-MRCM, FedGAN, LG-Fed, FedMRI), and adaptive models (GIMP, FedGIMP). Inference time for adaptive models is high compared to non-adaptive models that use a single forward pass, and it is relatively more proximate to LORAKS. Yet, this computational load enables substantial benefits in reconstruction performance. For instance, FedGIMP achieves 10.4dB PSNR, 14.7% SSIM improvement over LORAKS, and 6.0dB PSNR, 3.1% SSIM improvement over the top-contending non-adaptive model across single- and multi-reconstruction tasks at R=3× with matched imaging operators across sites and training/test sets. Furthermore, FedGIMP enhances tissue depiction visually with lower artifacts and higher acuity than competing methods. Note also that FedGIMP's performance increases across inference iterations, so the number of iterations can be adjusted to maximize performance per available compute time in specific applications. Thus, FedGIMP offers a flexible

trade-off between reconstruction quality and compute time, which might facilitate practical use.

VI. DISCUSSION

Multi-site imaging data collected under diverse protocols/devices can contain heterogeneity in the image distribution and the imaging operator across sites, as well as across the training-test sets [51]. Recent studies have proposed latent-space alignment or split-network approaches based on conditional reconstruction models to address across-site heterogeneity [35], [36]. Yet, conditional models are susceptible to domain shifts in the imaging operator pertained to undersampled data [55]. In contrast, FedGIMP decouples the undersampling characteristics from the prior to improve reliability against heterogeneity in the imaging operator. Experiments on multi-site datasets demonstrate that FedGIMP yields superior performance against federated conditional models under various imaging scenarios with varying acceleration rates, sampling density across sites, and across the training-test sets. Therefore, FedGIMP can improve flexibility in multi-site collaborations by permitting heterogeneous protocols.

A practical concern for MRI reconstruction is the computational cost of training and inference. Here, we primarily considered adversarial architectures for competing methods. Previous conditional methods additionally compute either a cross-site alignment loss on latent-space representations [35], or a weighted-contrastive loss across sites based on local encoder weights [36]. Furthermore, while conditional models are retrained for each configuration of acceleration rate and sampling density, FedGIMP trains a single MRI prior to reconstruct with various different imaging operators. Thus, FedGIMP offers a simpler training procedure. During inference, conditional models offer fast processing in a single forward-pass, whereas FedGIMP uses an iterative prior adaptation that elevates computational burden. Still, FedGIMP's inference time is generally compatible with clinically-adopted implementations of traditional reconstruction methods, implying practical feasability [60], [66]. To improve efficiency in time-critical applications, prior adaptation across an MRI volume can be accelerated by parallel computation via distributing cross-sections across multiple GPUs.

GAN models are commonly trained to generate new image samples from low-dimensional random variables [56]. While the stochasticity of generated images may be undesirable for reconstruction, previous studies have employed GAN-based methods to recover MRI images without notable hallucinations [14], [16], [18], [32], [41], [42]. To prevent artificial structures, these methods have incorporated data-consistency modules. Similarly in FedGIMP, the subject-specific prior adaptation based on data-consistency constrains the reconstruction by incorporating information regarding coil sensitivities, sampling patterns and measured k-space data from the actual subject. Under this control procedure, we have not encountered artificial structures based on our visual inspections, as also indicated by high performance metrics across various datasets. Multi-site data analyzed here for single- and multi-coil reconstructions generally contained a considerable

number of subjects. While the In-House dataset had relatively fewer subjects, we analyzed it given the scarcity of public datasets providing multi-coil, multi-contrast MRI data. It remains important future work to systematically validate the proposed method and its anatomical fidelity on broader patient cohorts.

FL methods transfer model weights across sites instead of MRI data to lower privacy risks. Still, security concerns can arise from backdoor attacks where an adversary poisons training updates to corrupt models and elicits diagnostically-inaccurate reconstructions [22]. Previously proposed non-adaptive models freezing model weights following training can be more sensitive to model corruption. In contrast, FedGIMP adapts its MRI prior to each test sample, so it can potentially alleviate corruption during inference [59]. Learning-based models can also be vulnerable to inference attacks on models aiming to leak sensitive information about training data [22]. Differential privacy between training and synthetic samples in adversarial models substantially improves for large and diverse training datasets as encountered in FL settings [67]. Furthermore, FedGIMP uses a shared generator without direct access to data and unshared discriminators that are not communicated [68]. Nonetheless, resilience against inference attacks can be improved by adopting differential-privacy procedures [69], [70]. Future studies are warranted to systematically characterize the privacy-preserving abilities of FedGIMP, including the benefits of local discriminators.

