Medical Image Analysis 78 (2022) 102429

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Progressively volumetrized deep generative models for data-efficient contextual learning of MR image recovery

Mahmut Yurt[a,b,1], Muzaffer Özbey[a,b,1], Salman U.H. Dar[a,b], Berk Tinaz[a,b,c], Kader K. Oguz[b,d], Tolga Çukur[a,b,e,*]

[a] Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey
[b] National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara 06800, Turkey
[c] Department of Electrical and Computer Engineering, University of Southern California, Los Angeles 90089, USA
[d] Department of Radiology, Hacettepe University, Ankara 06100, Turkey
[e] Neuroscience Program, Bilkent University, Ankara 06800, Turkey

ARTICLE INFO

Article history:
Received 10 February 2021
Revised 14 March 2022
Accepted 18 March 2022
Available online 26 March 2022

Keywords:
MRI
Generative adversarial networks
Synthesis
Reconstruction
Cross-section
Model complexity
Context

ABSTRACT

Magnetic resonance imaging (MRI) offers the flexibility to image a given anatomic volume under a multitude of tissue contrasts. Yet, scan time considerations put stringent limits on the quality and diversity of MRI data. The gold-standard approach to alleviate this limitation is to recover high-quality images from data undersampled across various dimensions, most commonly the Fourier domain or contrast sets. A primary distinction among recovery methods is whether the anatomy is processed per volume or per cross-section. Volumetric models offer enhanced capture of global contextual information, but they can suffer from suboptimal learning due to elevated model complexity. Cross-sectional models with lower complexity offer improved learning behavior, yet they ignore contextual information across the longitudinal dimension of the volume. Here, we introduce a novel progressive volumetrization strategy for generative models (ProvoGAN) that serially decomposes complex volumetric image recovery tasks into successive cross-sectional mappings task-optimally ordered across individual rectilinear dimensions. ProvoGAN effectively captures global context and recovers fine-structural details across all dimensions, while maintaining low model complexity and improved learning behavior. Comprehensive demonstrations on mainstream MRI reconstruction and synthesis tasks show that ProvoGAN yields superior performance to state-of-the-art volumetric and cross-sectional models.

1. Introduction

Magnetic resonance imaging (MRI) is a clinically preferred modality that produces volumetric images of a given anatomy under diverse tissue contrasts (Bauer et al., 2013). As MR acquisitions are intrinsically slow, there has been persistent interest in recovery methods to improve quality and diversity of images derived from accelerated imaging protocols (Wang et al., 2020b; Ye, 2019). Two mainstream MRI recovery problems with pervasive applications are reconstruction and synthesis (Griswold et al., 2002; Pruessmann et al., 1999; Lustig et al., 2008; Jog et al., 2015; Yurt et al., 2021; Van Nguyen et al., 2015; Roy et al., 2016; 2013). While reconstruction aims to recover high-quality images from undersampled k-space acquisitions (Lustig et al., 2007), synthesis aims to recover high-quality images of unacquired tissue contrasts from images of collected contrasts (Roy et al., 2011). Learning-based models have offered performance leaps in both recovery tasks, given their ability to solve inverse problems (Yi et al., 2019; Litjens et al., 2017; Choi et al., 2020). However, the trade-off between sensitivity to spatial context and model complexity introduces a dilemma regarding the use of volumetric versus cross-sectional recovery models (Singh et al., 2020). The primary aim of this study is to introduce a novel volumetrization approach to achieve the contextual sensitivity of volumetric models while maintaining on par complexity with cross-sectional models.

Among learning-based models, a native recovery approach is to perform a single-shot global mapping between source and target volumes (Lan et al., 2020; Malavé et al., 2020; Yang et al., 2020; Yu et al., 2018; 2019; Chen et al., 2021a; Sood et al., 2021; Küstner et al., 2020; El-Rewaidy et al., 2020; Chong and Ho, 2021). Volumetric models leverage spatial correlations across all dimensions to better capture contextual information (Lan et al., 2020; Malavé et al., 2020; Yang et al., 2020; Yu et al., 2018). Introduc-

* Corresponding author at: Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey.
E-mail address: cukur@ee.bilkent.edu.tr (T. Çukur).
[1] Denotes equal contribution.

https://doi.org/10.1016/j.media.2022.102429
1361-8415/

tion of these contextual priors can theoretically lead to more consistent and accurate recovery across the volume. However, three-dimensional (3D) models involve substantially more parameters than their two-dimensional (2D) counterparts (Singh et al., 2020; Yurt et al., 2021). Furthermore, each volume constitutes a single training sample for a 3D model, whereas it would yield several tens of samples for a 2D model. Taken together, these factors render heavier demand for training data and impair the learning process for volumetric models (Singh et al., 2020).

A less demanding approach in terms of training data for learning-based MRI recovery is to perform a spatially-localized mapping between individual cross-sections (Akçakaya et al., 2019; Chartsias et al., 2018; Chen et al., 2021c; Hyun et al., 2018; Joyce et al., 2017; Lee et al., 2020; Li et al., 2021; Mardani et al., 2019; Quan et al., 2018; Schlemper et al., 2017; Wang et al., 2021; Zhan et al., 2021a,b; Cheng et al., 2018; Mardani, et al., 2017). Volumes are split along a specific rectilinear orientation, and cross-sectional models are then trained to learn the 2D mapping (Dar et al., 2019; 2020; Hammernik et al., 2017; Han et al., 2018; Lee et al., 2019; Sevetlidis et al., 2016). Since a lower-dimensional mapping is to be learned, cross-sectional models are of lower complexity and have reduced demand for training data (Yurt et al., 2021). This facilitates the learning process, and often results in more detailed mappings along the transverse dimensions within cross-sections compared to 3D models. Yet, 2D models do not fully utilize context across the longitudinal dimension, even when simultaneously processing multiple neighboring cross-sections (Yang et al., 2020; Yu et al., 2018; Dar et al., 2019). This results in inconsistency and errors across separately recovered cross-sectional images (Lan et al., 2020; Malavé et al., 2020; Yu et al., 2019).

An effective alternative to either approach is to build hybrid architectures that bridge 2D and 3D models. A group of studies in this domain have proposed aggregated models that fuse the outputs of parallel streams, where the streams are cross-sectional models in three orthogonal orientations (Wei et al., 2019; Peng et al., 2020). Pseudo target volumes are first recovered separately by the 2D streams, and a 3D fusion network then produces the final target volume (Wei et al., 2019; Peng et al., 2020). Other studies have instead proposed transfer of learned model weights from 2D to 3D models (Shan et al., 2018; Liu et al., 2018). A 2D model is first pretrained for a cross-sectional recovery task at a selected orientation, the learned weights are then used to initialize the convolutional kernels in 3D models (Shan et al., 2018; Liu et al., 2018). While both approaches can improve learning behavior, they involve a volumetric processing component that elevates memory requirements and places practical constraints on model complexity, potentially limiting sensitivity to detailed image features.

Here, we propose a novel progressive volumetrization strategy for deep generative models (ProvoGAN) for contextual learning of MR image recovery. To improve learning efficiency by lowering model complexity, ProvoGAN serially decomposes volumetric recovery tasks into a sequence of cross-sectional subtasks (e.g., axial, coronal, sagittal) for the first time in literature. [2] For a given subtask in a selected orientation, the source volume is split across the respective longitudinal dimension, and a 2D model is trained to map between cross-sectional source and target images. The predicted pseudo cross-sections are reformatted into a volume and then input to the next subtask as spatial priors (Fig. 1). This progressive nature empowers ProvoGAN to recover fine-structural details in each orientation while ensuring contextual consistency across the volume. Furthermore, the progression order of the subtasks is adaptively optimized to enhance task-specific performance.

[2] We presented a preliminary conception of the idea in the IEEE International Symposium on Biomedical Imaging (ISBI) on April 4, 2020.

To ensure a high degree of realism, we primarily employ ProvoGAN to volumetrize a recent conditional generative adversarial network based on the ResNet architecture (Dar et al., 2019). Note that ProvoGAN can be viewed as a model-agnostic strategy, so it can be extended to volumetrize other 2D network models as also demonstrated here. Comprehensive demonstrations are provided for mainstream reconstruction and synthesis tasks in multi-contrast MRI protocols. Our results indicate that ProvoGAN yields enhanced recovery performance compared to cross-sectional, volumetric, and hybrid approaches in terms of image quality. Importantly, ProvoGAN maintains these performance benefits while at the same time offering reduced model complexity and improved learning behavior.

Contributions

- To our knowledge, ProvoGAN is the first volumetrized model for MRI recovery that serially decomposes a global 3D mapping into a sequence of progressive 2D mappings.
- ProvoGAN maximizes task performance via adaptive ordering of the progression sequence of 2D mappings across rectilinear orientations.
- ProvoGAN embodies a model-agnostic learning strategy, so it can be implemented to volumetrize various 2D network architectures.
- Demonstrations on mainstream reconstruction and synthesis tasks indicate that ProvoGAN yields superior performance to several prior 2D, 3D and hybrid models.

2. Methods

2.1. Generative adversarial networks

Generative adversarial networks (GAN) are generative models composed of two subnetworks. The first subnetwork is a generator (G) that aims to synthesize fake samples closely mimicking a target data distribution, while the second subnetwork is a discriminator (D) that aims to detect whether a given data sample has been drawn from the target distribution or not (Goodfellow et al., 2014). These subnetworks are trained alternately in a two player zero-sum min-max game in an adversarial setup:

$$L_{\mathrm{GAN}} = \mathrm{E}_y[\log(D(y))] + \mathrm{E}_z[\log(1 - D(G(z)))] \quad (1)$$

where L_{GAN} is the adversarial loss function, E denotes expectation, z denotes a random noise vector sampled from a prior distribution, and y denotes an arbitrary real sample drawn from the target domain. In practice, the log-likelihood terms are replaced with squared-loss terms to improve stability (Mao et al., 2017):

$$L_{\mathrm{GAN}} = -\mathrm{E}_y[(D(y) - 1)^2] - \mathrm{E}_z[D(G(z))^2] \quad (2)$$

where D is trained to maximize L_{GAN}, whereas G is trained to minimize it.

While the basic GAN model synthesizes target data samples given a random noise input, recent studies on computer vision (Isola et al., 2017; Zhu et al., 2017) and medical imaging (Dar et al., 2019; Lee et al., 2019; Li et al., 2019; Olut et al., 2018; Sharma and Hamarneh, 2019; Yang et al., 2018b; Yu et al., 2018; 2019) have demonstrated that conditional GAN (cGAN) models (Mirza and Osindero, 2014) are highly effective in image-to-image translation tasks. The central aim in these tasks is to synthesize data samples from the target image domain, given data samples from a separate source image domain. The cGAN model is therefore modified to condition both G and D on the source domain image:

$$L_{\mathrm{cGAN}} = -\mathrm{E}_{x,y}[(D(x, y) - 1)^2] - \mathrm{E}_x[D(x, G(x))^2] \quad (3)$$

where x denotes the source domain image, and y denotes the target domain image. When paired images from the source and target

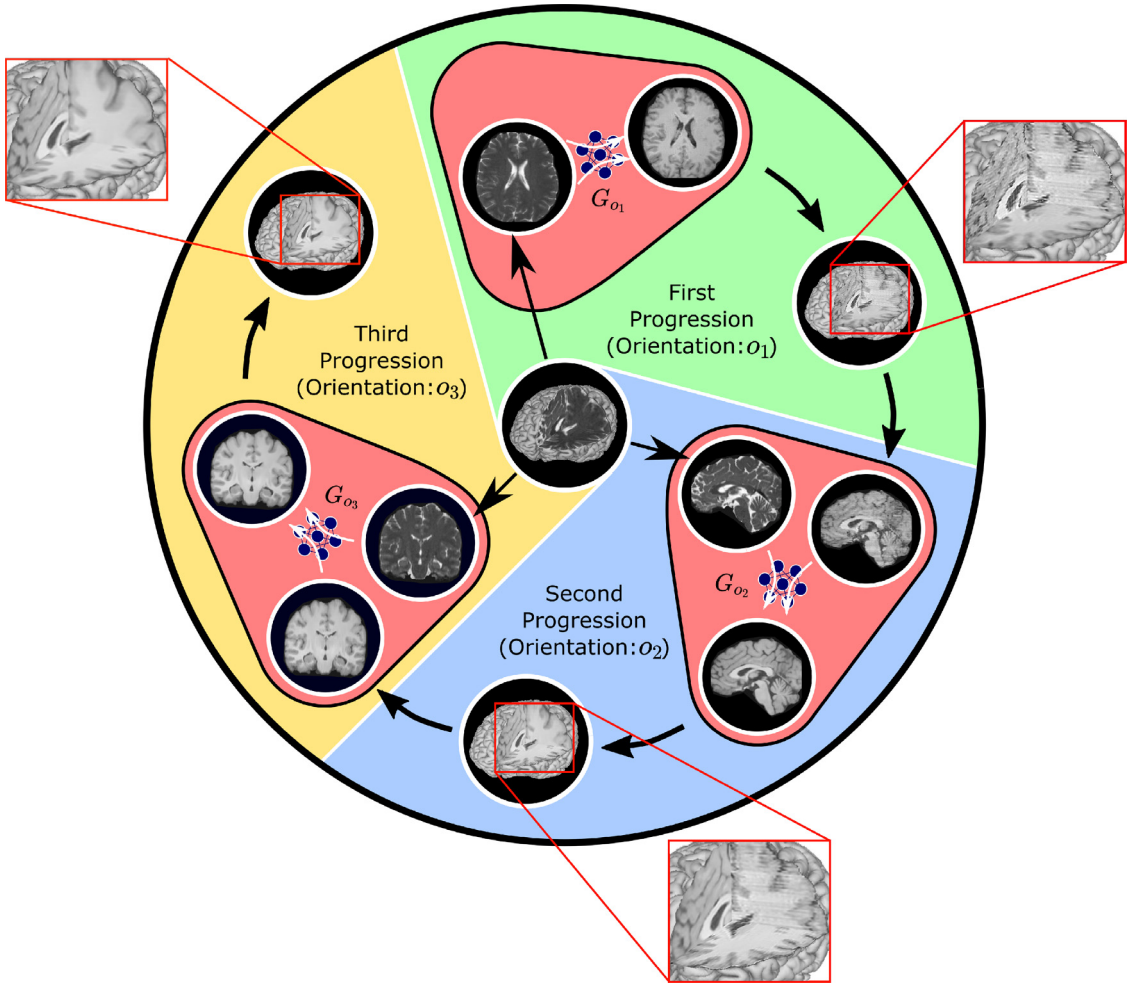

Fig. 1. ProvoGAN decomposes complex volumetric image recovery tasks into a cascade of progressive cross-sectional subtasks defined across the rectilinear orientations (axial, coronal, and sagittal). Given a specific order of progression sequence (axial → sagittal → coronal is given here for demonstration), ProvoGAN first learns a cross-sectional mapping in the first orientation, and processes cross-sections within the entire source volume to estimate the target volume. This volumetric estimate is then divided into cross-sections in the second orientation, and a separate cross-sectional model is learned in the second orientation. The volumetric estimate from the second progression is then fed onto the final progression in which a third cross-sectional model is learned for final recovery (see Supp. Fig. 1 for further details). The sequential implementation of the progressive cross-sectional models enables ProvoGAN to gradually improve capture of fine-structural details in each orientation and to ensure global contextual consistency within the volume while at the same time manifesting reduced model complexity and improved learning behavior of cross-sectional mapping.

domains are available, a pixel-wise loss between the ground truth and synthesized images can also be included:

$$L_{\text{cGAN}} = -\text{E}_{x,y}[(D(x,y)-1)^2] - \text{E}_x[D(x,G(x))^2] + \text{E}_{x,y}[||y - G(x)||_1] \quad (4)$$

The pixel-wise loss is typically based on the mean-absolute error to reduce sensitivity to outliers and alleviate undesirable smoothing. The mapping learned by the cGAN model grows more accurate as the statistical dependence between source and target domains gets stronger (Yurt et al., 2021).

2.2. MR image recovery via volumetric GANs

As MR images are intrinsically volumetric, a comprehensive approach for 3D MR image recovery is to use volumetric GAN (vGAN) models that perform a global mapping between source and target volumes (Lan et al., 2020; Malavé et al., 2020; Yang et al., 2020; Yu et al., 2018). To learn this mapping, vGAN models commonly employ complex generator G_V and discriminator D_V modules containing 3D convolutional kernels. The loss function is defined over the entire volume in an adversarial setup:

$$L_{\text{vGAN}} = -\text{E}_{X,Y}[(D_V(X,Y)-1)^2] - \text{E}_X[D_V(X,G_V(X))^2] \quad (5)$$

$$+ \text{E}_{X,Y}[||Y - G_V(X)||_1] \quad (6)$$

where X denotes the source and Y denotes the target volumetric images. For MRI reconstruction, X is typically the Fourier reconstruction of undersampled acquisitions, and Y is the fully-sampled reference volume. For MRI synthesis, X is the source contrast volume, and Y is the target contrast volume. Note that, in MRI reconstructions, an additional constraint is introduced to enforce consistency of acquired and recovered k-space data:

$$F_u(G_V(X)) := F_u(X) \quad (7)$$

where F_u denotes the partial Fourier operator that is defined at the acquired k-space points.

