Deformation-Compensated Learning for Image Reconstruction Without Ground Truth

Weijie Gan, *Graduate Student Member, IEEE*, Yu Sun, *Student Member, IEEE*, Cihat Eldeniz, Jiaming Liu, *Student Member, IEEE*, Hongyu An, and Ulugbek S. Kamilov, *Senior Member, IEEE*

***Abstract*—Deep neural networks for medical image reconstruction are traditionally trained using high-quality ground-truth images as training targets. Recent work on *Noise2Noise (N2N)* has shown the potential of using multiple noisy measurements of the same object as an alternative to having a ground-truth. However, existing N2N-based methods are not suitable for learning from the measurements of an object undergoing nonrigid deformation. This paper addresses this issue by proposing the *deformation-compensated learning (DeCoLearn)* method for training deep reconstruction networks by compensating for object deformations. A key component of DeCoLearn is a deep registration module, which is jointly trained with the deep reconstruction network without any ground-truth supervision. We validate DeCoLearn on both simulated and experimentally collected *magnetic resonance imaging (MRI)* data and show that it significantly improves imaging quality.**

***Index Terms*—Inverse problems, image reconstruction, deep learning, magnetic resonance imaging (MRI).**

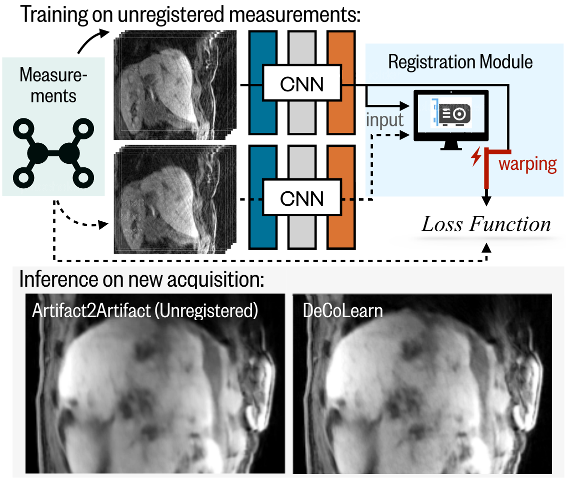


Fig. 1. The conceptual illustration of DeCoLearn for CS-MRI [1]. DeCoLearn trains a *convolutional neural network (CNN)* on unregistered measurements using a registration module that corrects for object deformation. This example highlights the improvement of DeCoLearn over an identical deep reconstruction network trained on the same measurements but without deformation compensation.

I. INTRODUCTION

THE recovery of a high-quality image from a set of noisy measurements is fundamental in medical imaging. For instance, it is essential in *compressed sensing magnetic resonance imaging (CS-MRI)* [1], which aims at obtaining diagnostic-quality images from severely undersampled k-space measurements. The recovery is traditionally formulated as an *inverse problem* that leverages a forward model characterizing the physics of data acquisition and a regularizer imposing prior knowledge on the solution. Many regularizers have been proposed to date, including those based on transform-domain sparsity, low-rank penalty, and dictionary learning [2]–[5].

Deep learning (DL) has recently gained popularity in medical image reconstruction [6]–[10]. A widely-used DL strategy is based on training a *convolutional neural network (CNN)* to map a low-quality image to its desired high-quality counterpart. However, this simple supervised DL approach is impractical in applications where it is difficult to collect a sufficient number of high-quality training images. This limitation has motivated the research on "ground-truth-free" DL schemes that rely exclusively on the information available in the corrupted data itself [11]–[15]. In this study, we focus on the line of work based on *Noise2Noise (N2N) [12]*, which has shown that one can train a CNN without ground-truth

---

Manuscript received 17 February 2022; accepted 24 March 2022. Date of publication 28 March 2022; date of current version 31 August 2022. This work was supported in part by the National Science Foundation (NSF) under Award CCF-2043134 and in part by the Washington University Institute of Clinical and Translational Sciences from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) under Grant UL1TR002345. *(Corresponding author: Ulugbek S. Kamilov.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Washington University in St. Louis Institutional under Application No. 202001098.

Weijie Gan and Yu Sun are with the Department of Computer Science & Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: weijie.gan@wustl.edu; sun.yu@wustl.edu).

Cihat Eldeniz is with the Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: cihat.eldeniz@wustl.edu).

Jiaming Liu is with the Department of Electrical & System Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: jiaming.liu@wustl.edu).

Hongyu An is with the Department of Neurology, Department of Biomedical Engineering, Mallinckrodt Institute of Radiology, Saint Louis, MO 63130 USA, and also with the Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: hongyuan@wustl.edu).

Ulugbek S. Kamilov is with the Department of Computer Science & Engineering and Electrical & Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: kamilov@ieee.org).

Digital Object Identifier 10.1109/TMI.2022.3163018

by using only pairs of noisy observations of the same object. Recent extensions to N2N have investigated the potential of this strategy in a variety of imaging scenarios [16]–[24].