Here we aggregated local models using the relatively common and simple FedAvg algorithm, which has been successfully employed in many previous FL studies [29]. Model aggregation was performed via a weighted linear combination, where the weight for each site is taken as the ratio of the number of local training samples to the number of all training samples across sites. Performance improvements might be possible with data-driven combination based on model comparisons [71] or neural networks [72]. However, data-driven model combination typically involves storage and computation of additional models, and transfer of additional model updates. Future work is warranted to examine the comparative benefits of unlearned versus learned aggregation of local models in MRI reconstruction.

FedGIMP trains an MRI prior over the distribution of high-quality MR images, uninformed regarding the reconstruction task. Several lines of development can be considered for improved performance. First, multi-supervised networks concurrently operating in image and k-space domains have recently been introduced for MRI reconstruction [11], [15]. FedGIMP might also benefit from a hybrid of image and k-space MR priors. Second, self-supervised training of parallel networks based on contrastive learning has been proposed [46], [48]. Contrastive learning strategies might alleviate the need for fully-sampled acquisitions for training the MRI prior in FedGIMP. In principle, the federated prior can be adapted for other inverse problems such as MRI super-resolution or synthesis. A simple approach would be to train task-specific models using a synthetic dataset generated via the MRI prior. Alternatively, the trained prior can serve

as a non-adaptive plug-and-play regularizer in optimization problems [54]. To adapt the prior during inference, the imaging operator for the reconstruction task would have to be replaced with corresponding operators in target tasks.

VII. Conclusion

Here, we introduced a novel MRI reconstruction based on federated learning of a generative MRI prior and inference adaptation via injection of subject-specific imaging operators onto this prior. Benefits over state-of-the-art federated and traditional methods were demonstrated in multi-site MRI datasets. Improved generalization against domain shifts renders FedGIMP a promising candidate for multi-site collaborations in accelerated MRI. FedGIMP might also be used for physics-based reconstruction in other modalities such as CT, PET, or ultrasound by modifying its imaging operator.

References

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[2] B. Zhao, W. Lu, T. K. Hitchens, F. Lam, C. Ho, and Z.-P. Liang, "Accelerated MR parameter mapping with low-rank and sparsity constraints," *Magn. Reson. Med.*, vol. 74, no. 2, pp. 489–498, 2015.

[3] S. Wang et al., "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 514–517.

[4] K. Hammernik et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2017.

[5] K. Kwon, D. Kim, and H. Park, "A parallel MR imaging method using multilayer perceptron," *Med. Phys.*, vol. 44, no. 12, pp. 6209–6224, Dec. 2017.

[6] S. U. H. Dar, M. Özbey, A. B. Çatlı, and T. Çukur, "A transfer-learning approach for accelerated mri using deep neural networks," *Magn. Reson. Med.*, vol. 84, no. 2, pp. 663–685, 2020.

[7] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magn. Reson. Med.*, vol. 80, no. 3, pp. 1189–1205, Sep. 2018.

[8] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 42, no. 3, pp. 521–538, Mar. 2020.

[9] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "Scan-specific robust artificial-neural-networks for K-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 439–453, Jun. 2019.

[10] L. Xiang et al., "Deep-learning-based multi-modal fusion for fast MR reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2105–2114, Jul. 2019.

[11] S. Wang et al., "DIMENSION: Dynamic MR imaging with both K-space and spatial prior knowledge obtained via multi-supervised network training," *NMR Biomed.*, vol. 35, no. 4, p. e4131, Apr. 2022.

[12] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for MR image reconstruction," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 647–658.

[13] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2018.

[14] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction with cyclic loss in generative adversarial Networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1488–1497, Aug. 2018.

[15] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, and D. Hwang, "KIKI-Net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2188–2201, 2018.

[16] M. Mardani et al., "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[17] D. Polak et al., "Joint multi-contrast variational network reconstruction (jVN) with application to rapid 2D and 3D imaging," *Magn. Reson. Med.*, vol. 84, no. 3, pp. 1456–1469, Sep. 2020.