Due to their 3D nature, vGAN models can better incorporate contextual information across MRI volumes by leveraging spatial correlations across separate cross-sections (Yang et al., 2020; Yu et al., 2018; 2019). This contextual prior can lead to elevated consistency across the volume and increased accuracy in recovery performance. That said, learning in 3D network models is inherently more difficult since they involve substantially more parameters

(Singh et al., 2020). The learning process might be further impaired by data scarcity as the entire volume of each subject is taken as a single training sample (Singh et al., 2020). These limitations often cause vGAN models to settle on suboptimal parameter sets, compromising recovery performance.

2.3. MR image recovery via cross-Sectional GANs

A more focused approach for 3D MRI recovery is based on cross-sectional GAN (sGAN) models that perform localized mappings between 2D cross-sectional images within source and target volumes (Dar et al., 2019; Yurt et al., 2021; Shin et al., 2018; Sharma and Hamarneh, 2019). These 2D images are typically taken to be individual cross-sections within the volume in a specific rectilinear orientation, i.e., axial, sagittal or coronal. To learn this 2D mapping, sGAN models employ relatively simpler generator G_S and discriminator D_S modules containing 2D convolutional kernels. The loss function is defined for individual cross-sections in an adversarial setup with a pixel-wise loss:

$$L_{\text{sGAN}} = -\mathrm{E}_{x_o^i, y_o^i}[(D_S(x_o^i, y_o^i) - 1)^2] - \mathrm{E}_{x_o^i}[D_S(x_o^i, G_S(x_o^i))^2] + \mathrm{E}_{x_o^i, y_o^i}[||y_o^i - G_S(x_o^i)||_1] \tag{8}$$

where x_o^i and y_o^i denote the ith cross-sections within the source and target volumes in orientation o. As with sGAN models, $x_o^i - y_o^i$ are taken as cross-sectional images for undersampled and fully-sampled acquisitions in MRI reconstruction, and $x_o^i - y_o^i$ are taken as cross-sectional images of source and target contrasts in MRI synthesis. Consistency between acquired and recovered data can again be enforced during reconstruction via the following procedure:

$$F_u(G_S(x_o^i)) := F_u(x_o^i) \tag{9}$$

where F_u denotes the partial Fourier operator that is defined at the acquired k-space points. Once the mapping between the source and target cross-sections is learned, cross-sections of the target volumes are independently generated, and then the target volumes are recovered by concatenating the generated cross-sections.

Due to their 2D nature, sGAN models are less complex and so they naturally have lower demand for data Yurt et al. (2021). Individual cross-sections within a subject's volume are taken as separate training samples, expanding the effective size of the dataset. As a result, more detailed cross-sectional mapping can be learned. However, this advantage comes at the expense of neglecting global contextual information across the volume (Yang et al., 2020; Yu et al., 2018; 2019). Therefore, sGAN models might suffer from inconsistency or inaccuracy of recovered images across cross-sections.

2.4. Progressively volumetrized GAN

Here, a novel architecture is proposed to address the limitations of volumetric and cross-sectional GAN models. The proposed model, named progressively volumetrized GAN (ProvoGAN), decomposes complex volumetric image recovery tasks into a series of simpler cross-sectional tasks (Fig. 1). The cross-sectional recovery tasks are defined in separate orientations, and are implemented sequentially via cascaded 2D GAN models. We consider rectilinear cross-sections of volumetric MRI datasets in this study, so the selected orientations are axial, coronal and sagittal. Given a specific order of the three orientations (o_1, o_2, o_3), ProvoGAN first learns a 2D recovery model in orientation o_1. The entire source volume is processed by this model to estimate the target volume. Afterwards, this volumetric estimate is separated into cross-sections in orientation o_2, and a separate 2D recovery model is trained. The estimated target volume for o_2 is then fed onto the final stage, where a third 2D recovery model is trained in orientation o_3.

The cascaded 2D models in ProvoGAN are trained sequentially in the three rectilinear orientations, where the 2D model weights at earlier orientations are frozen upon training. This learning strategy empowers ProvoGAN to progressively recover fine-structural details at each orientation, while bypassing the need for computationally expensive calculation of error gradients across the entire volume and across all orientations. Therefore, ProvoGAN offers the ability to efficiently capture global contextual information without drastically elevating computational demand. At the same time, this step-wise training can increase sensitivity to progression order. Therefore, progression order across orientations is adaptively tuned to maximize performance in specific tasks. Detailed formulation of the ProvoGAN model is provided below.

First progression: ProvoGAN first learns a cross-sectional mapping between the source-target volumes in o_1 via a generator (G_{o_1}) and a discriminator (D_{o_1}). The source and target cross-sections in o_1 are extracted with a division block (d_{o_1}).

$$\begin{aligned} x_{o_1}^i &\in \{x_{o_1}^1, x_{o_1}^2 \ldots, x_{o_1}^I\} = d_{o_1}(X) \\ y_{o_1}^i &\in \{y_{o_1}^1, y_{o_1}^2 \ldots, y_{o_1}^I\} = d_{o_1}(Y) \end{aligned} \tag{10}$$

where X denotes the source volume, Y denotes the target volume, $x_{o_1}^i$ denotes the ith cross-section of the source volume in o_1, $y_{o_1}^i$ denotes the ith cross-section of the target volume in o_1, and I denotes the total number of cross-sections within the volumes in o_1. G_{o_1} then learns to recover the cross-sections of the target volume from the corresponding cross-sections of the source volume.

$$\hat{y}_{p_1,o_1}^i = G_{o_1}(x_{o_1}^i) \tag{11}$$

where $\hat{y}_{p_1,o_1}^i$ denotes the ith cross-section of the target volume in o_1 recovered via the first progression. Meanwhile, D_{o_1} learns to distinguish between the real and fake cross-sections.

$$D_{o_1}(x_{o_1}^i, m) \in [0, 1] \tag{12}$$

where m is either generated ($\hat{y}_{p_1,o_1}^i$) or ground truth ($y_{o_1}^i$) target cross-section. To simultaneously train G_{o_1} and D_{o_1}, a loss function (L_{o_1}) consisting of adversarial and pixel-wise losses is used.

$$L_{o_1} = -\mathrm{E}_{x_{o_1}^i, y_{o_1}^i}[(D_{o_1}(x_{o_1}^i, y_{o_1}^i) - 1)^2] - \mathrm{E}_{x_{o_1}^i}[D_{o_1}(x_{o_1}^i, G_{o_1}(x_{o_1}^i))^2] + \mathrm{E}_{x_{o_1}^i, y_{o_1}^i}[||y_{o_1}^i - G_{o_1}(x_{o_1}^i)||_1] \tag{13}$$

Once G_{o_1} and D_{o_1} are properly trained, cross-sections in o_1 for the target volume are independently generated, and then combined with a concatenation block (c_{o_1}) to recover the entire target volume.

$$\hat{Y}_{p_1} = c_{o_1}(\hat{y}_{p_1,o_1}^1, \ldots, \hat{y}_{p_1,o_1}^I) \tag{14}$$

where $\hat{Y}_{p_1}$ denotes the target volume recovered after the first progression.

Second progression: Having learned the cross-sectional mapping in o_1, ProvoGAN then learns a separate recovery model in the second orientation o_2 to gradually enhance capture of fine-structural details and spatial correlations. The prediction for the target volume generated in the first progression is also incorporated as an input to the generator G_{o_2} to leverage global contextual priors.

$$\hat{y}_{p_2,o_2}^j = G_{o_2}(x_{o_2}^j, \hat{y}_{p_1,o_2}^j) \tag{15}$$

where $x_{o_2}^j$ denotes the jth cross-section of the source volume in o_2, $\hat{y}_{p_1,o_2}^j$ denotes the jth cross-section in o_2 of the target volume recovered in the first progression, and $\hat{y}_{p_2,o_2}^j$ denotes the jth cross-section in o_2 of the target volume recovered in the second progression. Meanwhile, discriminator D_{o_2} learns to distinguish between the generated and real cross-sections.

Third progression: Lastly, ProvoGAN learns a cross-sectional mapping in the third orientation o_3. As in the second progression, the prediction from the previous progression is incorporated into the mapping as prior information. Therefore, the third generator G_{o_3} receives as input the cross-sections in o_3 of the source volume and the previously recovered volume:

$$\hat{y}^k_{p_3,o_3} = G_{o_3}(x^k_{o_3}, \hat{y}^k_{p_2,o_3}) \tag{16}$$

where $x^k_{o_3}$ denotes the kth cross-section of the source volume in o_3, $\hat{y}^k_{p_2,o_3}$ denotes the kth cross-section in o_3 of the target volume recovered in the second progression, and $\hat{y}^k_{p_3,o_3}$ denotes the kth cross-section in o_3 of the target volume recovered in the third progression. Meanwhile, discriminator D_{o_3} learns to distinguish between the generated and real cross-sections. The final output volume $\hat{Y}_{p_3}$ of the proposed method is recovered by combining the generated cross-sections in o_3 via a concatenation block c_{o_3}:

$$\hat{Y}_{p_3} = c_{o_3}(\hat{y}^1_{p_3,o_3}, \ldots, \hat{y}^K_{p_3,o_3}) \tag{17}$$

where K denotes the total number of cross-sections in o_3. Note that, in MRI reconstruction, an additional constraint is introduced after each progression to enforce consistency of the acquired and recovered k-space data via the following procedure.

$$F_u(\hat{Y}_{p_n}) := F_u(X) \tag{18}$$

where F_u denotes the partial Fourier operator defined on the sampling mask utilized to acquire X, and n denotes the ongoing progression index. Meanwhile, an additional consistency between the progressions is enforced in the form of residual learning for MRI synthesis, where the generator models in the second and third progressions learn to predict the cross-sectional residuals between the target volume and the previously synthesized target volume.

$$\hat{y}_{p_n,o_n} = \hat{y}_{p_{n-1},o_n} + G_{o_n}(x_{o_n}, \hat{y}_{p_{n-1},o_n}) \tag{19}$$

2.5. Datasets

We demonstrated the proposed ProvoGAN approach on a public brain dataset, an in vivo knee dataset, and an in vivo brain dataset. The public dataset, IXI (https://brain-development.org/ixi-dataset/), consisted of coil-combined magnitude multi-contrast brain MR images of healthy subjects. The in vivo knee dataset (Epperson et al., 2013) consisted of multi-coil complex knee MR images of healthy subjects. The in vivo brain dataset contained multi-contrast brain MR images of both healthy subjects and glioma patients. Further details about each dataset are provided below.

IXI eataset: T_1-, T_2-, and proton-density (PD-) weighted brain MR images of 52 subjects were used, where 37 subjects were reserved for training, 5 for validation, and 10 for testing. T_1-weighted images were acquired sagittally with repetition time = 9.813 ms, echo time = 4.603 ms, flip angle = 8°, spatial resolution = $0.94 \times 0.94 \times 1.2$ mm^3, and matrix size = $256 \times 256 \times 150$. T_2-weighted images were acquired axially with repetition time = 8178 ms, echo time = 100 ms, flip angle = 90°, spatial resolution = $0.94 \times 0.94 \times 1.20$ mm^3, and matrix size = $256 \times 256 \times 150$. PD-weighted images were acquired axially with repetition time = 8178.34 ms, echo time = 8 ms, flip angle = 90°, spatial resolution = $0.94 \times 0.94 \times 1.2$ mm^3, and matrix size = $256 \times 256 \times 150$. Since the images of separate contrasts were spatially unregistered in this dataset, T_2- and PD-weighted images were registered onto T_1-weighted images using FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002) via an affine transformation. For synthesis images were further registered onto the Montreal Neurological Institute (MNI) template of T_1-weighted images with an isotropic resolution of 1 mm^3. Registration was performed based on mutual information loss.

In vivo knee dataset: PD-weighted multi-coil knee MR images of 20 subjects were used, where 12 subjects were reserved for training, 3 for validation, and 5 for testing. Images were sagittally acquired with 8 receive coils, repetition time = 1550 ms, echo time = 25.661 ms, spatial resolution = $0.5 \times 0.5 \times 0.6$ mm^3, and matrix size = $320 \times 320 \times 256$. MRI scans were performed in the Richard M. Lucas Center at Stanford University, California, United States on 3T GE scanners.

In vivo brain dataset: T_1-weighted, contrast enhanced T_1-weighted (T_{1c}), T_2-weighted, and FLAIR coil-combined brain MR images of 11 healthy subjects, 12 glioma patients with homogenous tumor, and 62 glioma patients with heterogenous tumor were used. 55 subjects were reserved for training (healthy: 8, homogenous: 7, heterogenous: 40), 15 for validation (healthy: 2, homogenous: 2, heterogenous: 11), and 15 for testing (healthy: 2, homogenous: 2, heterogenous: 11). Data augmentation was performed to prevent class imbalance among the three subject groups. Augmentation was achieved by rotating the volumes around their longitudinal axis by a random angle in the range $[-10°, 10°]$, and repeated 10 times for healthy subjects, 9 times for glioma patients with homogenous tumor, and performed once for glioma patients with heterogenous tumor. MRI exams were performed in the Department of Radiology at Hacettepe University, Ankara, Turkey, on Siemens and Philips scanners under a diverse set of protocols with varying spatial resolution across both contrast sets and subjects. Specifically, the prescribed resolutions included $1 \times 1 \times 1$ mm^3, $0.9 \times 0.9 \times 1.5$ mm^3, $0.9 \times 0.9 \times 1.8$ mm^3, $0.9 \times 0.9 \times 2.2$ mm^3 for T_1- and T_{1c}-weighted images, and $0.3 \times 0.3 \times 5$ mm^3, $0.4 \times 0.4 \times 5$ mm^3, $0.5 \times 0.5 \times 5$ mm^3, $0.6 \times 0.6 \times 5$ mm^3, $0.7 \times 0.7 \times 5$ mm^3 for T_2-weighted and FLAIR images. For demonstrations, all images were registered onto the MNI template of T_1-weighted images with an isotropic resolution of 1 mm^3. Registration was performed via FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002) using affine transformation based on mutual information loss. Imaging protocols were approved by the local ethics committee at Hacettepe University. All participants provided written informed consent.

For MRI reconstruction, volumes in the IXI and in vivo knee datasets were retrospectively undersampled with variable-density sampling patterns for acceleration factors ($R = 4, 8, 12, 16$). A sampling density function across k-space was taken a bi-variate normal distribution with mean at the center of k-space. The variance of the distribution was adjusted to achieve the expected sampling rate given R. The in-plane orientation was designated as axial. For MRI synthesis, all brain images were further skull stripped using FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002) with functional intensity threshold of 0.5, and vertical gradient intensity threshold of 0.

2.6. Competing methods

To demonstrate the performance of ProvoGAN in MR image recovery, we compared it against several state-of-the-art 3D models (vGAN, SC-GAN, REPLICA), 2D models (sGAN, RefineGAN, SPIRiT, SparseMRI), and hybrid models (M^{3}NET, TransferGAN). Baselines implemented for both reconstruction and synthesis included sGAN, vGAN, M^{3}NET, and TransferGAN. Meanwhile, task-specific baselines were RefineGAN, SPIRiT, and SparseMRI in MRI reconstruction, and SC-GAN and REPLICA in MRI synthesis.

The main effect that we seeked in comparing ProvoGAN against sGAN and vGAN was the benefit of progressive volumetrization over purely 2D or 3D processing. To improve reliability of these comparisons, we wanted to control for potential confounds from secondary factors such as network architecture or loss function. Therefore, the sGAN and vGAN models embodied consistent generator-discriminator architectures and loss functions with ProvoGAN (see Supp. Text 1,2 and Supp. Fig. 1,2 for details).

vGAN: A learning-based volumetric GAN model that performs a global one-shot mapping between source and target volumes (see Section 2.2). vGAN was implemented with a ResNet-based generator and a PatchGAN discriminator. *sGAN* A learning-based cross-sectional GAN model that performs a localized mapping between cross-sections of the source and target volumes (see Section 2.3). sGAN contained a ResNet-based generator and a PatchGAN discriminator.

RefineGAN: A learning-based cross-sectional GAN model proposed for MRI reconstruction (Quan et al., 2018). RefineGAN uses a cycle-consistency loss for acquired k-space samples in addition to adversarial and pixel-wise image loss to improve reconstruction quality. The overall architecture and loss terms were taken from Quan et al. (2018), but a ResNet-based generator was implemented to enable fair comparisons against ProvoGAN as it was observed here to yield higher reconstruction quality. *SC-GAN* A learning-based volumetric GAN model proposed for MRI synthesis (Lan et al., 2020). SC-GAN leverages self-attention modules to improve capture of long-range spatial dependencies. SC-GAN was implemented with a U-Net based generator and a PatchGAN discriminator as described in Lan et al. (2020), where the encoder and decoder components in the generator and the intermediate layer in the discriminator contained a self-attention module.