Despite recent progress, current N2N-based methods inherently assume that the object is stationary across all the measurements. This assumption limits their ability to exploit measurements of an object undergoing nonrigid deformation. To overcome this limitation, we propose a new *deformation-compensated learning (DeCoLearn)* method that uses multiple measurements of a deformation-affected object by integrating a deep registration [25] module into the deep architecture for an end-to-end training. DeCoLearn enables training without any ground-truth supervision by adopting recent ideas from self-supervised deep registration [26]–[29]. The key contributions of this work are as follows:

- DeCoLearn extends N2N and its more recent variant *Artifact2Artifact (A2A)* [13] to enable learning directly in the measurement domain (e.g., k-space for MRI) from undersampled and noisy measurements without any fully sampled ground-truth. It is trained by transforming the reconstructed images back to the measurement domain and minimizing the difference between the predicted measurements and the measured raw data.
- DeCoLearn can use information from multiple measurements of an object undergoing nonrigid deformation, which enables it to leverage information that is not suitable for direct N2N/A2A training. This capability is achieved by integrating a deep registration module into the final architecture (see Fig. 2), which is trained end-to-end on unregistered, noisy, and subsampled measurements. Note that the registration module is only necessary during training, since image reconstruction can be performed by using only the reconstruction module.
- We extensively validate DeCoLearn on both simulated and experimentally collected MRI data. Our simulation results show that DeCoLearn quantitatively outperforms several baseline methods and matches the performance of oracle method that has the knowledge of the true object motion. Our results on experimentally collected data show that DeCoLearn leads to significant quality improvements by using additional measurements not suitable for traditional N2N-based learning.

This paper extends the preliminary work presented in the conference paper [30]. While [30] considered 2D single-coil uniformly-sampled MRI data, the DeCoLearn algorithm in this paper considers 3D multi-coil non-uniformly sampled MRI. Additionally, while the method in [30] was validated only on simulated data, here we present results on experimentally collected MRI data where deformations correspond to breathing. This paper also provides an expanded discussion of related work, new technical details, as well as new figures and tables.

II. BACKGROUND

A. Imaging Inverse Problems

We consider the problem of recovering an unknown image $\boldsymbol{x} \in \mathbb{C}^n$ from its noisy measurements $\boldsymbol{y} \in \mathbb{C}^m$ specified by the linear system

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{e} \in \mathbb{C}^m$ is noise and $\boldsymbol{H} \in \mathbb{C}^{m\times n}$ is the measurement operator that characterizes the response of the imaging system. For instance, $\boldsymbol{H}$ in parallel CS-MRI with a dynamic object can be represented as

$$\boldsymbol{H}_i^{(t)} = \boldsymbol{P}^{(t)}\boldsymbol{F}\boldsymbol{S}_i, \tag{2}$$

where $\boldsymbol{F}$ denotes the Fourier transform operator, $\boldsymbol{P}^{(t)}$ refers to a k-space sampling operator at time t, and $\boldsymbol{S}_i$ is the matrix of the pixel-wise sensitivity map of the ith coil. We assume that $\boldsymbol{S}_i$ is fixed over time. When $m < n$, the problem is an ill-posed inverse problem, which can be conventionally formulated as regularized optimization

$$\mathop{\arg\min}_{\boldsymbol{x}\in\mathbb{C}^n} \; \mathcal{D}(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x}), \tag{3}$$

where $\mathcal{D}$ is the data-fidelity term that quantifies consistency with the observed data $\boldsymbol{y}$ and $\mathcal{R}$ is a regularizer that encodes prior knowledge on $\boldsymbol{x}$. For example, two widely-used functions in imaging are the least-squares and total variation (TV)

$$\mathcal{D}(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\right\|_2^2 \quad \text{and} \quad \mathcal{R}(\boldsymbol{x}) = \tau \left\|\boldsymbol{D}\boldsymbol{x}\right\|_1, \tag{4}$$

where $\tau > 0$ controls the regularization strength and $\boldsymbol{D}$ is the discrete gradient operator [5].

In the past few years, DL has gained popularity for solving imaging inverse problems due to its excellent performance (see reviews in [6]–[10]). One widely-used DL approach is based on training a CNN $\mathrm{h}_{\boldsymbol{\theta}}(\cdot)$, with parameters $\boldsymbol{\theta} \in \mathbb{R}^p$, to compute a regularized inverse of $\boldsymbol{H}$ by mapping corrupted images to their clean target versions. The training can be formulated as an optimization problem

$$\mathop{\arg\min}_{\boldsymbol{\theta}} \sum_i \mathcal{L}(\mathrm{h}_{\boldsymbol{\theta}}(\boldsymbol{H}_i^{\dagger}\boldsymbol{y}_i), \boldsymbol{x}_i), \tag{5}$$

where $\boldsymbol{H}^{\dagger}$ is a pseudoinverse of $\boldsymbol{H}$, $\mathcal{L}$ is a loss function, and i indexes the samples in the training set. Popular choices for $\mathcal{L}$ include the ℓ_1 and ℓ_2 norms. For example, prior work on DL for CS-MRI has trained the CNN by mapping the zero-filled images to their corresponding fully-sampled ground-truth images [31]–[33]. While traditional DL relies on generic CNN architectures (such as UNet [34]), recent work has also explored the integration of DL and model-based optimization. For example, *plug-and-play priors (PnP)* [35] and *regularization by denoisers (RED)* [36] refer to a related family of algorithms that use pre-trained deep denoisers as imaging priors [37]–[40]. The recent publication [41] has reviewed PnP/RED in the context of image reconstruction for MRI. *Deep unrolling* is another widely-used strategy inspired by LISTA [42], where the iterations of a regularized optimization are interpreted as layers of a CNN and trained in an end-to-end fashion [31]–[33], [42]–[45].

Our work contributes to this broad area by providing a new DL method that does not require clean ground-truth images as training targets. While this work focuses on traditional model-free DL architectures, our method is fully compatible with the latest model-based architectures.