[18] S. U. H. Dar, M. Yurt, M. Shahdloo, M. E. Ildiz, B. Tinaz, and T. Cukur, "Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1072–1087, Oct. 2020.

[19] C.-M. Feng, Z. Yang, H. Fu, Y. Xu, J. Yang, and L. Shao, "DONet: Dual-octave network for fast MR image reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 1, 2021, doi: 10.1109/TNNLS.2021.3090303.

[20] P. Guo, J. M. J. Valanarasu, P. Wang, J. Zhou, S. Jiang, and V. M. Patel, "Over-and-under complete convolutional RNN for MRI reconstruction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2021, pp. 13–23.

[21] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 141–151, Jan. 2020.

[22] G. A. Kaissis, M. R. Makowski, D. Rückeert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Mach. Intell.*, vol. 2, no. 6, pp. 305–311, Jun. 2020.

[23] W. Li et al., "Privacy-preserving federated brain tumour segmentation," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2019, pp. 133–141.

[24] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, 2019, pp. 92–104.

[25] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, p. 119, Sep. 2020.

[26] H. R. Roth et al., "Federated learning for breast density classification: A real-world implementation," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*. Switzerland: Springer, 2020, pp. 18–191.

[27] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101765.

[28] Q. Liu, C. Chen, J. Qin, Q. Dou, and P. Heng, "FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1013–1023.

[29] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[30] J. Wang et al., "A field guide to federated optimization," 2021, *arXiv:2107.06917.*

[31] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[32] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, "Inverse GANs for accelerated MRI reconstruction," in *Proc. SPIE*, vol. 11138, Sep. 2019, pp. 381–392.

[33] F. Knoll et al., "FastMRI: A publicly available raw K-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiology: Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190007.

[34] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson, "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 116–128, 2019.

[35] P. Guo, P. Wang, J. Zhou, S. Jiang, and V. M. Patel, "Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning," 2021, *arXiv:2103.02148.*

[36] C.-M. Feng, Y. Yan, S. Wang, Y. Xu, L. Shao, and H. Fu, "Specificity-preserving federated learning for MR image reconstruction," 2021, *arXiv:2112.05752.*

[37] B. Zhu, J. Z. Liu, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[38] S. Biswas, H. K. Aggarwal, and M. Jacob, "Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM," *Magn. Reson. Med.*, vol. 82, no. 1, pp. 485–494, 2019.

[39] Q. Chang et al., "Synthetic learning: Learn from distributed asynchronized discriminator GAN without sharing medical image data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13853–13863.

[40] P. Guo, P. Wang, R. Yasarla, J. Zhou, V. M. Patel, and S. Jiang, "Anatomic and molecular MR image synthesis using confidence guided CNNs," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2832–2844, Oct. 2021.

[41] G. Oh, B. Sim, H. Chung, L. Sunwoo, and J. C. Ye, "Unpaired deep learning for accelerated MRI using optimal transport driven CycleGAN," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1285–1296, 2020.

[42] K. Lei, M. Mardani, J. M. Pauly, and S. S. Vasanawala, "Wasserstein GANs for MR imaging: From paired to unpaired training," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 105–115, Jan. 2021.

[43] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Ugurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magn. Reson. Med.*, vol. 84, no. 6, pp. 3172–3191, Dec. 2020.

[44] P. Huang et al., "Deep MRI reconstruction without ground truth for training," in *Proc. 27th Annu. Meeting (ISMRM)*, 2019, p. 4668.

[45] J. I. Tamir, S. X. Yu, and M. Lustig, "Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data," in *Proc. Int. Soc. Magn. Reson. Med.*, 2019, p. 0660.

[46] C. Hu, C. Li, H. Wang, Q. Liu, H. Zheng, and S. Wang, *Self-Supervised Learning for MRI Reconstruction With a Parallel Network Training Framework*. Cham, Switzerland: Springer, 2021, pp. 382–391.

[47] F. Liu, R. Kijowski, G. El Fakhri, and L. Feng, "Magnetic resonance parameter mapping using model-guided self-supervised deep learning," *Magn. Reson. Med.*, vol. 85, no. 6, pp. 3211–3226, Jun. 2021.