M^{3}NET: A learning-based hybrid model proposed for MRI segmentation (Wei et al., 2019). First, M^{3}NET separately learns orthogonal cross-sectional mappings in three rectilinear orientations (i.e., axial, coronal, sagittal). Using these 2D mappings as parallel streams, it fuses their outputs with a 3D fusion module to recover the target volume. The overall architecture, 3D fusion module, and loss functions were adopted from Wei et al. (2019), where 2D models were implemented with ResNet-based generators and PatchGAN discriminators as they were observed to yield enhanced performance in this study.

TransferGAN: A learning-based hybrid GAN model proposed for low-dose CT denoising (Shan et al., 2018). TransferGAN pretrains a 2D model for image recovery in a specific orientation, and then performs domain transfer from 2D onto 3D by transferring model weights. The transfer learning procedure was implemented as described in Shan et al. (2018), with 2D-3D models implemented as conditional GANs using ResNet-based generators and PatchGAN discriminators for fair comparison against ProvoGAN.

SparseMRI: A compressed sensing-based cross-sectional method for single-coil MRI reconstruction (Lustig et al., 2007). SparseMRI enforces transform domain sparsity as prior information during reconstruction from undersampled acquisitions. Here, SparseMRI was implemented as described in Lustig et al. (2007).

SPIRiT: A compressed sensing-based cross-sectional method for multi-coil MRI reconstruction (Lustig and Pauly, 2010). SPIRiT employs k-space interpolation kernels to estimate missing k-space samples. Here, SPIRiT was implemented as described in Lustig and Pauly (2010).

REPLICA: A compressed sensing-based volumetric method for multi-contrast MRI synthesis (Jog et al., 2017). REPLICA performs a nonlinear intensity transformation in multi-resolution feature space via a regression ensemble based on random forests. Here, REPLICA was implemented as described in Jog et al. (2017).

In single-coil reconstruction, learning-based models were trained to recover a magnitude image given real and imaginary parts of the undersampled image. In multi-coil reconstruction, learning-based models were first trained to recover a coil-combined magnitude image given real and imaginary parts of coil-combined Fourier reconstructions of undersampled acquisitions. A complex image was then formed by mapping the phase of the coil-combined undersampled image onto the predicted magnitude image. Coil combination was performed using sensitivity maps estimated via ESPIRiT (Uecker et al., 2014). A multi-coil complex image was obtained by projecting the coil-combined network prediction onto individual coils with the estimated sensitivity maps. Data-consistency was enforced in Fourier domain using the multi-coil complex images. In synthesis, learning-based models were trained to recover the magnitude image of the target contrast given magnitude images of the source contrasts.

The volumetric vGAN, SC-GAN, and REPLICA methods received as input volumetric source images. The cross-sectional sGAN-A, sGAN-C, sGAN-S, RefineGAN, SPIRiT, and SparseMRI methods received as input individual cross-sections of source volumes. M^{3}NET received cross-sectional inputs, aggregated them across the volume and finally processed the entire volume. TransferGAN received cross-sectional inputs during pretraining of the 2D model, and instead received volumetric inputs during training of the 3D model. Details regarding the dimensionality of input data to each method are provided in Supp. Text 3.

ProvoGAN, vGAN, and sGAN were implemented in Python 2.7 using PyTorch 0.4 and NumPy 1.14 libraries. Implementations of RefineGAN and SC-GAN were adopted from Quan et al. (2018) and Lan et al. (2020) respectively, and performed in Python 3.6 using PyTorch 1.10 and Numpy 1.19 libraries. Implementations of M^{3}NET and TransferGAN were adapted from Wei et al. (2019) and Shan et al. (2018) respectively, and performed in Python 2.7 using the PyTorch 0.4 and Numpy 1.14 libraries. SparseMRI and SPIRiT were implemented in MATLAB using the toolboxes available at https://people.eecs.berkeley.edu/~mlustig/Software.html. REPLICA was also implemented in MATLAB using the toolbox available at https://github.com/jcreinhold/replica. All implementations were run on workstations equipped with Intel(R) Core(TM) i7-7800X @ 3.50 GHz and i7-6850K @ 3.60 GHz CPUs, and nVidia GeForce GTX 1080 Ti and RTX 2080 Ti GPUs. Quantitative performance assessments based on PSNR and SSIM were performed in Python 2.7 using the Scikit-image 0.14 library. A toolbox to implement ProvoGAN and competing deep-learning models is publicly available at https://github.com/icon-lab/ProvoGAN.

2.7. Experiments

Task-specific progression order in ProvoGAN: Experiments were performed on ProvoGAN to optimize its progression order across the rectilinear orientations for specific tasks. To do this, multiple independent ProvoGAN models were trained while varying the progression order: 1) A → C → S, 2) A → S → C, 3) C → A → S, 4) C → S → A, 5) S → A → C, 6) S → C → A, where A denotes the axial, C denotes the coronal, and S denotes the sagittal orientation. Performance of these models were evaluated on the validation set via PSNR measurements. The experiments were performed separately for all synthesis and reconstruction tasks, and the progression orders optimized for specific tasks were used in all evaluations thereafter.

MRI reconstruction: Reconstruction experiments were performed on the IXI and in vivo knee datasets to compare ProvoGAN against sGAN, vGAN, RefineGAN, SparseMRI, and SPIRiT. In the IXI dataset, the proposed and competing methods were demonstrated separately for single-coil reconstruction of T_1- and T_2-weighted images with four distinct acceleration factors ($R = 4, 8, 12, 16$). Meanwhile, in the in vivo knee dataset, the proposed and competing methods were demonstrated for multi-coil reconstruction of PD-weighted images again with ($R = 4, 8, 12, 16$). Note that a single sGAN model was trained in the axial orientation (sGAN-A) given the axial readout direction.

MRI synthesis: Synthesis experiments were performed on the IXI and in vivo brain datasets to demonstrate ProvoGAN against sGAN, vGAN, SC-GAN, and REPLICA. All synthesis experiments were conducted on coil-combined magnitude images. In the IXI dataset, three synthesis tasks were consid-

ered: 1) T_2, PD $\rightarrow$ T_1, 2) T_1, PD $\rightarrow$ T_2, 3) T_1, T_2 $\rightarrow$ PD. In the in vivo brain dataset, four synthesis tasks were considered: 1) T_2, FLAIR, T_{1c} $\rightarrow$ T_1, 2) T_1, FLAIR, T_{1c} $\rightarrow$ T_2, 3) T_1, T_2, T_{1c} $\rightarrow$ FLAIR, 4) T_1, T_2, FLAIR $\rightarrow$ T_{1c}. For each task, three independent sGAN models were implemented to recover target cross-sections in separate orientations: sGAN-A for the axial, sGAN-C for the coronal, sGAN-S for the sagittal orientation.

Progressive volumetrization versus hybrid models: Experiments were conducted on the IXI dataset to demonstrate ProvoGAN against M^{3}NET and TransferGAN. Reconstruction experiments were conducted for T_1- and T_2-weighted image recovery tasks at four distinct acceleration factors ($R = 4, 8, 12, 16$). Meanwhile, synthesis experiments were conducted for the many-to-one recovery tasks of T_2, PD $\rightarrow$ T_1, T_1, PD $\rightarrow$ T_2, and T_1, T_2 $\rightarrow$ PD.

Radiological evaluation: To assess the clinical value of the recovered images, an expert radiologist (25+ years of experience) gave opinion scores to the images while blinded to the method name and order of presentation. Reconstructed images were evaluated for single-coil reconstructions of T_1- and T_2-weighted acquisitions at $R = 8$ in the IXI dataset, and multi-coil reconstructions of PD-weighted acquisitions at $R = 8$ in the in vivo knee dataset. Synthesized images were evaluated for T_2, PD $\rightarrow$ T_1 in IXI and T_1, T_2, T_{1c} $\rightarrow$ FLAIR in the in vivo brain datasets. From each recovered volume, intermediate axial, coronal, and sagittal cross-sections were randomly selected, and the image quality was rated as the similarity to ground truth images on a five-point scale (5: perfect match, 4: good, 3: moderate, 2: limited, 1: very poor, 0: unacceptable).

Multi-cross-section models: To demonstrate the benefit of leveraging contextual priors by incorporating multiple neighboring cross-sections at the input level, variants of ProvoGAN and sGAN, referred to as ProvoGAN(multi) and sGAN(multi), were implemented, which receive as input n_c consecutive cross-sections to recover the corresponding central cross-section in the target volume. Here $n_c = 3$ was selected as higher number of cross-sections did not yield a notable benefit in recovery performance (Dar et al., 2019). Experiments were performed on the IXI dataset for reconstruction of T_1- and T_2-weighted images with distinct acceleration factors ($R = 4, 8, 12, 16$), and for many-to-one synthesis tasks (T_2, PD $\rightarrow$ T_1, T_1, PD $\rightarrow$ T_2, T_1, T_2 $\rightarrow$ PD). The ordering of the progressions across the orientations in ProvoGAN was optimized via PSNR measurements in the validation set. Three separate sGAN(multi) models were implemented in each individual rectilinear orientation: sGAN(multi)-A for the axial, sGAN(multi)-C for the coronal, and sGAN(multi)-S for the sagittal orientation.

Cross-sectional models of varying complexity: An additional analysis was performed on ProvoGAN and sGAN to examine recovery performance as a function of the complexity of convolutional layers. Several variants of ProvoGAN and sGAN were implemented while the number of network weights in individual convolutional layers were scaled by $n_f \in \{1/16, 1/9, 1/4, 1, 4, 9, 16\}$, where the kernel size, number of layers, number of hidden units were kept fixed but the number of filters were modified. This resulted in seven distinct ProvoGAN and sGAN pairs: ProvoGAN(n_f)-sGAN(n_f). Experiments were performed on the IXI dataset for single-coil reconstruction of T_1-weighted acquisitions undersampled at $R = 8$ and a many-to-one synthesis task of T_2, PD $\rightarrow$ T_1. The ordering of the progressions across the orientations in ProvoGAN(n_f) was optimized via PSNR measurements in the validation set. Separate sGAN models were trained in the individual rectilinear orientations (axial, coronal, and sagittal) for each model complexity level: sGAN(n_f)-A, sGAN(n_f)-C, sGAN(n_f)-S.

Data efficiency: The complexity of volumetric models can elevate the amount of data samples required for successful training. Instead, ProvoGAN comprises sequential cross-sectional models that are of lower complexity and that can be trained effectively with fewer data. Experiments were conducted to comparatively demonstrate the data efficiency of ProvoGAN against vGAN. To do this, models were trained using data from varying number of subjects ($n_T \in \{5, 15, 25\}$), yielding ProvoGAN(n_T) and vGAN(n_T). T_1- and T_2-weighted reconstructions in the IXI dataset at $R = 4, 8, 12, 16$ were considered. T_2, PD $\rightarrow$ T_1, T_1, PD $\rightarrow$ T_2, and T_1, T_2 $\rightarrow$ PD synthesis tasks were considered. To prevent potential confounds, the optimal progression orders determined for the original ProvoGAN models were retained.

Generalizability of progressive volumetrization: Experiments were conducted to demonstrate the generalizability of the proposed progressive volumetrization to another network architecture. Demonstrations were performed on the IXI dataset for T_1, T_2 $\rightarrow$ PD, T_1, PD $\rightarrow$ T_2, and T_2, PD $\rightarrow$ T_1 synthesis tasks. A recent state-of-the-architecture, SC-GAN, with a U-Net backbone using intermittent self-attention layers (Lan et al., 2020) was considered. Variants of sGAN, vGAN, and ProvoGAN were implemented based on this architecture: sSC-GAN, vSC-GAN, ProvoSC-GAN. Again, three separate sSC-GAN models were trained in each orientation: sSC-GAN-A in the axial, sSC-GAN-C in the coronal, and sSC-GAN-S in the sagittal orientation.

Statistical assessments: PSNR, SSIM, and opinion scores were utilized to quantitatively evaluate the recovery quality of the methods under comparison. Since the performance measurements from these metrics followed a non-normal distribution ($p < 0.05$ with Shapiro–Willks test), significance of differences in quantitative metrics were evaluated using non-parametric statistical tests. Assessments of progression order in ProvoGAN were performed via Kruskal–Wallis tests, whereas performance comparisons among competing methods were performed via Wilcoxon signed-rank tests.

3. Results

3.1. Task-specific progression order

ProvoGAN serially decomposes a given volumetric recovery task into cross-sectional mappings in three rectilinear orientations. Subsequent 2D mappings are residually learned based on outputs from earlier progressions. Please note that spatial distribution of the tissues and the correlations between the source-target images may vary uniquely across orientations for each recovery task. In this setup, if an earlier 2D model yields relatively higher artifacts, the task difficulty for the remaining progressions would be elevated. Contrarily, initiating the progression at a different orientation with lower artifacts can reduce task difficulty for the remaining stages. Therefore, we predicted that the progression sequence in ProvoGAN can significantly affect task-specific recovery performance.

To test this prediction, we performed reconstruction and synthesis experiments separately on the IXI, in vivo brain, and in vivo knee datasets (see Section 2.5 for details). We comparatively evaluated performance of multiple independent ProvoGAN models for the six possible permutations of the progression sequence: 1) A $\rightarrow$ C $\rightarrow$ S, 2) A $\rightarrow$ S $\rightarrow$ C, 3) C $\rightarrow$ A $\rightarrow$ S, 4) C $\rightarrow$ S $\rightarrow$ A, 5) S $\rightarrow$ A $\rightarrow$ C, 6) S $\rightarrow$ C $\rightarrow$ A, where A denotes the axial, C denotes the coronal, and S denotes the sagittal orientation. Here, we considered volumetric PSNR measurements between the recovered and reference target volumes within the validation set. The highest and lowest performing ProvoGAN models yield an average PSNR difference of 3.44 dB for single-coil reconstruction tasks in IXI and 3.42 dB for multi-coil reconstruction tasks in the in vivo knee dataset (see Supp. Tables 1,2 for details). Meanwhile, the average PSNR difference between the highest and lowest performing ProvoGAN models is 1.46 dB for synthesis in IXI, and 1.01 dB for synthesis in the in vivo brain dataset (see Supp. Tables 3,4). Optimization of the progression order enables a significant per-

Table 1
Quality of Reconstruction in the IXI Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the reconstructed and ground truth images in the test set in the IXI dataset are given as mean ± std for the test set. The measurements are reported for zero-filled images (ZF), the proposed ProvoGAN and competing sGAN, vGAN, RefineGAN, and SparseMRI reconstruction methods for four distinct acceleration factors ($R = 4, 8, 12, 16$). Boldface indicates the best performing method.

		ProvoGAN		sGAN		vGAN		RefineGAN		SparseMRI		ZF	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
$R = 4$	T_1	**35.25**	**96.73**	33.85	93.21	30.82	88.61	31.04	92.13	26.25	72.16	24.59	64.96
		±1.78	**±0.57**	± 1.29	± 0.78	± 1.24	± 0.93	± 1.74	± 0.94	± 0.80	± 1.92	± 1.17	± 3.07
	T_2	**35.50**	**96.08**	32.95	86.44	33.09	93.13	33.96	94.00	28.08	82.44	27.54	75.14
		±2.62	**±1.07**	± 1.50	± 1.26	± 1.73	± 1.20	± 0.91	± 0.57	± 0.73	± 1.42	± 1.04	± 2.09
$R = 8$	T_1	**31.38**	**94.93**	30.08	91.18	29.71	88.73	27.63	91.81	25.92	72.26	23.04	62.53
		±1.26	**±0.86**	± 1.32	± 1.01	± 0.88	± 1.23	± 1.56	± 0.68	± 0.36	± 1.52	± 1.05	± 3.32
	T_2	**33.49**	**95.92**	32.24	90.47	31.35	92.57	30.70	93.82	26.55	79.35	26.67	74.16
		±2.21	**±1.01**	± 2.14	± 0.95	± 0.66	± 0.86	± 1.51	± 0.40	± 0.51	± 1.39	± 0.87	± 2.10
$R = 12$	T_1	**29.67**	**92.48**	27.34	86.23	27.56	82.70	27.48	88.56	23.84	60.77	20.76	52.72
		±0.91	**±0.90**	± 1.06	± 1.36	± 1.20	± 1.70	± 1.08	± 1.07	± 0.28	± 1.74	± 1.04	± 3.62
	T_2	**30.41**	**91.98**	28.48	79.50	27.97	85.76	29.16	91.49	24.67	70.16	24.60	66.90
		± 1.03	**±1.40**	± 1.06	± 2.15	± 0.79	± 1.74	± 1.54	± 0.87	± 0.59	± 1.67	± 0.70	± 2.12
$R = 16$	T_1	**29.15**	**91.40**	26.73	85.23	25.47	79.56	22.37	84.31	23.42	61.64	20.95	52.85
		±1.09	**±1.09**	± 1.53	± 1.74	± 1.07	± 2.16	± 1.07	± 1.36	± 0.47	± 1.86	± 1.07	± 3.46
	T_2	**30.66**	**93.74**	29.05	83.38	27.66	85.47	28.12	91.07	24.37	69.15	24.44	66.70
		±1.60	**±1.35**	± 1.04	± 1.22	± 0.50	± 1.57	± 1.26	± 0.84	± 0.69	± 1.60	± 0.76	± 2.19

Table 2
Quality of Reconstruction in the In vivo Knee Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the reconstructed and ground truth images in the test set in the in vivo knee dataset are given as mean ± std for the test set. The measurements are reported for zero-filled images (ZF), the proposed ProvoGAN and competing sGAN, vGAN, RefineGAN, and SPIRiT methods for four distinct acceleration factors ($R = 4, 8, 12, 16$). Boldface indicates the best performing method.