B. Deep Image Reconstruction Without Ground Truth

There is a growing interest in DL image reconstruction to reduce the dependence on high-quality ground-truth training targets. One widely-adopted framework is *N2N* [12], where the CNN $\mathrm{h}_{\boldsymbol{\theta}}$ is trained on a group of noisy images $\{\widehat{\boldsymbol{x}}_{ij}\}$, with j indexing different realizations of the same underlying image i. There have been multiple extensions of the original method [16]–[24] with applications to numerous medical imaging problems, including motion-resolved MRI [13], [17], cryo-transmission electron microscopy (cryo-TEM) [22] and optical coherence tomography angiography (OCTA) [21]. *A2A* [13] is one of the extensions of N2N that showed excellent performance using multiple noisy and artifact-corrupted images $\{\widehat{\boldsymbol{x}}_{ij}\}$ obtained directly from sparsely-sampled MR measurements. In A2A, ij denotes the j^{th} MRI acquisition of the subject i with each acquisition consisting a different undersampling pattern and noise realization. The whole dataset $\{\widehat{\boldsymbol{x}}_{ij}\}$ is assumed to compliment the information missing in each individual measurement, therefore enabling training of the CNN $\mathrm{h}_{\boldsymbol{\theta}}$ to predict clean images. The underlying assumption of N2N/A2A is that the expected value of the images $\{\widehat{\boldsymbol{x}}_{ij}\}_j$ still matches the ground-truth $\boldsymbol{x}_i$ [12]. The CNN in A2A is trained by minimizing a loss function

$$\underset{\boldsymbol{\theta}}{\arg\min} \sum_{i,j,j'} \mathcal{L}\big(\mathrm{h}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{x}}_{ij}),\ \widehat{\boldsymbol{x}}_{ij'}\big). \tag{6}$$

Recent works [15], [46] have shown the potential of training a model-based deep network without ground-truth by dividing a single k-space MRI acquisition into two subsets and using both subsampled sets of measurements as training targets. The same training strategy has been extended to the "zero-shot" learning and achieved excellent performance when training and testing datasets are highly inconsistent [47]. A similar strategy has also been used for denoising in 3D parallel-beam tomography by splitting a stack of noisy sinograms along the angular axis [23]. Two recent papers considered the inclusion of image deformation into the training of a deep image denoiser [20], [24]. In [20], a pre-trained registration network is used for training a video denoising network. In [24], a deep network is trained along with a deep deformation network to remove common types of noise in medical images, including additive white Gaussian noise (AWGN), Rician noise, and Poisson noise. The key difference of our work is that it goes beyond denoising by considering general inverse problems and using training labels directly in the k-space for MRI.

Noise2Void [14] and *Noise2Self* [48] are a related class of methods that use a single noisy copy of each training image in the dataset [49], [50]. Self2Self [51] extends this idea to use only a single noisy image as a training sample. These methods have been shown to achieve excellent performance in the context of image denoising. Since N2V-type methods learn only from a single image, they are expected to be suboptimal when dealing with structured artifacts, such as aliasing or streaks. We empirically verify this limitation of N2V in the context of accelerated MRI in Section IV.

Another related line of work is on *deep image prior (DIP)* [52], where a CNN is used for image reconstruction without any training on external data [53]–[55]. DIP exploits the architecture of the CNN to regularize the reconstruction by mapping random but fixed latent inputs to noisy measurements. A recent method TDDIP [54] extends DIP to dynamic MRI by compensating for the object motion by encoding the motion trajectory into the input latent variable. DIP is fundamentally different from DeCoLearn since it is not an end-to-end DL model and needs to solve a nonconvex optimization problem for each reconstruction task.

Our work contributes to this area by enabling the use of information from the measurements of an object undergoing nonrigid deformation. It not only allows our method to use more information for training, but also addresses the assumptions of stationarity and artifact incoherence in the prior work. It is worth mentioning that while in this paper we use a traditional CNN as the deep reconstruction network for DeCoLearn, the method itself is fully compatible with any model-based DL architectures [15].

C. Deep Image Registration

Let $\boldsymbol{r}$ and $\boldsymbol{m}$ denote a reference image and its deformed counterpart, respectively. Deformable image registration aims to obtain a registration field $\widehat{\boldsymbol{\phi}}^{m\rightarrow r}$ that maps the coordinates of $\boldsymbol{m}$ to those of $\boldsymbol{r}$ by comparing the content of the corresponding images. Deformable image registration has been widely-used in many applications, such as motion tracking [56] and image segmentation [57], [58]. The registration field $\widehat{\boldsymbol{\phi}}^{m\rightarrow r}$ is often characterized by a displacement vector field $\widehat{\boldsymbol{v}}^{m\rightarrow r}$ that represents coordinate offsets from $\boldsymbol{m}$ to $\boldsymbol{r}$, $\widehat{\boldsymbol{\phi}}^{m\rightarrow r} = \boldsymbol{I} + \widehat{\boldsymbol{v}}^{m\rightarrow r}$, where $\boldsymbol{I}$ denotes an identity transformation [59].

Recently, there has been considerable interest in developing DL methods for deformable image registration [25], especially methods that require no knowledge of the ground-truth transformation for training [26]–[29]. The corresponding self-supervised methods train a CNN $\mathrm{g}_{\boldsymbol{\varphi}}$, with parameters $\boldsymbol{\varphi} \in \mathbb{R}^k$, by mapping an input image pair $\{\boldsymbol{m}, \boldsymbol{r}\}$ to a deformation field $\widehat{\boldsymbol{\phi}}^{m\rightarrow r} = \mathrm{g}_{\boldsymbol{\varphi}}(\boldsymbol{m}, \boldsymbol{r})$ that can be used for registration [25]. The CNN is trained on a set of image pairs $\big\{\boldsymbol{m}_i, \boldsymbol{r}_i\big\}$ by minimizing the following loss function