[48] S. Wang et al., "PARCEL: Physics-based unsupervised contrastive representation learning for parallel MR imaging," 2022, *arXiv:2202.01494*.

[49] S. C. Karayumak et al., "Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters," *NeuroImage*, vol. 184, pp. 180–200, Jan. 2019.

[50] B. E. Dewey et al., "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes," *Magn. Reson. Imag.*, vol. 64, pp. 160–170, Dec. 2019.

[51] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "FedBN: Federated learning on non-IID features via local batch normalization," 2021, *arXiv:2102.07623*.

[52] Z. Yan, J. Wicaksana, Z. Wang, X. Yang, and K.-T. Cheng, "Variation-aware federated learning with multi-source decentralized medical image data," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2615–2628, Jul. 2021.

[53] S. Park, G. Kim, J. Kim, B. Kim, and J. C. Ye, "Federated split task-agnostic vision transformer for COVID-19 CXR diagnosis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 1–14.

[54] K. C. Tezcan, C. F. Baumgartner, R. Luechinger, K. P. Pruessmann, and E. Konukoglu, "MR image reconstruction using deep density priors," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1633–1642, Jul. 2019.

[55] Y. Korkmaz, S. U. H. Dar, M. Yurt, M. Ozbey, and T. Cukur, "Unsupervised MRI reconstruction via zero-shot learned adversarial transformers," *IEEE Trans. Med. Imag.*, vol. 41, no. 7, pp. 1747–1763, Jul. 2022.

[56] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[57] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," 2017, *arXiv:1703.06868*.

[58] M. Uecker et al., "ESPIRiT—An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magn. Reson. Med.*, vol. 71, no. 3, pp. 990–1001, Mar. 2014.

[59] H. Kumar Aggarwal and M. Jacob, "Model adaptation for image reconstruction using generalized stein's unbiased risk estimator," 2021, *arXiv:2102.00047*.

[60] J. P. Haldar and J. Zhuo, "P-LORAKS: Low-rank modeling of local K-space neighborhoods with parallel imaging data," *Magn. Reson. Med.*, vol. 75, no. 4, p. 1499, 2016.

[61] M. Rasouli, T. Sun, and R. Rajagopal, "FedGAN: Federated generative adversarial networks for distributed data," 2020, *arXiv:2006.07228*.

[62] P. Pu Liang et al., "Think locally, act globally: Federated learning with local and global representations," 2020, *arXiv:2001.01523*.

[63] B. H. Menze and et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2004.

[64] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "Coil compression for accelerated imaging with Cartesian sampling," *Magnetic Reson. Med.*, vol. 69, no. 2, pp. 82–571, 2013.

[65] D. Chen, N. Yu, Y. Zhang, and M. Fritz, "GAN-leaks: A taxonomy of membership inference attacks against generative models," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA: ACM, Oct. 2020, p. 343.

[66] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast ℓ_1-SPIRiT compressed sensing parallel imaging MRI: Scalable parallel implementation and clinically feasible runtime," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1250–1262, Jun. 2012.

[67] Q. Feng, C. Guo, F. Benitez-Quiroz, and A. Martinez, "When do GANs replicate? On the choice of dataset size," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 6701–6710.

[68] T. Han et al., "Breaking medical data sharing boundaries by using synthesized radiographs," *Sci. Adv.*, vol. 6, no. 49, Dec. 2020, Art. no. eabb7973.

[69] L. Zhang, B. Shen, A. Barnawi, S. Xi, N. Kumar, and Y. Wu, "FedDP-GAN: Federated differentially private generative adversarial networks framework for the detection of COVID-19 pneumonia," *Inf. Syst. Frontiers*, vol. 23, no. 6, pp. 1403–1415, Dec. 2021.

[70] A. Ziller, D. Usynin, M. Knolle, K. Hammernik, D. Rueckert, and G. Kaissis, "Complex-valued deep learning with differential privacy," 2021, *arXiv:2110.03478*.

[71] M. Zhang, K. Sapra, S. Fidler, S. Yeung, and J. M. Alvarez, "Personalized federated learning with first order model optimization," 2021, *arXiv:2012.08565*.

[72] X. Ma, J. Zhang, S. Guo, and W. Xu, "Layer-wised model aggregation for personalized federated learning," 2022, *arXiv:2205.03993*.