	ProvoGAN		sGAN		vGAN		RefineGAN		SPIRiT		ZF	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
$R = 4$	**40.75**	**95.74**	40.34	95.69	36.80	92.79	40.31	95.21	39.46	95.35	32.17	93.50
	±1.35	**±0.94**	± 1.43	± 0.87	± 1.69	± 1.36	± 1.50	± 1.15	± 1.39	± 1.12	± 2.38	± 1.85
$R = 8$	**39.45**	**95.13**	38.73	93.73	30.83	87.54	38.22	92.43	35.61	93.16	29.85	90.81
	±2.15	**±1.08**	± 1.01	± 0.98	± 1.44	± 2.80	± 1.43	± 1.58	± 2.70	± 1.60	± 1.77	± 1.88
$R = 12$	**36.99**	**93.63**	36.76	91.99	29.03	88.10	36.21	90.82	33.58	91.66	28.65	89.17
	±1.29	**±0.95**	± 1.18	± 0.97	± 1.49	± 1.78	± 0.88	± 1.47	± 3.17	± 1.86	± 1.25	± 1.64
$R = 16$	**37.86**	**92.27**	35.11	89.34	28.99	86.00	35.42	89.70	32.28	90.69	27.93	87.94
	±0.50	**±1.45**	± 1.06	± 1.61	± 2.21	± 2.28	± 0.55	± 1.13	± 2.63	± 1.85	± 1.25	± 1.60

formance increase in both reconstruction ($p < 0.05$, Kruskal–Wallis test) and synthesis ($p < 0.05$). Therefore, the optimal orders were utilized for each recovery task in all evaluations thereafter unless otherwise stated.

Note that brain and knee MRI acquisitions were undersampled across the two phase-encoding dimensions in the axial plane (A/P, L/R) in the reconstruction experiments, so the reconstruction task in the axial plane is relatively more difficult. Accordingly, there is a general performance increase in progression orders that leave the axial orientation towards later stages of ProvoGAN, and this effect is particularly emphasized towards higher acceleration rates R (see Supp. Tables 1,2). For synthesis, a factor that contributes to task difficulty is the level of structural details in the target contrast. Images in the IXI dataset and T_2-weighted and FLAIR images in the in vivo brain dataset have relatively higher spatial resolution in the axial plane, but broader voxel dimensions in the longitudinal direction. Accordingly, a general performance increase is observed in progression orders that leave the axial orientation towards later stages, and these effects are more accentuated when the target is T_1- and T_2-weighted images that have relatively better capture of structural details compared to other contrasts such as PD- or T_{1c}-weighted (see Supp. Tables 3,4).

3.2. Accelerated MRI reconstruction

Next, we performed comprehensive experiments on the IXI and in vivo knee datasets for accelerated MRI reconstruction. We comparatively demonstrated the recovery quality of ProvoGAN against state-of-the-art cross-sectional (sGAN, RefineGAN, SparseMRI, and SPIRiT), and volumetric (vGAN) models (see Section 2.6 for details). We first assessed the performance of the competing methods quantitatively based on volumetric PSNR and SSIM measurements between the reconstructed and high-quality reference images in the test set. We considered single-coil reconstruction tasks in the IXI dataset for T_1- and T_2-weighted images with distinct acceleration factors ($R = 4, 8, 12, 16$). The proposed ProvoGAN model offers enhanced recovery performance compared to competing methods ($p < 0.05$), where it achieves in the range of [1.85,6.55] dB higher PSNR and [3.26,23.17] % higher SSIM (see Table 1). We then considered multi-coil reconstruction of PD-weighted images in the in vivo brain dataset with $R = 4, 8, 12, 16$. ProvoGAN again maintains superior performance to the competing methods ($p < 0.05$), where it achieves in the range of [1.03,7.35] dB higher PSNR and [1.48,5.59] % higher SSIM (see Table 2).

To corroborate quantitative assessments, we visually examined the reconstructed volumes from individual methods to identify the nature of reconstruction errors ProvoGAN alleviates. Representative results from the competing methods are shown in Fig. 2 for IXI and in Fig. 3 for the in vivo knee dataset. Overall, cross-sectional models (sGAN, RefineGAN, SparseMRI, SPIRiT) that perform 2D mapping via compressed sensing or deep learning suffer from discontinuity artifacts across individually recovered cross-sections and retrograded capture of fine-structural details. Meanwhile, the volumetric vGAN model performing 3D mapping suffers from loss of spatial resolution within the reconstructed volumes due to noticeable over-smoothing. In contrast, ProvoGAN reconstructs the target vol-

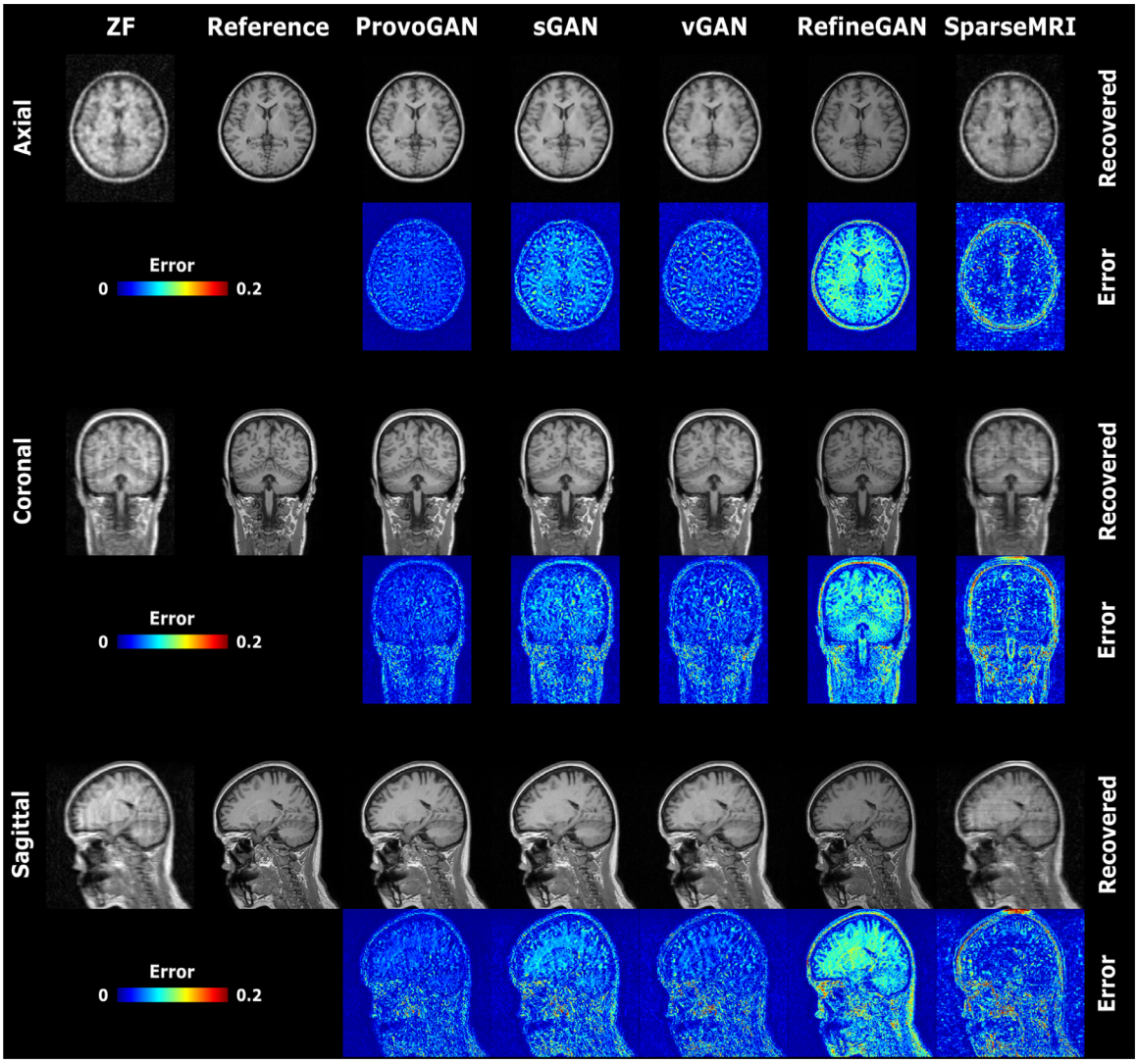

Fig. 2. The proposed ProvoGAN method is demonstrated on the IXI dataset for single-coil reconstruction of T_1-weighted acquisitions undersampled at $R = 8$. Representative results are displayed for all competing methods together with the zero-filled (ZF) undersampled source images (first column) and the reference target images (second column). The top two rows display results for the axial, the middle two rows for the coronal, and the last two rows for the sagittal orientation. Error was taken as the absolute difference between the reconstructed and reference images (see colorbar). Overall, the proposed ProvoGAN method offers delineation of tissues with higher acuity compared to the volumetric (vGAN) model, and alleviates undesirable discontinuities compared to cross-sectional models (sGAN, RefineGAN, SparseMRI) by improving reconstruction performance in all orientations.

umes with higher consistency across the cross-sections in all orientations and offers sharper delineation of brain and knee tissues. Taken together, these findings clearly outline ProvoGAN's potential to mitigate the limitations of volumetric and cross-sectional models for accelerated MRI reconstruction.

3.3. Multi-contrast MRI synthesis

We further conducted experiments on the IXI and in vivo brain datasets for multi-contrast MRI synthesis to demonstrate ProvoGAN against state-of-the-art cross-sectional (sGAN) and volumetric (vGAN, SC-GAN, REPLICA) models (see Section 2.6 for details). We again measured volumetric PSNR and SSIM between the synthesized and reference target images for quantitative performance evaluation. In the IXI dataset, we considered synthesis tasks of T_2, PD $\rightarrow$ T_1, T_1, PD $\rightarrow$ T_2, and T_1, T_2 $\rightarrow$ PD. ProvoGAN outperforms the competing methods in all tasks ($p < 0.05$), where it achieves in the range of [1.20,2.90] dB higher PSNR and [2.08,4.37] % higher SSIM (see Table 3). In the in vivo brain dataset, we considered synthesis tasks of T_2, FLAIR, T_{1c} $\rightarrow$ T_1, T_1, FLAIR, T_{1c} $\rightarrow$ T_2, T_1, T_2, T_{1c} $\rightarrow$ FLAIR, and T_1, T_2, FLAIR $\rightarrow$ T_{1c}. ProvoGAN again yields enhanced recovery performance in all tasks compared to the competing methods ($p < 0.05$), where it maintains [0.59,5.01] dB higher PSNR and [1.96,5.36] % higher SSIM (see Table 4). Note that the in vivo brain dataset was acquired under a diverse set of scanning protocols with varying spatial resolution where T_2-weighted and FLAIR images were acquired with larger slice thickness (see Section 2.5). Here, models were built to synthesize T_1-weighted and T_{1c}-weighted images, where the thick-slices data were on the input side. Models were also built to synthesize T_2-weighted and FLAIR images where thick-slice data were on the output side (see Section 2.7). In both cases, ProvoGAN enhances recovery performance compared to sGAN. On average, ProvoGAN achieves 1.69 dB higher PSNR and 4.18% higher SSIM when thick-slice data are on the input side, and 0.70 dB higher PSNR and 1.85% higher SSIM when thick-slice data are on the output side.

The superior synthesis quality offered by ProvoGAN is clearly visible in representative results displayed in Fig. 4 for the IXI dataset and Fig. 5 for the in vivo brain dataset. These results indicate that the cross-sectional sGAN-A, sGAN-C, and sGAN-S mod-

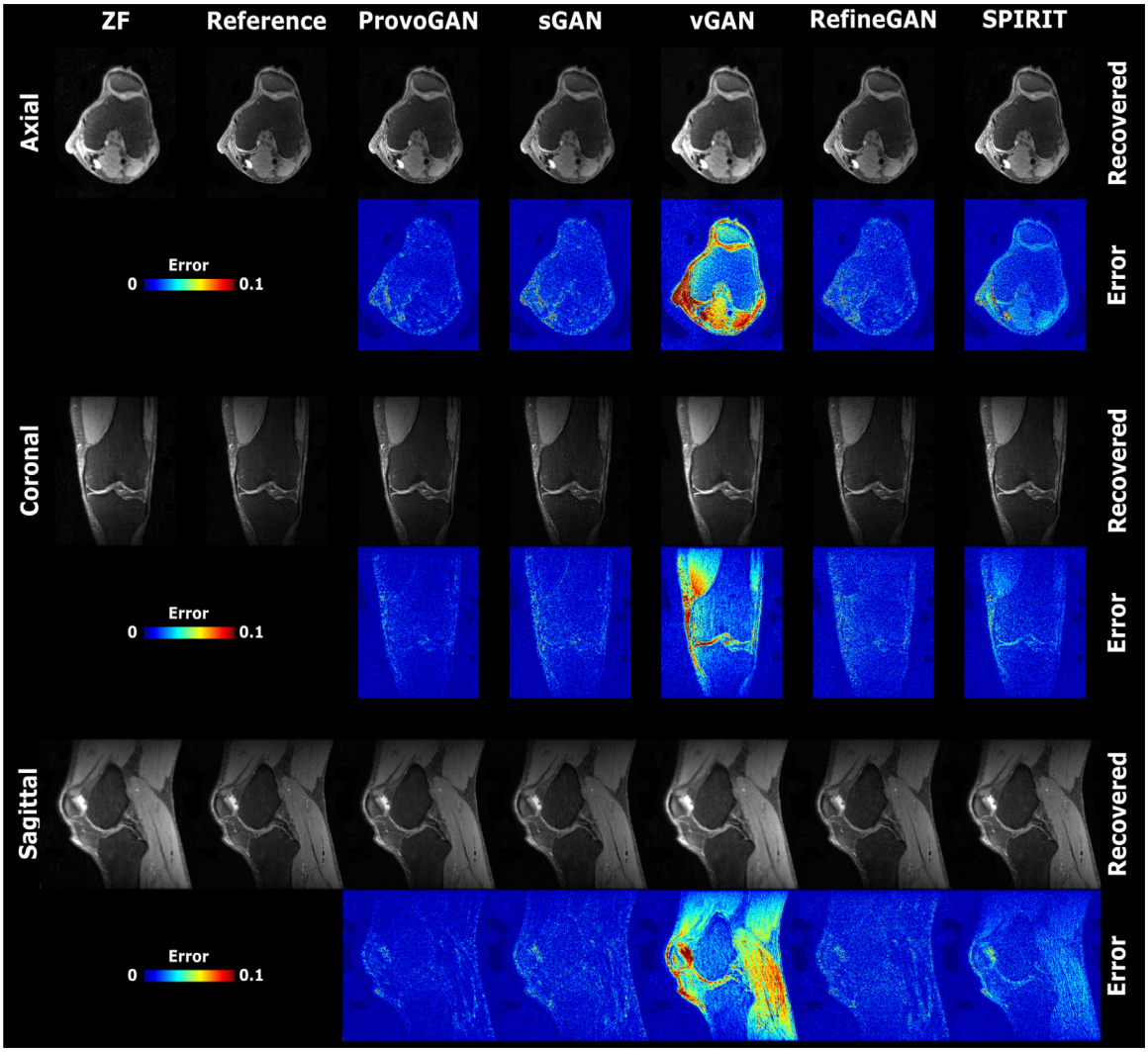

Fig. 3. The proposed method is demonstrated on the in vivo multi-coil knee dataset for reconstruction at an acceleration ratio of $R = 8$. Representative results are displayed for all competing methods together with the zero-filled (ZF) undersampled source images (first column) and the reference target images (second column). The top two rows display results for the axial, the middle two rows for the coronal, and the last two rows for the sagittal orientation. Error was taken as the absolute difference between the reconstructed and reference images (see colorbar). Overall, ProvoGAN achieves sharper tissue depiction compared to vGAN, and alleviates undesirable discontinuities compared to cross-sectional models (sGAN, RefineGAN, SPIRiT) by improving reconstruction performance in all orientations.