$$\underset{\boldsymbol{\varphi}}{\arg\min} \sum_{i} \mathcal{L}_{\mathrm{d}}(\boldsymbol{m}_i \circ \widehat{\boldsymbol{\phi}}_i^{m\rightarrow r}, \boldsymbol{r}_i) + \mathcal{L}_{\mathrm{r}}(\widehat{\boldsymbol{\phi}}_i^{m\rightarrow r}), \tag{7}$$

where $\circ$ is the warping operator that transforms the coordinates of $\boldsymbol{m}_i$ based on the registration field $\widehat{\boldsymbol{\phi}}_i^{m\rightarrow r}$. The term $\mathcal{L}_{\mathrm{d}}$ penalizes the discrepancy between $\boldsymbol{m}_i$ after transformation and its reference $\boldsymbol{r}_i$, while $\mathcal{L}_{\mathrm{r}}$ regularizes the local spatial variations in the estimated registration field. In order to use the standard gradient methods for minimizing this loss function, the warping operator needs to be differentiable and is often implemented as the *Spatial Transform Network (STN)* [60].

Our work seeks to leverage the recent progress in deep image registration to enable a novel methodology for training deep reconstruction networks on deformation-affected datasets.

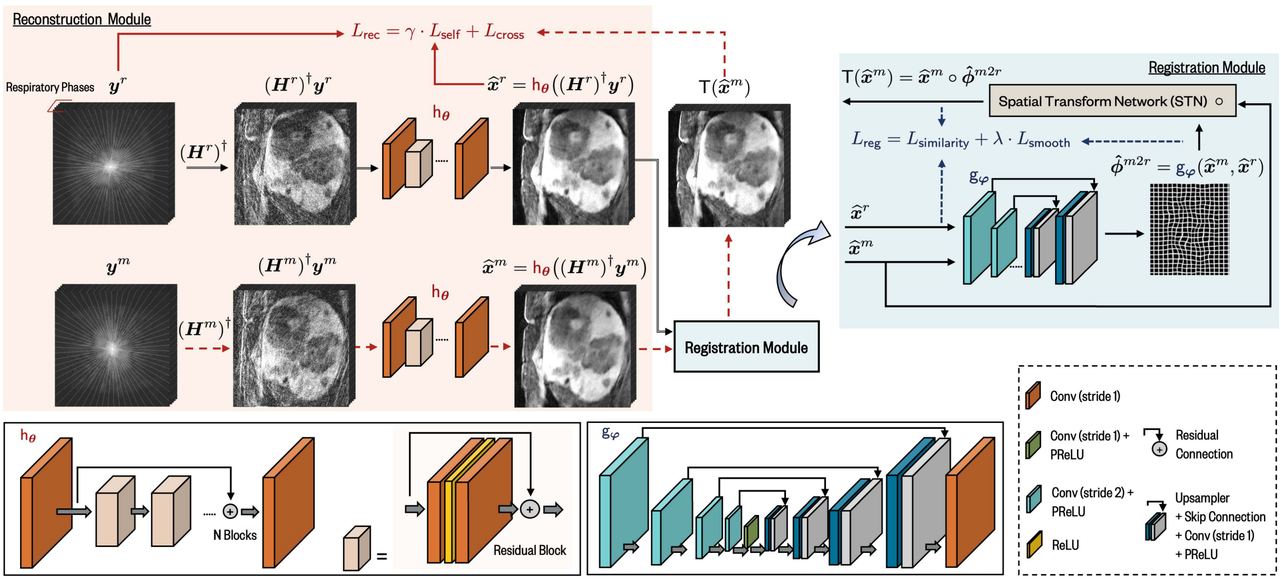


Fig. 2. The proposed method jointly trains two CNN modules: $\mathsf{h}_{\boldsymbol{\theta}}$ for image reconstruction and $\mathsf{g}_{\boldsymbol{\varphi}}$ for image registration. Inputs are the measurement pairs of the same object but at different motion states. The zero-filled images are passed through $\mathsf{h}_{\boldsymbol{\theta}}$ to remove artifacts due to noise and undersampling. The output images are then used in $\mathsf{g}_{\boldsymbol{\varphi}}$ to obtain the motion field characterizing the directional mapping between their coordinates. We implement the warping operator as the *Spatial Transform Network (STN)* to register one of the reconstructed images to the other. We train the whole network end-to-end without any ground-truth images or transformations.

D. Motion-Compensated Reconstruction

Motion-compensated (MoCo) reconstruction refers to a class of methods for reconstructing dynamic object from their noisy measurements [61]–[71]. MoCo methods seek to leverage data redundancy over the motion dimension during reconstruction. For example, traditional model-based MoCo methods include an additional regularizer in the motion dimension [61]–[63] or enforce spatial smoothness in the images at different motion phases using *motion vector fields (MVFs)* [64]–[66]. MVFs can be obtained by registering images of the reconstructed object at different motion states or via joint optimization using multi-task optimization [67]–[69]. Recent methods have also used DL to estimate MVFs by training a self-supervised network on reconstructed images [70] or by jointly updating both MVFs and images in a supervised fashion [71].

DeCoLearn is a complementary paradigm to the traditional MoCo image reconstruction. The primary focus of DeCoLearn is to enable *learning* given pairs of measurements of objects undergoing deformations. Thus, unlike MoCo methods, DeCoLearn does not specifically target *sequential* data. DeCoLearn can be used both as a traditional (non-MoCo) algorithm on 2D/3D spatial images or extended to explicitly take into account the motion/temporal dimension of the signal.

III. PROPOSED METHOD

In this section, we introduce the technical details of the proposed method. We start by describing the overall architecture, followed by the details of each module.