Table 3
Quality of Synthesis in the IXI Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the synthesized and ground truth images in the test set in the IXI dataset are given as mean ± std. The measurements are provided for the proposed and competing methods for all synthesis tasks: 1) T_2, PD → T_1, 2) T_1, PD → T_2, 3) T_1, T_2 → PD. sGAN-A denotes the sGAN model trained in the axial orientation, sGAN-C in the coronal orientation, and sGAN-S in the sagittal orientation. Boldface indicates the highest performing method.

	ProvoGAN		sGAN-A		sGAN-C		sGAN-S		vGAN		SC-GAN		REPLICA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T_2, PD → T_1	**24.15**	**90.33**	23.20	85.81	22.58	86.60	23.65	87.71	23.35	85.48	22.58	85.32	21.14	86.30
	±2.80	**±4.47**	± 2.08	± 3.95	± 2.11	±4.05	±1.98	±4.15	2.89	4.18	2.99	4.00	±4.15	± 3.80
T_1, PD → T_2	**28.97**	**94.17**	27.64	92.49	27.74	92.67	27.93	93.28	25.97	90.61	25.29	89.81	26.98	92.51
	±2.91	**±4.16**	± 2.59	± 4.20	± 2.67	± 4.31	± 2.19	± 2.88	± 1.81	± 4.04	± 1.95	± 4.31	± 2.37	± 4.57
T_1, T_2 → PD	**29.81**	**95.41**	27.69	93.64	29.00	94.21	27.12	92.67	26.17	90.70	26.36	92.12	26.96	94.10
	±2.96	**±2.75**	2.20	± 3.00	± 2.41	± 2.99	± 1.61	± 2.95	± 1.41	± 2.47	± 1.41	± 2.59	± 2.93	± 3.03

els suffer from suboptimal recovery in the longitudinal dimension due to independent synthesis of cross-sections. Meanwhile volumetric vGAN, SC-GAN, and REPLICA models suffer from poor recovery of fine-structural details and loss of spatial resolution in the target images due to increased model complexity. In comparison to cross-sectional baselines, ProvoGAN alleviates discontinuity artifacts by pooling global contextual information via progressive execution of cross-sectional mappings. In comparison to volumetric baselines, ProvoGAN offers sharper and improved tissue depiction particularly near tumor regions due to its improved learning behavior. Overall, these findings demonstrate ProvoGAN's utility for diverse synthesis tasks in multi-contrast MRI exams.

3.4. Demonstrations against hybrid models

Having demonstrated the superior performance of ProvoGAN against several state-of-the-art cross-sectional and volumetric models, we conducted additional experiments to comparatively

Table 4
Quality of Synthesis in the In vivo Brain Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the synthesized and ground truth images in the test set of the in vivo brain dataset are given as mean ± std. The measurements are provided for proposed and competing methods for all many-to-one synthesis tasks: 1) T_2, FLAIR, $T_{1c} \rightarrow T_1$, 2) T_1, FLAIR, $T_{1c} \rightarrow T_2$, 3) T_1, T_2, $T_{1c} \rightarrow$ FLAIR, 4) T_1, T_2, FLAIR $\rightarrow T_{1c}$. sGAN-A denotes the sGAN model trained in the axial orientation, sGAN-C in the coronal orientation, and sGAN-S in the sagittal orientation. Boldface indicates the highest performing method.

	ProvoGAN		sGAN-A		sGAN-C		sGAN-S		vGAN		SC-GAN		REPLICA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T_2, FLAIR, $T_{1c} \rightarrow T_1$	**26.92** ±**4.55**	**94.24** ±**3.41**	24.17 ± 3.83	88.23 ± 4.68	25.31 ± 4.12	90.78 ± 4.09	26.22 ± 3.09	91.17 ± 3.03	22.73 ± 3.69	87.73 ± 3.60	21.70 ± 2.96	85.60 ±3.45	17.14 ± 4.43	83.34 ± 7.43
T_1, FLAIR, $T_{1c} \rightarrow T_2$	**26.87** ±**2.40**	**92.79** ±**4.22**	25.67 ± 1.75	89.76 ± 3.20	25.98 ± 2.18	90.95 ± 4.10	26.85 ± 2.38	92.11 ± 4.11	25.48 ± 1.82	89.75 ± 3.93	24.48 ± 1.72	88.63 ± 3.54	24.68 ± 1.94	89.06 ± 3.16
T_1, T_2, $T_{1c} \rightarrow$ FLAIR	**25.52** ±**2.22**	**90.39** ±**3.06**	24.50 ± 1.84	87.21 ± 2.73	24.95 ± 2.03	88.09 ± 3.18	24.81 ± 2.21	88.39 ± 2.67	22.94 ± 1.61	85.67 ± 3.23	23.07 ± 2.23	85.88 ±2.62	22.70 ±2.79	87.63 ± 2.97
T_1, T_2, FLAIR $\rightarrow T_{1c}$	**29.67** ±**2.23**	**94.14** ±**2.09**	28.53 ± 1.98	91.70 ± 2.34	28.48 ± 2.11	89.91 ± 3.09	28.75 ± 2.23	92.06 ± 2.44	27.21 ± 1.46	89.11 ± 1.74	27.57 ± 1.68	90.19 ± 2.00	24.44 ± 2.49	90.09 ± 3.03

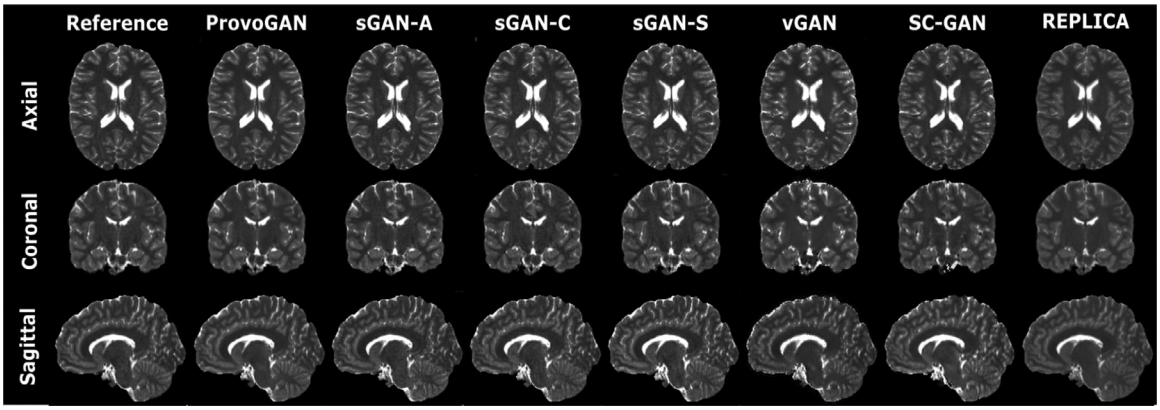

Fig. 4. The proposed method is demonstrated on the IXI dataset for T_2-weighted image synthesis from T_1- and PD-weighted images. Representative results are displayed for all competing methods together with the reference target images (first column). The first row displays results for the axial orientation, the second row for the coronal orientation, and the third row for the sagittal orientation. Overall, the proposed method delineates tissues with higher spatial resolution compared to volumetric vGAN, SC-GAN, and REPLICA models, and alleviates discontinuity artifacts by improving synthesis performance in all orientations compared to cross-sectional sGAN-A, sGAN-C, and sGAN-S models.

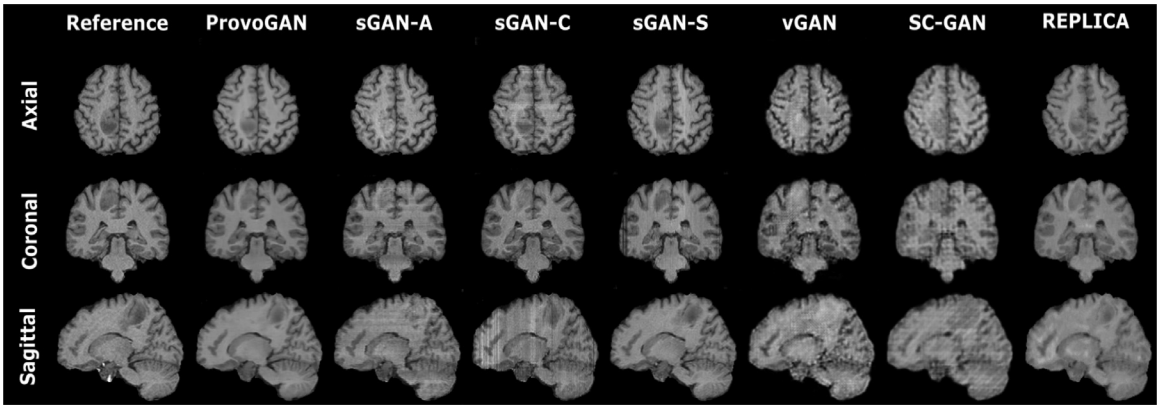

Fig. 5. The proposed method is demonstrated on the in vivo brain dataset for T_1-weighted image synthesis from T_2-, T_{1c}-weighted and FLAIR images. Representative results are displayed for all competing methods together with the reference target images (first column). The first row displays results for the axial, the second row for the coronal, and the third row for the sagittal orientation. Overall, the proposed method delineates tissues with higher spatial resolution compared to volumetric vGAN, SC-GAN, and REPLICA models, and alleviates discontinuity artifacts by improving synthesis performance in all orientations compared to cross-sectional sGAN-A, sGAN-C, and sGAN-S models. Meanwhile, the proposed method achieves more accurate depictions for tumor regions, which are suboptimally recovered by the competing methods.

evaluate it against alternative volumetrization methods. In particular, ProvoGAN was compared with hybrid models based on fusion (M^3NET) and transfer learning strategies (TransferGAN) that both involve a mixture of cross-sectional and volumetric mappings (see Section 2.6 for details). Experiments were performed on the IXI dataset for accelerated MRI reconstruction and multi-contrast MRI synthesis. For reconstruction, T_1- and T_2-weighted image recovery tasks at four distinct acceleration factors ($R = 4, 8, 12, 16$) were examined. Table 5 reports performance measurements for the methods under comparison. ProvoGAN yields superior performance compared to both hybrid models in all reconstruction tasks ($p < 0.05$), where on average it achieves 1.87 dB higher PSNR and 4.34% higher SSIM compared to M^3NET, and 1.83 dB higher PSNR and 5.02% higher SSIM compared to TransferGAN. Meanwhile, T_1, $T_2 \rightarrow$ PD, T_1, PD $\rightarrow T_2$, T_2, PD $\rightarrow T_1$ recovery tasks were considered for synthesis. The respective measurements

Table 5
Comparison of Volumetrization Approaches for Reconstruction in the IXI Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the reconstructed and ground truth images in the test set in the IXI dataset are given as mean ± std. The measurements are reported for the proposed ProvoGAN and competing M^{3}NET and TransferGAN methods for four distinct acceleration factors ($R = 4, 8, 12, 16$). Boldface indicates the best performing method.

		ProvoGAN		M^{3}NET		TransferGAN	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
$R = 4$	T_1	**35.25**	**96.73**	31.25	92.07	31.80	89.11
		±1.78	**±0.57**	± 1.87	± 1.18	± 0.88	± 1.10
	T_2	**35.50**	**96.08**	33.85	92.88	34.18	94.12
		±2.62	**±1.07**	± 2.73	± 1.12	± 1.30	± 0.84
$R = 8$	T_1	**31.38**	**94.93**	29.61	90.78	30.37	88.44
		±1.26	**±0.86**	± 1.03	± 1.05	± 1.22	± 0.99
	T_2	**33.49**	**95.92**	32.12	93.09	31.69	93.49
		±2.21	**±1.01**	± 2.25	± 1.32	± 0.83	± 0.71
$R = 12$	T_1	**29.67**	**92.48**	28.55	87.61	28.79	85.31
		±0.91	**±0.90**	± 1.05	± 1.46	± 0.89	± 1.54
	T_2	**30.41**	**91.98**	30.29	88.31	29.18	90.49
		± 1.03	**±1.40**	± 1.00	± 1.08	± 0.81	± 0.94
$R = 16$	T_1	**29.15**	**91.40**	25.83	85.59	27.61	83.31
		±1.09	**±1.09**	± 1.13	± 1.90	± 0.88	± 1.65
	T_2	**30.66**	**93.74**	29.02	88.03	27.28	88.84
		±1.60	**±1.35**	± 1.24	± 1.71	± 1.69	± 1.36

Table 6
Comparison of Volumetrization Approaches for Synthesis in the IXI Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the synthesized and ground truth images in the test set in the IXI dataset are given as mean ± std. The measurements are reported for the proposed ProvoGAN and competing M^{3}NET and TransferGAN methods for all synthesis tasks: 1) T_2, PD → T_1, 2) T_1, PD → T_2, 3) T_1, T_2 → PD. Boldface indicates the highest performing method.

	ProvoGAN		M^{3}NET		TransferGAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T_2, PD → T_1	**24.15**	**90.33**	20.85	86.81	23.84	87.09
	±2.80	**±4.47**	± 4.08	± 5.00	± 3.37	± 4.09
T_1, PD → T_2	**28.97**	**94.17**	23.79	89.82	27.78	93.10
	±2.91	**±4.16**	± 1.23	± 3.12	± 2.47	± 3.98
T_1, T_2 → PD	**29.81**	**95.41**	29.48	94.84	28.71	94.45
	±2.96	**±2.75**	± 2.24	± 2.41	± 1.89	± 2.52

are reported in Table 6. We find that ProvoGAN again maintains enhanced recovery performance in all synthesis tasks ($p < 0.05$), where it achieves an average of 2.94 dB higher PSNR and 2.81% higher SSIM compared to M^{3}NET, and 0.87 dB higher PSNR and 1.76% higher SSIM compared to TransferGAN.

Quantitative improvements that ProvoGAN offers are also visible in representative images displayed in Supp. Fig. 3 for reconstruction and in Fig. 6 for synthesis. The M^{3}NET model that performs 3D fusion of 2D model outputs at separate orientations moderately increases contextual sensitivity, but suffers from residual discontinuity artifacts and over-smoothing. Meanwhile, the TransferGAN model that transfers pretrained weights from a 2D model to condition the final 3D model improves learning behavior, but it suffers from elevated model complexity leading to loss of spatial resolution and structural details. In contrast, ProvoGAN yields enhanced recovery performance in all orientations with higher contextual consistency and sensitivity to structural details in the recovered images.

3.5. Radiological evaluation

Quantitative performance assessments in MRI recovery tasks clearly indicate that ProvoGAN outperforms competing volumetric and cross-sectional models in terms of image quality metrics (i.e., PSNR, SSIM). Yet, an important question concerns to what extent these quantitative improvements will benefit diagnostic assessments. Given its ability to effectively capture global context as well as fine structural details, we hypothesized that ProvoGAN will recover MR images of equivalent or higher diagnostic value than competing models. To test this hypothesis, radiological evaluations were performed on images recovered via ProvoGAN, sGAN and vGAN, as well as SC-GAN, REPLICA for synthesis, and RefineGAN, compressed-sensing methods (SparseMRI and SPIRIT) for reconstruction (see Section 2.7 for details). Opinion scores for all methods in axial, coronal, and sagittal orientations denoted as (OS_A, OS_C, OS_S) are reported in Fig. 7a–c for reconstruction and in Fig. 7d,e for synthesis. Across reconstruction tasks, ProvoGAN achieves average opinion scores of (4.13, 3.97, 3.93) where sGAN yields (3.33, 2.77, 2.67), vGAN yields (2.00, 1.93, 1.70), RefineGAN yields (3.63, 3.67, 3.13) and compressed-sensing reconstructions yield (2.00, 2.27, 2.13). Across synthesis tasks, ProvoGAN achieves average opinion scores of (3.73, 3.85, 4.10) whereas vGAN yields (1.50, 1.63, 1.33), SC-GAN yields (1.47, 1.43, 1.40) and REPLICA yields (2.40, 2.17, 1.83). Meanwhile, transverse sGAN models [3] maintain (2.90, 3.23, 3.38) and longitudinal sGAN models [4] yield (1.64, 1.81, 1.47). Overall, ProvoGAN outperforms all competing methods in synthesis ($p < 0.05$, Wilcoxon signed-ranked test) and reconstruction ($p < 0.05$) tasks, except for RefineGAN and SPIRIT in the in-vivo knee dataset where the three methods perform similarly ($p > 0.05$). In synthesis, ProvoGAN surpasses not only longitudinal sGAN models in transverse dimensions but also transverse sGAN models in transverse dimensions for which they have been optimized.

Radiological evaluations were also performed on recovered images from M^{3}NET and TransferGAN, as competing volumetrization baselines. Opinion scores for all volumetrization approaches in axial, coronal, and sagittal orientations denoted as (OS_A, OS_C, OS_S) are reported for the IXI dataset in Supp. Fig. 4. Across reconstruction tasks, ProvoGAN achieves average opinion scores of (4.00, 3.95, 3.90) where M^{3}NET yields (2.95, 3.25, 3.50) and TransferGAN yields (2.80, 2.75, 2.8). For synthesis, ProvoGAN achieves average opinion scores of (4.00, 4.3, 4.2) where M^{3}NET yields (2.90, 3.20, 3.00) and TransferGAN yields (1.90, 2.10, 1.9). Overall, ProvoGAN outperforms competing volumetrization methods in both synthesis ($p < 0.05$) and reconstruction ($p < 0.05$) tasks. Taken together, these findings strongly suggest that ProvoGAN can offer improved diagnostic quality in accelerated multi-contrast MRI protocols.

3.6. Complexity of cross-sectional mappings

Model complexity in deep neural networks depends on several architectural choices, including the number of layers, number of filters in each layer, and kernel size. To minimize bias in performance comparisons, here we aligned the architectural designs as closely as possible among the competing methods. To do this, the number of layers, number of filters, and kernel size were all kept fixed across methods, except that 2D convolutional kernels were used in sGAN and ProvoGAN whereas 3D convolutional kernels were used in vGAN (see Supp. Figs. 1,2). The precise parameter values were guided by the demanding vGAN model. We selected the parameter set that resulted in maximal model complexity while still allowing us to fit a single vGAN model into the

[3] In radiological evaluation, an sGAN model is called transverse for those opinion scores given for the orientation where that sGAN model is trained, e.g., sGAN-A for OS_A.

[4] In radiological evaluation, an sGAN model is called longitudinal for those opinion scores given for the orientation where that sGAN model is not trained, e.g., sGAN-A for OS_C or OS_S.

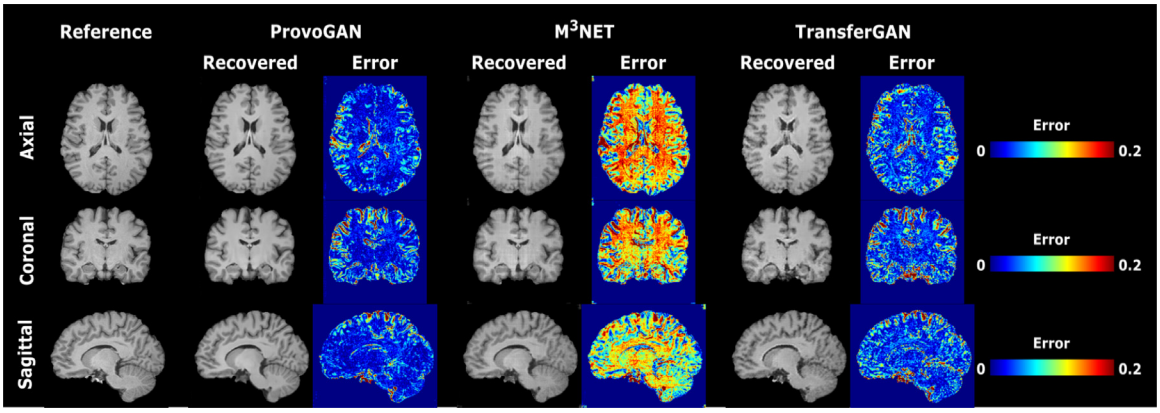

Fig. 6. The proposed ProvoGAN method is demonstrated on the IXI dataset against hybrid models (M^{3}NET and TransferGAN) for T_1-weighted image synthesis from T_2- and PD-weighted images. Representative results are displayed for all methods under comparison together with the ground truth target images (first column). The first row displays results for the axial orientation, the second row for the coronal orientation, and the third row for the sagittal orientation. Error was taken as the absolute difference between the reconstructed and reference images (see colorbar). Overall, the proposed method offers sharper and more accurate delineation of tissues than the competing methods. Furthermore, ProvoGAN better alleviates residual discontinuity artifacts compared to M^{3}NET.

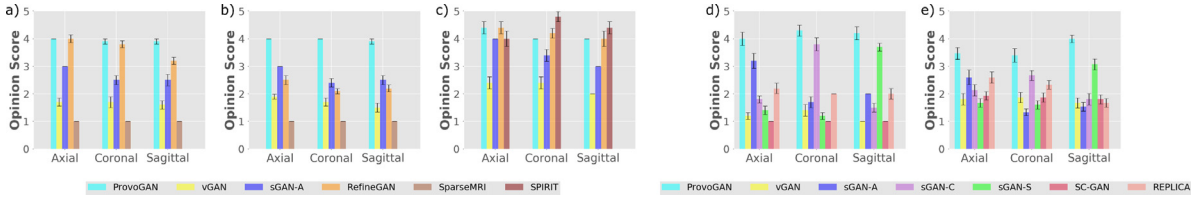

Fig. 7. Methods were compared in terms of radiological opinion scores for three reconstruction tasks: a) single-coil reconstruction of T_1-weighted images undersampled by $R = 8$ in the IXI dataset, b) single-coil reconstruction of T_2-weighted images undersampled by $R = 8$ in the IXI dataset, c) multi-coil reconstruction of PD-weighted images undersampled by $R = 8$ in the in vivo knee dataset, and for two synthesis tasks: d) many-to-one synthesis task of T_2, PD $\rightarrow$ T_1 in the IXI dataset, e) many-to-one synthesis task of T_1, T_2, T_{1c} $\rightarrow$ FLAIR in the in vivo brain dataset. The quality of the recovered axial, coronal, and sagittal cross-sections were rated by an expert radiologist by assessing their similarity to the reference cross-sections via a five-point scale (0: unacceptable, 1: very poor, 2: limited, 3: moderate, 4: good, 5: perfect match). Figure legends denote the colors used for marking the methods under comparison.

VRAM of the GPUs used to conduct the experiments here. Thus, it is reasonable to consider that vGAN is at its performance limits (see Supp. Tables 5,6). That said, a relevant question is whether and how the relative performance benefits of ProvoGAN over sGAN change with model complexity. To examine this issue, we performed separate experiments on reconstruction and synthesis tasks (see Section 2.7 for details) while systematically varying the complexity of the convolutional layers in both models by a factor of $n_f \in \{1/16, 1/9, 1/4, 1, 4, 9, 16\}$. This resulted in seven distinct pairs of models: ProvoGAN(n_f)-sGAN(n_f) with n_f-fold change in number of learnable network weights. PSNR and SSIM measurements between the recovered and reference volumes reported in Supp. Table 7 demonstrate that ProvoGAN achieves superior reconstruction performance to sGAN at all complexity levels, with on average 1.42 dB higher PSNR and 3.20% higher SSIM ($p < 0.05$, Wilcoxon signed-rank test). Meanwhile, PSNR and SSIM measurements reported in Supp. Table 8 indicate that ProvoGAN increases synthesis performance on average by 1.22 dB in PSNR, and 2.82% in SSIM compared to sGAN across complexity levels ($p < 0.05$). Taken together, these findings suggest that the benefits of ProvoGAN over sGAN in MRI recovery tasks are reliable across variations in complexity of network layers.

An alternative approach to help improve performance of cross-sectional models without substantially altering model complexity would be to admit inputs from multiple neighboring cross-sections. Given several neighboring cross-sections as input, this would enable a 2D model to incorporate local context in the vicinity of the central cross-section. To examine the utility of this approach in cross-sectional processing, we implemented multi-cross-section variants of the two methods, namely ProvoGAN(multi) and sGAN(multi). Both variants received as input three consecutive cross-sections and learned to recover the central cross-section of the target volume. We postulated that while this approach might increase sGAN performance to a limited degree, ProvoGAN that leverages broad spatial priors across all orientations should still yield superior performance. To test this prediction, we performed comprehensive experiments on the IXI dataset for reconstruction and synthesis tasks (see Section 2.7 for details). PSNR and SSIM measurements were performed between the recovered and reference target volumes (see Supp. Tables 9,10). Overall, ProvoGAN enhances recovery performance compared to sGAN(multi) in both tasks ($p < 0.05$), where it achieves on average 1.48 dB higher PSNR and 6.87% higher SSIM in reconstruction, and 0.87 dB higher PSNR and 1.71% higher SSIM in synthesis. These findings reveal that ProvoGAN outperforms cross-sectional mappings implemented with extended spatial priors across the longitudinal dimension. Note that ProvoGAN and ProvoGAN(multi) perform similarly across tasks ($p > 0.05$), whereas sGAN(multi) generally yields on par or higher performance than sGAN. This result suggests that sGAN processing each cross-section independently suffers from loss of spatial context across the longitudinal dimension, whereas sGAN(multi) improves performance by incorporating short-range context across this dimension. In contrast, ProvoGAN(multi) captures limited additional information from multiple neighboring cross-sections, since ProvoGAN readily captures global context across the volume.

3.7. Data efficiency

Volumetric models characteristically involve a substantial amount of parameters that result in heavier demand for training data for successful model training. Instead, ProvoGAN com-

Table 7
Data Efficiency of ProvoGAN for Reconstruction in the IXI Dataset: Volumetric PSNR (dB) and SSIM (%) measurements between the reconstructed and ground truth images in the IXI dataset are given as mean ± std across the test set. Measurements are reported for ProvoGAN and vGAN trained with varying number of subjects, and at four distinct acceleration factors ($R = 4, 8, 12, 16$). ProvoGAN(n_T) and vGAN(n_T) denote models trained with $n_t \in \{5, 15, 25\}$ subjects.

		$R = 4$		$R = 8$		$R = 12$		$R = 16$	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T_1	ProvoGAN(25)	34.37 ± 1.51	95.47 ± 0.63	31.63 ± 1.23	94.79 ± 1.02	29.07 ± 0.99	88.03 ± 1.61	28.29 ± 0.92	86.73 ± 1.58
	ProvoGAN(15)	34.04 ± 1.18	94.00 ± 0.84	31.13 ± 1.15	94.42 ± 1.05	28.96 ± 0.94	87.74 ± 1.61	27.96 ± 1.17	86.44 ± 1.50
	ProvoGAN(5)	33.65 ± 1.33	92.93 ± 1.03	30.72 ± 1.40	93.28 ± 0.98	28.97 ± 0.99	89.83 ± 1.39	28.21 ± 1.08	88.10 ± 1.35
	vGAN(25)	28.90 ± 1.29	83.27 ± 1.36	26.75 ± 1.29	81.91 ± 1.84	27.14 ± 0.89	79.92 ± 2.40	24.95 ± 1.43	77.50 ± 2.73
	vGAN(15)	28.52 ± 1.71	81.67 ± 1.59	26.26 ± 1.44	80.33 ± 2.22	24.59 ± 1.39	75.38 ± 2.79	23.98 ± 1.28	74.68 ± 2.53
	vGAN(5)	26.61 ± 1.25	74.71 ± 2.33	24.43 ± 1.14	73.56 ± 2.20	23.64 ± 1.27	68.63 ± 3.16	23.64 ± 1.18	69.31 ± 2.63
T_2	ProvoGAN(25)	36.32 ± 2.35	96.76 ± 0.72	32.23 ± 1.46	93.83 ±0.96	30.02 ± 1.04	91.90 ± 1.20	30.59 ± 1.41	93.58 ± 1.27
	ProvoGAN(15)	35.89 ± 2.69	96.49 ± 0.99	32.05 ± 1.27	93.50 ± 0.98	29.89 ± 1.08	91.31 ± 1.48	28.99 ± 0.89	90.81 ± 1.25
	ProvoGAN(5)	35.83 ± 3.03	96.00 ± 1.46	31.45 ±1.33	92.82 ± 1.26	29.54 ± 0.82	89.91 ± 1.53	28.53 ± 0.89	89.28 ± 1.80
	vGAN(25)	31.14 ±1.21	88.65 ±1.57	29.75 ± 0.73	89.11 ± 1.28	27.47 ± 0.60	84.13 ± 1.71	27.29 ± 0.79	85.72 ± 1.33
	vGAN(15)	30.25 ± 1.07	86.21 ± 1.87	28.76 ± 0.56	86.58 ± 1.50	26.89 ± 0.50	83.13 ± 1.57	26.67 ± 0.44	81.47 ± 1.81
	vGAN(5)	28.70 ± 0.55	82.03 ± 2.24	27.63 ± 0.67	82.67 ± 1.50	25.54 ± 0.66	79.22 ± 1.85	21.62 ± 1.05	77.10 ± 1.74

prises more compact cross-sectional models that can be efficiently trained on limited datasets. To demonstrate the data efficiency of ProvoGAN, we trained independent ProvoGAN and vGAN models while varying the number of training subjects in $n_T = \{5, 15, 25\}$, yielding ProvoGAN(n_T) and vGAN(n_T). For reconstruction, T_1- and T_2-weighted image recovery tasks in the IXI dataset at four distinct acceleration factors ($R = 4, 8, 12, 16$) were considered. Table 7 lists reconstruction performance for all models. As expected, model performance drops for both ProvoGAN and vGAN as number of training subjects is reduced. That said, ProvoGAN(n_T) outperforms vGAN(n_T) at all n_T values and in all tasks ($p < 0.05$, Wilcoxon signed-rank test), with 4.47 dB higher PSNR and 11.70% higher SSIM on average. Furthermore, the performance drop due to training with $n_T = 5$ versus $n_T = 25$ is merely 0.70 dB PSNR and 1.12% SSIM for ProvoGAN, and 2.70 dB PSNR and 7.87% SSIM for vGAN. Therefore, ProvoGAN better maintains its reconstruction performance on limited datasets to the extent that ProvoGAN models trained with $n_T = 5$ outperform vGAN models trained with $n_T = 25$. For synthesis, many-to-one recovery tasks of T_2, PD $\rightarrow$ T_1, T_1, PD $\rightarrow$ T_2, and T_1, T_2 $\rightarrow$ PD in the IXI dataset were considered. Measurements reported in Supp. Table 11 again indicate that ProvoGAN(n_T) outperforms vGAN(n_T) at all n_T values and in all tasks ($p < 0.05$), except for PSNR in T_1-weighted image recovery. On average, ProvoGAN models yield 1.76 dB higher PSNR and 2.51% higher SSIM compared to corresponding vGAN models. Moreover, the performance drop due to training with $n_T = 5$ versus $n_T = 25$ is merely 0.68 dB PSNR and 1.63% SSIM for ProvoGAN, and 1.19 dB PSNR and 2.32% SSIM for vGAN. Taken together, these results suggest that the proposed progressive volumetrization approach offers enhanced data efficiency during model training compared to volumetric models.

The primary factor contributing to the enhanced data efficiency of ProvoGAN is the reduced number of parameters in 2D versus 3D network architectures that elicit improved learning behavior. For the various reconstruction and synthesis tasks examined here, Supp. Table 12 lists comparisons between sGAN, ProvoGAN and vGAN in terms of model complexity, memory load, number of floating point operations per second (FLOPS), and total training time. Compared to vGAN, ProvoGAN reduces model complexity by 3-fold, memory load by 20-fold, FLOPS by 80-fold approximately. Collectively, these benefits empower ProvoGAN to offer improved learning behavior and computationally efficient inference while enhancing the capture of global context in volumetric images. Compared to sGAN, each progression in ProvoGAN naturally maintains the same model complexity and memory load. Yet, it has 3-fold higher FLOPS and train duration than sGAN due to ProvoGAN's sequential learning and inference process across the three rectilinear orientations, which also pushes its train duration beyond that of vGAN. As such, ProvoGAN offers a favorable compromise between contextual sensitivity and data efficiency, albeit at the expense of a prolonged training procedure.

3.8. Generalizability of progressive volumetrization

Here we primarily implemented ProvoGAN on a recent conditional GAN architecture with a ResNet backbone within the generator (Dar et al., 2019). That said, progressive volumetrization can be viewed as a model-agnostic approach that can be adapted to various 2D network architectures. To illustrate the generalizability of ProvoGAN, we performed progressive volumetrization on another state-of-the-art architecture SC-GAN with a U-Net backbone injected with self-attention layers (Lan et al., 2020). Cross-sectional (sSC-GAN), volumetric (vSC-GAN) and volumetrized (ProvoSC-GAN) variants of this architecture were built. Demonstrations were performed for T_1, T_2 $\rightarrow$ PD, T_1, PD $\rightarrow$ T_2, and T_2, PD $\rightarrow$ T_1 synthesis tasks on the IXI dataset. Resulting PSNR and SSIM measurements are listed in Table 8, where ProvoSC-GAN achieves in the range [0.35,2.42] dB higher PSNR and [0.47,3.05] % higher SSIM compared to sSC-GAN and vSC-GAN ($p < 0.05$). The superior synthesis quality offered by ProvoSC-GAN is also visible in representative results displayed in Fig. 8. Specifically, sSC-GAN models manifest discontinuity artifacts across the respective longitudinal dimensions,

Table 8
Progressive Volumetrization of the SC-GAN Architecture: Volumetric PSNR (dB) and SSIM (%) measurements between the synthesized and ground truth images in the test set in the IXI dataset are given as mean ± std. Measurements are provided for proposed and competing methods for all many-to-one synthesis tasks: 1) T_2, PD → T_1, 2) T_1, PD → T_2, 3) T_1, T_2 → PD. sSC-GAN-A denotes the sSC-GAN model trained in the axial orientation, sSC-GAN-C in the coronal orientation, and sSC-GAN-S in the sagittal orientation. Boldface indicates the highest performing method.