A. Overall Model

Consider a pair of unregistered measurements $(\boldsymbol{y}^r, \boldsymbol{y}^m)$ obtained separately from the same object

$$\boldsymbol{y}^r = \boldsymbol{H}^r \boldsymbol{x}^r + \boldsymbol{e}^r \quad \text{and} \tag{8a}$$

$$\boldsymbol{y}^m = \boldsymbol{H}^m \boldsymbol{x}^m + \boldsymbol{e}^m \quad \text{with } \boldsymbol{x}^m = \boldsymbol{x}^r \circ \boldsymbol{\phi}^{r\to m}, \tag{8b}$$

where $(\boldsymbol{H}^r, \boldsymbol{H}^m)$ and $(\boldsymbol{e}^r, \boldsymbol{e}^m)$ denote distinct forward operators and noise vectors, respectively. Eq. (8b) models the object motion as a dense nonrigid transformation-field $\boldsymbol{\phi}^{r\to m}$ relative to $\boldsymbol{x}^r$. For example, $(\boldsymbol{y}^r, \boldsymbol{y}^m)$ can be two motion-affected accelerated MRI measurements of the same patient. Our method aims to train a deep neural network on a set of such pairs $\{(\boldsymbol{y}_i^r, \boldsymbol{y}_i^m)\}_i^N$, where $N \geq 1$ denotes the total number of training samples, without the need for ground-truth images ($\boldsymbol{x}_i^r$ and $\boldsymbol{x}_i^m$) or transformations ($\boldsymbol{\phi}_i^{r\to m}$).

Fig. 2 summarizes the data processing pipeline of DeCoLearn. It consists of a reconstruction module trained to form images from measurements, and a registration module for registering the reconstructed images onto each other. The trainable parameters of both modules are denoted as $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ in respective order. During training, we define two distinct loss functions $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$ as well as two Adam [72] optimizers Adam_{rec} and Adam_{reg} for each module. Given a mini-batch of training samples, the proposed training procedure alternatively minimizes the loss functions by fixing the trainable parameters of one module while training the other. Algorithm 1 summarizes the training strategy. Note that the registration module of DeCoLearn is only employed during training, since reconstruction during testing can be performed directly by using the reconstruction module alone.

B. Reconstruction Module

During training, the reconstruction module separately takes two measurements $\boldsymbol{y}^r$ and $\boldsymbol{y}^m$ described in (8) as inputs to produce two images $\widehat{\boldsymbol{x}}^r$ and $\widehat{\boldsymbol{x}}^m$ as outputs, respectively. The measurements are first mapped to the image domain by applying the pseudoinverse of their respective forward operators. We denote with $(\boldsymbol{H}^m)^\dagger \boldsymbol{y}^m$ and $(\boldsymbol{H}^r)^\dagger \boldsymbol{y}^r$ the resulting

Algorithm 1 DeCoLearn Training

```
Require: Initial parameters $\boldsymbol{\theta}^0$ and $\boldsymbol{\varphi}^0$, number of iterations $K$,
    and Adam [72] optimizers $\mathrm{Adam}_{\mathrm{reg}}$ and $\mathrm{Adam}_{\mathrm{rec}}$.
for number of training iterations $k = 1, 2, \ldots, K$ do
    Select a training mini-batch: $\boldsymbol{y}_i^r, \boldsymbol{y}_i^m, \boldsymbol{H}_i^r, \boldsymbol{H}_i^m$
    $\boldsymbol{\theta}^k \leftarrow \mathrm{Adam}_{\mathrm{rec}}(\boldsymbol{\theta}^{k-1}, \partial L_{\mathrm{rec}}/\partial\boldsymbol{\theta})$
    $\boldsymbol{\varphi}^k \leftarrow \mathrm{Adam}_{\mathrm{reg}}(\boldsymbol{\varphi}^{k-1}, \partial L_{\mathrm{reg}}/\partial\boldsymbol{\varphi})$
end for
return Learned parameters $\boldsymbol{\theta}^K$ and $\boldsymbol{\varphi}^K$.
```

artifact-corrupted images in the image domain. A CNN $\mathrm{h}_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta} \in \mathbb{R}^p$ is then trained to remove the artifacts from the corrupted images

$$\widehat{\boldsymbol{x}}^m = \mathrm{h}_{\boldsymbol{\theta}}\big((\boldsymbol{H}^m)^\dagger \boldsymbol{y}^m\big) \quad \text{and } \widehat{\boldsymbol{x}}^r = \mathrm{h}_{\boldsymbol{\theta}}\big((\boldsymbol{H}^r)^\dagger \boldsymbol{y}^r\big). \tag{9}$$

Our network is a customized version of the residual CNN used in the prior work on deep image reconstruction [13], [15], [73].

Since the underlying true images $\boldsymbol{x}^m$ and $\boldsymbol{x}^r$ are unregistered, their reconstructed versions $\widehat{\boldsymbol{x}}^m$ and $\widehat{\boldsymbol{x}}^r$ obtained from $\mathrm{h}_{\boldsymbol{\theta}}$ are also unregistered. Therefore, it is suboptimal to construct a loss function to directly compare the pixel-wise difference between $\widehat{\boldsymbol{x}}^m$ and $\widehat{\boldsymbol{x}}^r$. It is thus necessary to use the registration module to mitigate their potential misalignment. We define $\mathrm{T}(\widehat{\boldsymbol{x}}^r)$ and $\mathrm{T}(\widehat{\boldsymbol{x}}^m)$ as the images transformed according to the estimated deformation field (see details in Sec. III-C). In our notation, $\mathrm{T}(\widehat{\boldsymbol{x}}^r)$ denotes a transformed variant of $\widehat{\boldsymbol{x}}^r$ relative to $\widehat{\boldsymbol{x}}^m$.