	ProvoSC-GAN		sSC-GAN-A		sSC-GAN-C		sSC-GAN-S		vSC-GAN	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
T_2, PD → T_1	**23.64**	**88.27**	22.74	85.63	23.03	86.94	23.30	87.53	22.58	85.32
	±3.12	**±4.50**	± 2.27	± 3.74	± 2.42	± 3.80	± 2.95	± 4.46	± 2.99	± 4.00
T_1, PD → T_2	**28.20**	**93.26**	27.79	92.64	27.72	92.45	28.09	93.22	25.29	89.81
	±2.97	**±4.53**	± 2.71	± 4.44	± 2.74	± 4.53	± 2.86	± 4.35	± 1.95	± 4.31
T_1, T_2 → PD	**29.66**	**94.86**	28.83	94.02	28.93	93.93	29.05	94.21	26.36	92.12
	±2.37	**±2.96**	± 2.12	± 2.86	± 2.11	± 2.95	± 1.92	± 2.46	± 1.41	± 2.59

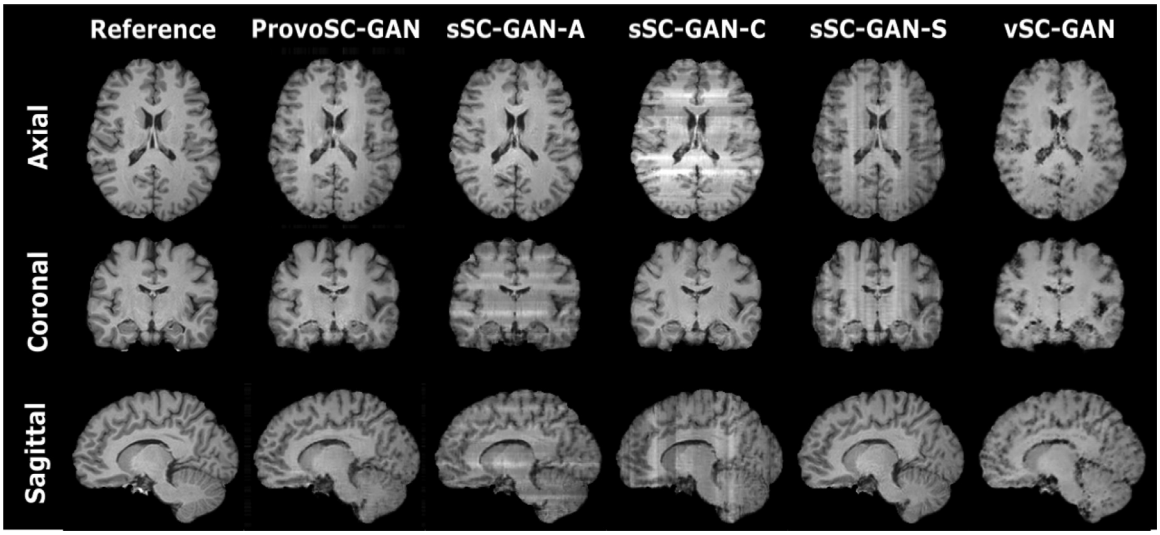

Fig. 8. Progressive volumetrization was performed on recently proposed SC-GAN architecture. Representative results for T_1-weighted image synthesis from T_2- and PD-weighted images in the IXI dataset are displayed. Results are shown for progressively volumetrized (ProvoSC-GAN), cross-sectional (sSC-GAN), and volumetric (vSC-GAN) models, along with the ground truth target images (first column). The first row displays results for the axial orientation, the second row for the coronal orientation, and the third row for the sagittal orientation. Overall, ProvoSC-GAN improves delineation of structural details compared to vSC-GAN, and enhances contextual consistency in the longitudinal dimensions compared to sSC-GAN models.

and vSC-GAN is suboptimal in recovering fine-structural details. In contrast, ProvoSC-GAN alleviates the limitations of both sSC-GAN and vSC-GAN models to enable more detailed and spatially-coherent tissue depiction. Taken together, these results strongly suggest that the proposed progressive volumetrization strategy can be extended to other network architectures while preserving its advantages against cross-sectional and volumetric mappings.

4. Discussion

Here, we introduced a progressively volumetrized deep generative model (ProvoGAN) for accelerated MRI that decomposes complex volumetric image recovery tasks into a series of cross-sectional mappings task-optimally ordered across individual rectilinear orientations. This progressive decomposition empowers ProvoGAN to learn both global contextual priors and fine-structural details in each orientation with enhanced data efficiency. Comprehensive evaluations on brain and knee MRI datasets illustrate the superior performance of ProvoGAN against state-of-the-art volumetric and cross-sectional models. Compared to volumetric models, ProvoGAN better captures fine structural details while at the same time maintaining lower instantaneous model complexity. As subtasks in ProvoGAN take single cross-sections as separate training samples, the effective size of the training set is expanded. Therefore, for a given model complexity, ProvoGAN demands an order of magnitude lower memory load than volumetric models. Compared to cross-sectional models, ProvoGAN mitigates discontinuity artifacts across the longitudinal dimensions and extends reliable capture of structural details from transverse onto longitudinal dimensions. Importantly, ProvoGAN offers this advanced recovery performance for the same budget of model complexity and memory load as cross-sectional models, albeit at the expense of a three-times prolonged training procedure due to sequential learning.

Several recent studies in medical image processing have focused on improving learning behavior in volumetric models. An earlier group of studies proposed spatially-focused 3D models to process volumetric patches during MRI recovery (Cordier et al., 2016; Huang et al., 2018; Jog et al., 2015; 2017; Roy et al., 2010; 2013; 2016; Vemulapalli et al., 2015; Ye et al., 2013). Patch-based models that restrict the spatial extent of network inputs-outputs can reduce model size to offer performance improvements. That said, a compact 3D patch incorporates context along the longitudinal axis at the expense of narrowing coverage in the in-plane dimensions. Since patches are processed independently, the predicted volumes might also manifest discontinuity artifacts. These limitations can undercut potential benefits of patch-based processing for 3D models. Later studies proposed hybrid models to bridge 2D and 3D models in an effort to combine their strengths (Wei et al., 2019; Peng et al., 2020; Shan et al., 2018; Liu et al., 2018). Among hybrid methods are fusion models that aggregate the outputs of parallel 2D models in multiple orientations (Peng et al., 2020; Wei et al., 2019). Fusion models employ a cascade of 2D

and 3D processing, so they incur high computational complexity, and sensitivity to fine structural details might be limited by the aggregation process across orientations. An alternative approach is transfer learning from 2D onto 3D models to facilitate model training (Shan et al., 2018; Liu et al., 2018). A full-scale 3D model is leveraged in transfer learning methods that lead to elevated model complexity, and a similar computational footprint to conventional 3D models. In contrast, ProvoGAN is composed of a sequence of 2D models, without any 3D module, resulting in substantially lower model complexity and computational load.

An alternative approach to volumetrization in medical imaging tasks has been to revise cross-sectional models to help them better incorporate spatial context. In (Zheng et al., 2018), enhanced spatial consistency during cardiac image segmentation was aimed by performing cross-sectional mapping on short-axis images sequentially across neighboring cross-sections. The segmentation map from the earlier cross-section was used to initialize the map for the current cross-section (Zheng et al., 2018). While benefits were demonstrated over 2D processing, this approach limits accumulation of contextual information to a single direction and to neighboring cross-sections. A different strategy for cardiac MRI segmentation was to perform cross-sectional mapping in short-axis orientation while latent representations captured via an autoencoder on a multitude of view orientations were fused at intermediate layers (Chen et al., 2019). The complexity of the resulting models scales with the number of additional views included, and this promising approach might be limited in applications where a multitude of different views on the same anatomy are unavailable. In (Prasoon et al., 2013; Wang et al., 2017; Xie et al., 2019), MR images at three rectilinear views that span across a target voxel were incorporated as inputs to a cross-sectional model during segmentation or classification tasks. Classifying a center voxel by fusing information across orientations might limit flow of contextual information from nearby voxels not covered by the input image views.

Our analyses involved brain and knee MRI datasets mostly collected at near-isotropic resolution. That said, the in vivo brain dataset was acquired under a diverse set of imaging protocols with varying spatial resolution. Note that images from T_2-weighted and FLAIR acquisitions had considerably poorer resolution in the longitudinal dimension, although they were registered to the MNI template with 1-mm isotropic resolution prior to modeling, In our experiments, models were built to synthesize T_1-weighted and T_{1c}-weighted images, where thick-slice acquisitions were on the input side. Models were also built to synthesize T_2-weighted and FLAIR images where thick-slice acquisitions were on the output side. We find that ProvoGAN offers enhanced recovery in both cases comprising a mixture of isotropic and anisotropic resolutions. Yet, contextual dependencies in the longitudinal dimension might be weaker for datasets uniformly acquired with thick slices, which in turn can limit the benefits of volumetrization. We plan to investigate this important issue in future studies by evaluating volumetrization performance on datasets with systematically varied slice thickness.

Several technical lines of development can be taken to further improve the performance and reliability of progressive volumetrization. In this study, ProvoGAN was independently demonstrated for mainstream MRI reconstruction and synthesis tasks. ProvoGAN can also be adopted for a joint reconstruction-synthesis task to further improve the utility and practicality of accelerated multi-contrast MRI protocols (Dar et al., 2020; Iglesias et al., 2021; Yurt et al., 2020). Here ProvoGAN was trained using a fully-supervised learning framework, which assumes the availability of datasets containing high-quality ground truth target images. However, compiling large datasets with high-quality references might prove difficult due to various concerns such as patient motion or examination costs (Liu et al., 2021). An alternative would be to train ProvoGAN in a self-supervised setting for reconstruction tasks (Yaman et al., 2020; Demirel et al., 2021; Cole et al., 2020; Korkmaz et al., 2022) or in a semi-supervised setting for synthesis tasks (Yurt et al., 2020) to alleviate dependency on high-quality training datasets. Another avenue of development concerns the generalization of ProvoGAN to work on nonrectilinear orientations (Ramzi et al., 2021; Sun et al., 2020; Motyka et al., 2021). While ProvoGAN was mainly demonstrated for rectilinear acquisitions in this work, similar decompositions can be viable for nonrectilinear sampling schemes in MRI such as radial and spiral acquisitions. Additionally, the number of total progressions in ProvoGAN can be adaptively modified together with the specific ordering of the orientations used in the progressions to enhance task-optimal recovery performance. Instead of performing a separate sequential training of each progression, an end-to-end training of the whole network can also be performed for improved performance by leveraging advanced model parallelism techniques (Zhu et al., 2020).

In this work, we demonstrated the proposed progressive volumetrization approach via a data-driven deep generative model that performs recovery in the image domain. Although image-to-image learning of deep models proves popular in MRI recovery tasks (Wang et al., 2021a; Yang et al., 2018a; Dai et al., 2020; Quan et al., 2018; Cole et al., 2021; Wang et al., 2020a), there are other successful approaches to MRI processing based on k-space-to-k-space learning (Han et al., 2020), k-space-to-image learning (Eo et al., 2018; Zhu et al., 2018; Akçakaya et al., 2019; Wang et al., 2019), or model-based learning with unrolled network architectures (Sun et al., 2016; Zhang and Ghanem, 2018; Aggarwal et al., 2018; Duan et al., 2019; Yang et al., 2018c). In principle, ProvoGAN can also be implemented to volumetrize models based on these recent powerful approaches. Thus, it remains important future work to investigate the potential benefits of progressive volumetrization to the contextual sensitivity of a broader family of recovery methods (Wang et al., 2021a; Yang et al., 2021; Chen et al., 2021a; Dalmaz et al., 2021; Tavaf et al., 2021; Güngör et al., 2021; Hu et al., 2021).

In summary, here we introduced a progressive volumetrization framework for deep network models to process 3D imaging datasets. The superior learning behavior of ProvoGAN was demonstrated for inverse problem solutions in two mainstream MRI tasks, reconstruction and synthesis. Yet, our framework can be adopted to other imaging modalities and tasks with minimal effort (Chung et al., 2020; Dewey et al., 2019; Wang et al., 2017; Xie et al., 2019; Qin et al., 2019; Singh et al., 2020; Zhao et al., 2020; Zheng et al., 2018; Zhou et al., 2020; Narnhofer et al., 2021). As the key idea of subtasking across cross-sectional orientations is domain general, ProvoGAN has further implications for computer vision applications that rely on 3D processing such as style transfer, semantic segmentation and video processing (Aberman et al., 2020; Chen et al., 2021b).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Mahmut Yurt: Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing – original draft. **Muzaffer Özbey:** Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing – original draft.

Salman U.H. Dar: Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing – original draft. **Berk Tinaz:** Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing – original draft. **Kader K. Oguz:** Formal analysis, Investigation, Validation, Writing – review & editing. **Tolga Çukur:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

Acknowledgments

This study was supported in part by a TUBITAK 1001 Research Grant (118E256), an EMBO Installation Grant (3028), a TUBA GEBIP 2015 fellowship, a BAGEP 2017 fellowship, and by nVidia under GPU grant.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2022.102429.

References

Aberman, K., Weng, Y., Lischinski, D., Cohen-Or, D., Chen, B., 2020. Unpaired motion style transfer from video to animation. ACM Trans. Graph. (TOG) 39 (4), 64. 1

Aggarwal, H.K., Mani, M.P., Jacob, M., 2018. MoDL: model-based deep learning architecture for inverse problems. IEEE Trans. Med. Imaging 38 (2), 394–405.

Akçakaya, M., Moeller, S., Weingärtner, S., Uğurbil, K., et al., 2019. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. Magn. Reson. Med. 81 (1), 439–453.

Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M., 2013. A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. 58 (13), R97–R129.

Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A., et al., 2018. Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans. Med. Imaging 37 (3), 803–814.

Chen, C., Biffi, C., Tarroni, G., Petersen, S., Bai, W., Rueckert, D., 2019. Learning shape priors for robust cardiac MR segmentation from multi-view images. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, Cham, pp. 523–531.

Chen, C., Raymond, C., Speier, B., Jin, X., Cloughesy, T. F., Enzmann, D., Ellingson, B. M., Arnold, C. W., 2021a. Synthesizing MR image contrast enhancement using 3D high-resolution ConvNets. arXiv:2104.01592.

Cheng, J. Y., Chen, F., Alley, M. T., Pauly, J. M., Vasanawala, S. S., 2018. Highly scalable image reconstruction using deep neural networks with bandpass filtering. arXiv:1805.03300v2.

Chen, Y., Firmin, D., Yang, G., 2021c. Wavelet improved GAN for MRI reconstruction. In: Medical Imaging 2021: Physics of Medical Imaging, vol. 11595. International Society for Optics and Photonics, p. 1159513.

Chen, J., Li, S., Liu, D., Lu, W., 2021b. Indoor camera pose estimation via style-transfer 3D models. Computer-Aided Civ. Infrastruct. Eng. 37 (3), 335–353.

Choi, Y., Uh, Y., Yoo, J., Ha, J.-W., 2020. StarGAN v2: diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8188–8197.

Chong, C.K., Ho, E.T.W., 2021. Synthesis of 3D MRI brain images with shape and texture generative adversarial deep neural networks. IEEE Access 9, 64747–64760.

Chung, H., Cha, E., Sunwoo, L., Ye, J. C., 2020. Two-stage deep learning for accelerated 3D time-of-flight MRA without matched training data. arXiv:2008.01362.

Cole, E., Cheng, J., Pauly, J., Vasanawala, S., 2021. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. Magn. Reson. Med. 86 (2), 1093–1109.

Cole, E. K., Pauly, J. M., Vasanawala, S. S., Ong, F., 2020. Unsupervised MRI reconstruction with generative adversarial networks. arXiv:2008.13065.

Cordier, N., Delingette, H., Le, M., Ayache, N., 2016. Extended modality propagation: image synthesis of pathological cases. IEEE Trans. Med. Imaging 35 (12), 2598–2608.

Dai, X., Lei, Y., Fu, Y., Curran, W.J., Liu, T., Mao, H., Yang, X., 2020. Multimodal MRI synthesis using unified generative adversarial networks. Med. Phys. 47 (12), 6343–6354.

Dalmaz, O., Yurt, M., Çukur, T., 2021. ResViT: residual vision transformers for multimodal medical image synthesis. arXiv:2106.16031.