The loss function $\mathcal{L}_{\mathrm{rec}}$ of $\mathrm{h}_{\boldsymbol{\theta}}$ has two components

$$\mathcal{L}_{\mathrm{rec}} = \mathcal{L}_{\mathrm{cross}} + \gamma \cdot \mathcal{L}_{\mathrm{self}}, \tag{10}$$

where the parameter $\gamma > 0$ controls the relative strength of each component. The function $\mathcal{L}_{\mathrm{cross}}$ is the main component that penalizes the difference between the raw data and the transformed reconstructed image at a different motion state

$$\mathcal{L}_{\mathrm{cross}} = \sum_{i=1}^{N} \mathcal{L}\big(\boldsymbol{y}_i^r, \boldsymbol{H}_i^r \mathrm{T}(\widehat{\boldsymbol{x}}_i^m)\big) + \mathcal{L}\big(\boldsymbol{y}_i^m, \boldsymbol{H}_i^m \mathrm{T}(\widehat{\boldsymbol{x}}_i^r)\big), \tag{11}$$

where $\boldsymbol{H}_i^m$ and $\boldsymbol{H}_i^r$ are the forward operators used to map the registered images back to the measurement domain. Eq. (11) maps pairs of measurements having the forms (8a) and (8b) by assuming that the deformations between them have been accounted for via the registration module. The function $\mathcal{L}_{\mathrm{self}}$ penalizes the discrepancy between the measurements estimated from a reconstructed image and the corresponding actual raw measurements

$$\mathcal{L}_{\mathrm{self}} = \sum_{i=1}^{N} \mathcal{L}\big(\boldsymbol{y}_i^r, \boldsymbol{H}_i^r \widehat{\boldsymbol{x}}_i^r\big) + \mathcal{L}\big(\boldsymbol{y}_i^m, \boldsymbol{H}_i^m \widehat{\boldsymbol{x}}_i^m\big). \tag{12}$$

Note that N2N/A2A can be seen as special cases of the proposed method where the potential deformations between the measurements are set to identity.

C. Registration Module

Our registration module builds on self-supervised deep image registration discussed in Sec. II-C, which consists of a CNN $\mathrm{g}_{\boldsymbol{\varphi}}$, customized from U-net [34] with trainable parameters $\boldsymbol{\varphi} \in \mathbb{R}^q$, and a *Spatial Transform Network (STN)* [60]. As its order-sensitive input, the network accepts a pair of reconstructed images $(\widehat{\boldsymbol{x}}^m, \widehat{\boldsymbol{x}}^r)$ estimated using $\mathrm{h}_{\boldsymbol{\theta}}$ and registers them onto each other. The network $\mathrm{g}_{\boldsymbol{\varphi}}$ uses two inputs in different orders to generate two motion fields

$$\hat{\boldsymbol{\phi}}^{m\to r} = \mathrm{g}_{\boldsymbol{\varphi}}(\widehat{\boldsymbol{x}}^m, \widehat{\boldsymbol{x}}^r) \quad \text{and } \hat{\boldsymbol{\phi}}^{r\to m} = \mathrm{g}_{\boldsymbol{\varphi}}(\widehat{\boldsymbol{x}}^r, \widehat{\boldsymbol{x}}^m) \tag{13}$$

that characterize two coordinate mappings with opposite directions relative to each other. For example, $\hat{\boldsymbol{\phi}}^{m\to r}$ denotes a directional mapping from the coordinates of $\widehat{\boldsymbol{x}}^m$ to those of $\widehat{\boldsymbol{x}}^r$. STN then transforms the coordinate of inputs based on the motion fields and obtains their registered variants

$$\mathrm{T}(\widehat{\boldsymbol{x}}^m) = \widehat{\boldsymbol{x}}^m \circ \hat{\boldsymbol{\phi}}^{m\to r} \quad \text{and } \mathrm{T}(\widehat{\boldsymbol{x}}^r) = \widehat{\boldsymbol{x}}^r \circ \hat{\boldsymbol{\phi}}^{r\to m}. \tag{14}$$

The loss function $\mathcal{L}_{\mathrm{reg}}$ for training $\boldsymbol{g}_{\boldsymbol{\varphi}}$ is specified as

$$\mathcal{L}_{\mathrm{reg}} = \mathcal{L}_{\mathrm{similarity}} + \lambda \cdot \mathcal{L}_{\mathrm{smooth}}, \tag{15}$$

where $\mathcal{L}_{\mathrm{similarity}}$ enforces similarity between registered images and their references, $\mathcal{L}_{\mathrm{smooth}}$ enforces spatial smoothness in the motion field, and $\lambda > 0$ is a regularization parameter. The function $\mathcal{L}_{\mathrm{similarity}}$ is given by

$$\mathcal{L}_{\mathrm{similarity}} = -\sum_i \big(\mathrm{LCC}(\mathrm{T}(\widehat{\boldsymbol{x}}^m), \widehat{\boldsymbol{x}}_i^r) + \mathrm{LCC}(\mathrm{T}(\widehat{\boldsymbol{x}}^r), \widehat{\boldsymbol{x}}_i^m)\big). \tag{16}$$

where LCC denotes the local cross-correlation (LCC) [29], which is known to be robust to intensity variations across different acquisitions [74]. While minimizing $L_{\mathrm{similarity}}$ enforces accurate alignment, it can also generate non-smooth registration fields that are not physically realistic [29]. Therefore, we include the function $\mathcal{L}_{\mathrm{smooth}}$ that imposes smoothness on the coordinate offsets $\hat{\boldsymbol{v}} = \hat{\boldsymbol{\phi}} - \boldsymbol{I}$

$$\mathcal{L}_{\mathrm{smooth}} = \sum_i \Big(\big\|\boldsymbol{D}\hat{\boldsymbol{v}}_i^{m\to r}\big\|^2 + \big\|\boldsymbol{D}\hat{\boldsymbol{v}}_i^{r\to m}\big\|^2 \Big). \tag{17}$$

IV. EXPERIMENTAL VALIDATION

We validate our method in the context of accelerated MRI. We consider three settings: *(a) 2D simulated measurements and simulated deformations*; *(b) 2D simulated measurements and real unknown deformations*; and *(c) 3D experimentally collected measurements and real unknown deformations*.

A. Setup

1) Baseline Methods: We used several well-known image reconstruction methods for comparison

(a) *TV/CS*: The traditional total variation regularization method is summarized in eq. (4). On the experimentally collected free-breathing MRI data, we replace the basic TV with the *compressed sensing (CS)* method from [76]. Similarly to the well-known XD-GRASP method [61], CS exploits regularization along the motion dimension to significantly boost reconstruction performance.

TABLE I

AVERAGE PSNR AND SSIM VALUES OBTAINED OVER THE TEST SET. THE TABLE HIGHLIGHTS THAT DECOLEARN OUTPERFORMS SEVERAL WELL-KNOWN BASELINE METHODS AT DIFFERENT ACCELERATION FACTORS AND SYNTHETIC DEFORMATION MAGNITUDES

Experiment of Simulated Measurement and Simulated Deformation

Schemes	PSNR						SSIM					
Synthetic Deformable with $\sigma =$	10		18		24		10		18		24	
Acceleration rate	x3	x4	x3	x4	x3	x4	x3	x4	x3	x4	x3	x4
Zero-Filled	28.20*	26.03*	28.19*	26.02*	28.28*	26.05*	0.772*	0.717*	0.772*	0.715*	0.774*	0.716*
Total Variation	33.01*	29.78*	32.96*	29.79*	33.18*	29.82*	0.942*	0.893*	0.941*	0.893*	0.944*	0.894*
N2V [14]	28.19*	26.07*	28.19*	26.03*	28.35*	26.04*	0.774*	0.719*	0.774*	0.716*	0.778*	0.717*
DIP [52]	32.64*	30.45*	32.87*	30.71*	33.05*	30.93*	0.913*	0.869*	0.915*	0.871*	0.915*	0.870*
Self-Supervised	31.41*	29.58*	31.28*	28.92*	31.62*	29.74*	0.925*	0.922*	0.942*	0.910*	0.946*	0.908*
SSDU [15], [46]	32.98*	30.37*	32.92*	30.87*	33.13*	30.98*	0.956†	0.939	0.954*	0.943	0.959†	0.944
DeCoLearn	**33.71**	**31.60**	**33.85**	**31.67**	**34.04**	**31.72**	**0.962**	**0.945**	**0.965**	**0.947**	**0.964**	**0.949**

Statistically significant differences compared with DeCoLearn are marked ($^*P < 0.0001;^\dagger P < 0.05$).

(b) *SSDU/Self-Supervised* [46][1]: A recent self-supervised method that trains a *deep unrolling network* by dividing each k-space MRI acquisition into two subsets and using them as training targets for each other. *Self-Supervised* is a variant of SSDU that uses the same reconstruction CNN as DeCoLearn. Having both methods allows to separate the influence of the deep unrolling architecture from that of the training scheme on the SSDU performance.

(c) *DIP/TDDIP* [54][2]: DIP is an image reconstruction method that uses an untrained CNN as a regularizer. We use an improved variant of DIP on our simulated data where two i.i.d. latent vectors are mapped to different measurements of the same subject. TDDIP is a recent extension of DIP that improves performance by taking into account the motion dimension in the image sequence. We use TDDIP on our experimentally-collected MRI data by sampling the latent inputs in the straight-line manifold due to the acyclic nature of the respiratory motion occurred in the dataset [54].

(d) *Noise2Void (N2V)* [14][3]: An alternative to N2N that trains image restoration CNNs by mapping noisy pixels to their randomly-selected neighbors. Unlike N2N, N2V does not require paired data, but inherently assumes that artifacts are spatially unstructured—an assumption that does not hold for aliasing and streaking artifacts in MRI.

We also performed an ablation study to highlight the influence of the registration module within DeCoLearn. The ablated methods can be divided into three categories.

- **Registration-free methods:**
  (i) *A2A (Unregistered)*: The most basic variant of A2A, trained directly on unregistered measurements. It can be interpreted as the worst-case scenario for DeCoLearn when no deformation-compensation is performed during training.

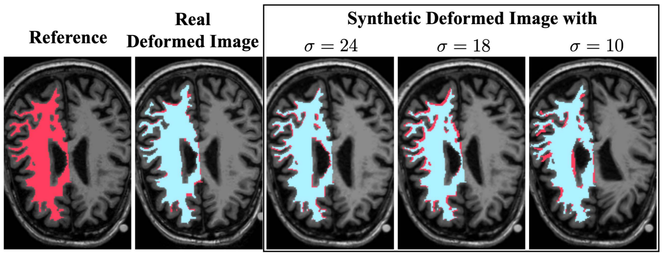


Fig. 3. Visual illustration of deformations in the simulated experiments. The red regions are segmentations in the reference, while the blue regions are the corresponding segmentations in the deformed counterparts. The synthetic deformations were generated by using the method in [75], where σ is inversely related to the deformation strengths. The *in vivo* deformation is due to normal aging and disease.