Dar, S.U.H., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Çukur, T., et al., 2019. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 90 (312), 2375–2388.

Dar, S.U.H., Yurt, M., Shahdloo, M., Ildiz, M.E., Tinaz, B., Çukur, T., 2020. Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. IEEE J. Sel. Top. Signal Process. 14 (6), 1072–1087.

Demirel, O. B., Yaman, B., Dowdle, L., Moeller, S., Vizioli, L., Yacoub, E., Strupp, J., Olman, C. A., Uğurbil, K., Akçakaya, M., 2021. Improved simultaneous multi-slice functional MRI using self-supervised deep learning. arXiv:2105.04532.

Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., Sotirchos, E.S., Saidha, S., Oh, J., Pham, D.L., Calabresi, P.A., van Zijl, P.C.M., Prince, J.L., 2019. DeepHarmony: a deep learning approach to contrast harmonization across scanner changes. Magn. Reson. Imaging 64, 160–170.

Duan, J., Schlemper, J., Qin, C., Ouyang, C., Bai, W., Biffi, C., Bello, G., Statton, B., O'regan, D.P., Rueckert, D., 2019. VS-Net: variable splitting network for accelerated parallel MRI reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 713–722.

El-Rewaidy, H., Neisius, U., Mancio, J., Kucukseymen, S., Rodriguez, J., Paskavitz, A., Menze, B., Nezafat, R., 2020. Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI. NMR Biomed. 33 (7), e4312.

Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.-J., Hwang, D., 2018. KIKI-Net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn. Reson. Med. 80 (5), 2188–2201.

Epperson, K., Sawyer, A.M., Lustig, M., Alley, M., Uecker, M., Virtue, P., Lai, P., Vasanawala, S., 2013. Creation of fully sampled MR data repository for compressed sensing of the knee. In: 2013 Meeting Proceedings. Section for Magnetic Resonance Technologists.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. In: Advances in Neural Information Processing Systems, pp. 2672–2680.

Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. 47 (6), 1202–1210.

Güngör, A., Askin, B., Soydan, D. A., Saritas, E. U., Top, C. B., Çukur, T., 2021. TranSMS: transformers for super-resolution calibration in magnetic particle imaging. arXiv:2111.02163.

Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F., et al., 2017. Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. 79 (6), 3055–3071.

Han, Y., Sunwoo, L., Ye, J.C., 2020. K-space deep learning for accelerated MRI. IEEE Trans. Med. Imaging 39 (2), 377–386.

Han, Y., Yoo, J., Kim, H.H., Shin, H.J., Sung, K., Ye, J.C., 2018. Deep learning with domain adaptation for accelerated projection-reconstruction MR. Magn. Reson. Med. 80 (3), 1189–1205.

Hu, Y., Xu, Y., Tian, Q., Chen, F., Shi, X., Moran, C.J., Daniel, B.L., Hargreaves, B.A., 2021. RUN-UP: accelerated multishot diffusion-weighted MRI reconstruction using an unrolled network with U-Net as priors. Magn. Reson. Med. 85 (2), 709–720.

Huang, Y., Shao, L., Frangi, A.F., 2018. Cross-modality image synthesis via weakly–coupled and geometry co-regularized joint dictionary learning. IEEE Trans. Med. Imaging 37 (3), 815–827.

Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K., 2018. Deep learning for undersampled MRI reconstruction. Phys. Med. Biol. 63 (13), 135007.

Iglesias, J.E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., González, R.G., Alexander, D.C., Golland, P., Edlow, B.L., Fischl, B., et al., 2021. Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. NeuroImage 118206.

Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.

Jog, A., Carass, A., Roy, S., Dzung, L.P., Prince, J., 2015. MR image synthesis by contrast learning on neighborhood ensembles. Med. Image Anal. 24, 63–76.

Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L., et al., 2017. Random forest regression for magnetic resonance image synthesis. Med. Image Anal. 35, 475–488.

Joyce, T., Chartsias, A., Tsaftaris, S.A., 2017. Robust multi-modal MR image synthesis. In: Medical Image Computing and Computer-Assisted Intervention, pp. 347–355.

Korkmaz, Y., Dar, S.U.H., Yurt, M., Özbey, M., Çukur, T., 2022. Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. IEEE Trans. Med. Imaging 1. doi:10.1109/TMI.2022.3147426.

Küstner, T., Fuin, N., Hammernik, K., Bustin, A., Qi, H., Hajhosseiny, R., Masci, P.G., Neji, R., Rueckert, D., Botnar, R.M., et al., 2020. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. Sci. Rep. 10 (1), 1–13.

Lan, H., Toga, A. W., Sepehrband, F., 2020. SC-GAN: 3D self-attention conditional GAN with spectral normalization for multi-modal neuroimaging synthesis. bioRxiv.

Lee, D., Kim, J., Moon, W., Ye, J.C., 2019. Collagan: collaborative GAN for missing image data imputation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2482–2491.

Lee, D., Moon, W.-J., Ye, J.C., 2020. Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks. Nat. Mach. Intell. 2, 34–42.

Li, G., Lv, J., Wang, C., 2021. A modified generative adversarial network using spatial and channel-wise attention for CS-MRI reconstruction. IEEE Access 9, 83185–83198.

Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019. DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer International Publishing, Cham, pp. 795–803.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, J., Aviles-Rivero, A.I., Ji, H., Schönlieb, C.-B., 2021. Rethinking medical image reconstruction via shape prior, going deeper and faster: deep joint indirect registration and reconstruction. Med. Image Anal. 68, 101930.

Liu, S., Xu, D., Zhou, S.K., Pauly, O., Grbic, S., Mertelmeier, T., Wicklein, J., Jerebko, A., Cai, W., Comaniciu, D., 2018. 3D anisotropic hybrid network: transferring convolutional features from 2D images to 3D anisotropic volumes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 851–858.

Lustig, M., Donoho, D., Pauly, J.M., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. 58 (6), 1182–1195.

Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M., 2008. Compressed sensing MRI. IEEE Signal Process. Mag. 25 (2), 72–82.

Lustig, M., Pauly, J.M., 2010. SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary *k*-space. Magn. Reson. Med. 64 (2), 457–471.

Malavé, M.O., Baron, C.A., Koundinyan, S.P., Sandino, C.M., Ong, F., Cheng, J.Y., Nishimura, D.G., 2020. Reconstruction of undersampled 3D non-cartesian image-based navigators for coronary MRA using an unrolled deep learning model. Magn. Reson. Med. 84 (2), 800–812.

Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Smolley, S.P., 2017. Least squares generative adversarial networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2813–2821.

Mardani, M., Gong, E., Cheng, J.Y., Vasanawala, S.S., Zaharchuk, G., Xing, L., Pauly, J.M., 2019. Deep generative adversarial neural networks for compressive sensing MRI. IEEE Trans. Med. Imaging 38 (1), 167–179.

Mardani, M., Monajemi, H., Papyan, V., Vasanawala, S., Donoho, D., Pauly, J. M., et al., 2017. Recurrent generative adversarial networks for proximal learning and automated compressive image recovery. arXiv:1711.10046.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv:1411.1784.

Motyka, S., Hingerl, L., Strasser, B., Hangel, G., Heckova, E., Agibetov, A., Dorffner, G., Gruber, S., Trattning, S., Bogner, W., 2021. K-space-based coil combination via geometric deep learning for reconstruction of non-Cartesian MRSI data. Magn. Reson. Med. 86 (5), 2353–2367.

Narnhofer, D., Effland, A., Kobler, E., Hammernik, K., Knoll, F., Pock, T., 2021. Bayesian uncertainty estimation of learned variational MRI reconstruction. arXiv:2102.06665.

Olut, S., Sahin, Y.H., Demir, U., Unal, G., 2018. Generative adversarial training for MRA image synthesis using multi-contrast MRI. In: PRedictive Intelligence in MEdicine (PRIME). Springer International Publishing, Cham, pp. 147–154.

Peng, C., Lin, W., Liao, H., Chellappa, R., Zhou, S., 2020. SAINT: spatially aware interpolation network for medical slice synthesis. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, pp. 7747–7756.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M., 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 246–253.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. 42 (5), 952–962.

Qin, C., Schlemper, J., Caballero, J., Price, A.N., Hajnal, J.V., Rueckert, D., 2019. Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging 38 (1), 280–290.

Quan, T.M., Nguyen-Duc, T., Jeong, W.-K., 2018. Compressed sensing MRI reconstruction with cyclic loss in generative adversarial networks. IEEE Trans. Med. Imaging 37 (6), 1488–1497.

Ramzi, Z., Starck, J.-L., Ciuciu, P., 2021. Density compensated unrolled networks for non-cartesian MRI reconstruction. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1443–1447.

Roy, S., Carass, A., Prince, J., 2011. A compressed sensing approach for MR tissue contrast synthesis. In: Information Processing in Medical Imaging, pp. 371–383.

Roy, S., Carass, A., Prince, J.L., 2013. Magnetic resonance image example-based contrast synthesis. IEEE Trans. Med. Imaging 32 (12), 2348–2363.

Roy, S., Carass, A., Shiee, N., Pham, D.L., Prince, J.L., 2010. MR contrast synthesis for lesion segmentation. In: IEEE International Symposium on Biomedical Imaging, pp. 932–935.

Roy, S., Chou, Y.-Y., Jog, A., Butman, J., Dzung, L.P., 2016. Patch based synthesis of whole head MR images: application to EPI distortion correction. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (Eds.), Simulation and Synthesis in Medical Imaging, vol. 9968, pp. 146–156.

Schlemper, J., Caballero, J., Hajnal, J.V., Price, A., Rueckert, D., et al., 2017. A deep cascade of convolutional neural networks for MR image reconstruction. In: International Conference on Information Processing in Medical Imaging, pp. 647–658.

Sevetlidis, V., Giuffrida, M.V., Tsaftaris, S.A., 2016. Whole image synthesis using a deep encoder-decoder network. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 127–137.

Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M.K., Sun, L., Cong, W., Wang, G., 2018. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE Trans. Med. Imaging 37 (6), 1522–1534.

Sharma, A., Hamarneh, G., 2019. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. IEEE Trans. Med. Imaging 39, 1170–1183.

Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 1–11.

Singh, S.P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., Gulyás, B., 2020. 3D deep learning on medical images: a review. Sensors 20 (18), 5097.

Sood, R.R., Shao, W., Kunder, C., Teslovich, N.C., Wang, J.B., Soerensen, S.J.C., Madhuripan, N., Jawahar, A., Brooks, J.D., Ghanouni, P., et al., 2021. 3D registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction. Med. Image Anal. 69, 101957.

Sun, C., Yang, Y., Cai, X., Salerno, M., Meyer, C.H., Weller, D., Epstein, F.H., 2020. Non–cartesian slice-GRAPPA and slice-SPIRiT reconstruction methods for multiband spiral cardiac MRI. Magn. Reson. Med. 83 (4), 1235–1249.

Sun, J., Li, H., Xu, Z., et al., 2016. Deep ADMM-net for compressive sensing MRI. Adv. Neural Inf. Process. Syst. 29, 1–9.

Tavaf, N., Torfi, A., Ugurbil, K., Van de Moortele, P.-F., 2021. GRAPPA-GANs for parallel MRI reconstruction. arXiv:2101.03135.

Uecker, M., Lai, P., Murphy, M.J., Virtue, P., Elad, M., Pauly, J.M., Vasanawala, S.S., Lustig, M., 2014. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. Magn. Reson. Med. 71 (3), 990–1001.

Van Nguyen, H., Zhou, K., Vemulapalli, R., 2015. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer International Publishing, Cham, pp. 677–684.

Vemulapalli, R., Nguyen, H.V., Zhou, S.K., 2015. Unsupervised cross-modal synthesis of subject-specific scans. In: IEEE International Conference on Computer Vision, pp. 630–638.

Wang, G., Gong, E., Banerjee, S., Martin, D., Tong, E., Choi, J., Chen, H., Wintermark, M., Pauly, J.M., Zaharchuk, G., 2020. Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. IEEE Trans. Med. Imaging 39 (10), 3089–3099.

Wang, G., Ye, J.C., De Man, B., 2020. Deep learning for tomographic image reconstruction. Nat. Mach. Intell. 2, 737–748.

Wang, S., Ke, Z., Cheng, H., Jia, S., Ying, L., Zheng, H., Liang, D., 2019. DIMENSION: dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. NMR Biomed. e4131.

Wang, S., Xiao, T., Liu, Q., Zheng, H., 2021. Deep learning for fast MR imaging: a review for learning reconstruction from incomplete k-space data. Biomed. Signal Process. Control 68, 102579.

Wang, S., Zhou, M., Gevaert, O., Tang, Z., Dong, D., Liu, Z., Jie, T., 2017. A multi--view deep convolutional neural networks for lung nodule segmentation. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1752–1755.

Wang, T., Lei, Y., Fu, Y., Wynne, J.F., Curran, W.J., Liu, T., Yang, X., 2021. A review on medical imaging synthesis using deep learning and its clinical applications. J. Appl. Clin. Med. Phys. 22 (1), 11–36.

Wei, J., Xia, Y., Zhang, Y., 2019. M3Net: A multi-model, multi-size, and multi-view deep neural network for brain magnetic resonance image segmentation. Pattern Recognit. 91, 366–378.

Xie, Y., Xia, Y., Zhang, J., Song, Y., Feng, D.D.F., Fulham, M., Cai, W., 2019. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE Trans. Med. Imaging 38, 991–1004.

Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M., 2020. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn. Reson. Med. 84 (6), 3172–3191.

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., Firmin, D., 2018. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans. Med. Imaging 37 (6), 1310–1321.

Yang, H., Lu, X., Wang, S.-H., Lu, Z., Yao, J., Jiang, Y., Qian, P., 2020. Synthesizing multi-contrast MR images via novel 3D conditional variational auto-encoding GAN. Mob. Netw. Appl. 26 (1), 415–424.

Yang, H., Lu, X., Wang, S.-H., Lu, Z., Yao, J., Jiang, Y., Qian, P., 2021. Synthesizing multi-contrast mr images via novel 3D conditional variational auto-encoding GAN. Mob. Netw. Appl. 26 (1), 415–424.

Yang, Q., Li, N., Zhao, Z., Fan, X., I-Chao Chang, E., Xu, Y., 2018b. MRI image-to-image translation for cross-modality image registration and segmentation. arXiv:1801.06940v1.

Yang, Y., Sun, J., Li, H., Xu, Z., 2018. Admm-csnet: a deep learning approach for image compressive sensing. IEEE Trans. Pattern Anal. Mach. Intell. 42 (3), 521–538.

Ye, D.H., Zikic, D., Glocker, B., Criminisi, A., Konukoglu, E., 2013. Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 606–613.

Ye, J.C., 2019. Compressed sensing MRI: a review from signal processing perspective. BMC Biomed. Eng. 1, 1–17.

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: a review. Med. Image Anal. 58, 101552.

Yu, B., Zhou, L., Wang, L., Fripp, J., Bourgeat, P., 2018. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: IEEE 15th International Symposium on Biomedical Imaging. IEEE, pp. 626–630.

Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P., 2019. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging 38, 1750–1762.

Yurt, M., Dar, S.U.H., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T., 2021. mustGAN: multi-stream generative adversarial networks for MR image synthesis. Med. Image Anal. 70, 101944.

Yurt, M., Dar, S. U. H., Tnaz, B., Özbey, M., Çukur, T., 2020. Semi-supervised learning of mutually accelerated multi-contrast MRI synthesis without fully-sampled ground-truths. arXiv:2011.14347.

Zhan, B., Li, D., Wang, Y., Ma, Z., Wu, X., Zhou, J., Zhou, L., 2021a. LR-cGAN: latent representation based conditional generative adversarial network for multi--modality MRI synthesis. Biomed. Signal Process. Control 66, 102457.

Zhan, B., Li, D., Wu, X., Zhou, J., Wang, Y., 2021b. Multi-modal MRI image synthesis via GAN with multi-scale gate mergence. IEEE J. Biomed. Health Inform. 26 (1), 17–26.

Zhang, J., Ghanem, B., 2018. ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1828–1837.

Zhao, C., Dewey, B.E., Pham, D.L., Calabresi, P.A., Reich, D.S., Prince, J.L., 2020. SMORE: a self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning. IEEE Trans. Med. Imaging 40 (3), 805–817.

Zheng, Q., Delingette, H., Duchateau, N., Ayache, N., 2018. 3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation. IEEE Trans. Med. Imaging 37 (9), 2137–2148.

Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., Summers, R. M., 2020. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises.

Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S., 2018. Image reconstruction by domain-transform manifold learning. Nature 555 (7697), 487–492.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International Conference on Computer Vision, pp. 2223–2232.

Zhu, W., Zhao, C., Li, W., Roth, H., Xu, Z., Xu, D., 2020. LAMP: large deep nets with automated model parallelism for image segmentation. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Springer International Publishing, Cham, pp. 374–384.