- **Pre-registration methods:** In this category, we explore the use of a fixed registration module that provides motion field estimates during the A2A training.
  (ii) *A2A (Affine)*: Uses Affine algorithms implemented in *advanced normalization tools (ANTS)* [77].
  (iii) *A2A (SyN)*: Similar to *A2A (Affine)*, but uses *Symmetric Normalization (SyN)* [74] algorithm instead.
  (iv) *A2A (VoxelMorph)*: Uses a deep registration method from [29] pre-trained on artifact-corrupted images.
- **Oracle-registration method:**
  (v) *A2A (Oracle)*: A2A (Oracle) is the idealized variant of DeCoLearn using the registration model that provides perfect results. In our simulations, we synthesized the *registered* data by applying different measurement operators on the same ground-truth image with no motion. Note that this method is *not* applicable to the experimental data as the ground-truth is unavailable.

2) Evaluation Metrics: In simulations, we implemented two widely-used quantitative metrics, *peak signal-to-noise ratio (PSNR)*, measured in dB and *structural similarity index (SSIM)*, relative to the ground-truth images used to synthesize the measurements. The quantitative results were statistically analyzed by comparing DeCoLearn to other image reconstruction methods. We used the non-parametric Friedman's test and the post-hoc test of the original FDR method of Benjamini

---

[1] We use the SSDU implementation at github.com/byaman14/SSDU.

[2] We use the TDDIP implementation at github.com/jaejun-yoo/TDDIP.

[3] We use the Noise2Void implementation at github.com/juglab/n2v.

TABLE II

QUANTITATIVE RESULTS OF AN ABLATION STUDY SHOWING INFLUENCE OF THE REGISTRATION MODULE. THE TABLE SHOWS THAT DECOLEARN NEARLY MATCHES THE PERFORMANCE OF THE IDEALIZED *A2A (ORACLE)* METHOD, WHICH USES THE TRUE DEFORMATIONS

Experiment of Simulated Measurement and Simulated Deformation

Schemes	PSNR						SSIM					
Synthetic Deformable with $\sigma =$	10		18		24		10		18		24	
Acceleration rate	x3	x4	x3	x4	x3	x4	x3	x4	x3	x4	x3	x4
A2A (Unregistered)	30.19	29.07	31.96	30.37	32.83	30.89	0.921	0.903	0.942	0.926	0.954	0.935
A2A (Affine)	30.42	29.14	32.50	30.67	33.42	31.20	0.922	0.900	0.950	0.932	0.959	0.940
A2A (SyN)	32.70	30.31	32.71	30.35	32.85	30.39	0.952	0.929	0.957	0.932	0.956	0.933
A2A (VoxelMorph)	32.44	30.26	33.06	30.67	33.16	31.03	0.950	0.928	0.958	0.936	0.957	0.938
DeCoLearn	**33.71**	**31.60**	**33.85**	**31.67**	**34.04**	**31.72**	**0.962**	**0.945**	**0.965**	**0.947**	**0.964**	**0.949**
A2A (Oracle)‡	34.17	31.89	34.20	31.91	34.29	31.93	0.965	0.948	0.965	0.948	0.966	0.949

‡: idealized algorithm, not available in practice.

TABLE III

AVERAGE PSNR AND SSIM VALUES OBTAINED OVER THE TEST SET. NOTE HOW DECOLEARN ACHIEVES BETTER PERFORMANCE THAN ALL THE METHODS AT DIFFERENT ACCELERATION FACTORS. THE DEFORMATIONS CONSIDERED IN THIS TABLE ARE *IN VIVO* DUE TO NORMAL AGING AND DISEASE

Experiment of Simulated Measurement and Real Deformation

Schemes	PSNR		SSIM	
Acceleration rate	x3	x4	x3	x4
Zero-Filled	27.85*	25.70*	0.757*	0.702*
Total Variation	32.72*	29.49*	0.943*	0.892*
N2V [14]	27.82*	25.69*	0.760*	0.703*
DIP [52]	31.76*	30.70†	0.903*	0.876*
Self-Supervised	31.16*	29.36*	0.950*	0.929*
SSDU [15], [46]	32.51*	30.18*	0.959†	0.945
DeCoLearn	**33.23**	**31.19**	**0.966**	**0.949**

Statistically significant differences compared with DeCoLearn are marked ($^*P < 0.0001$;$^\dagger P < 0.05$).

TABLE IV

QUANTITATIVE RESULTS FROM AN ABLATION STUDY EVALUATING THE INFLUENCE OF REGISTRATION. NOTE HOW DECOLEARN ACHIEVES COMPARABLE PERFORMANCE TO *A2A (ORACLE)*, WHICH, UNLIKE DECOLEARN, RELIES ON REGISTRATION INFORMATION OBTAINED FROM THE GROUND-TRUTH. THE DEFORMATIONS CONSIDERED IN THIS TABLE ARE *IN VIVO* DUE TO NORMAL AGING AND DISEASE

Experiment of Simulated Measurement and Real Deformation

Schemes	PSNR		SSIM	
Acceleration rate	x3	x4	x3	x4
A2A (Unregistered)	31.94	30.05	0.953	0.932
A2A (Affine)	29.97	28.87	0.944	0.925
A2A (SyN)	30.66	28.88	0.947	0.921
A2A (VoxelMorph)	30.36	28.97	0.943	0/927
DeCoLearn	**33.23**	**31.19**	**0.966**	**0.949**
A2A (Oracle)‡	33.85	31.52	0.966	0.949

‡: idealized algorithm, not available in practice.

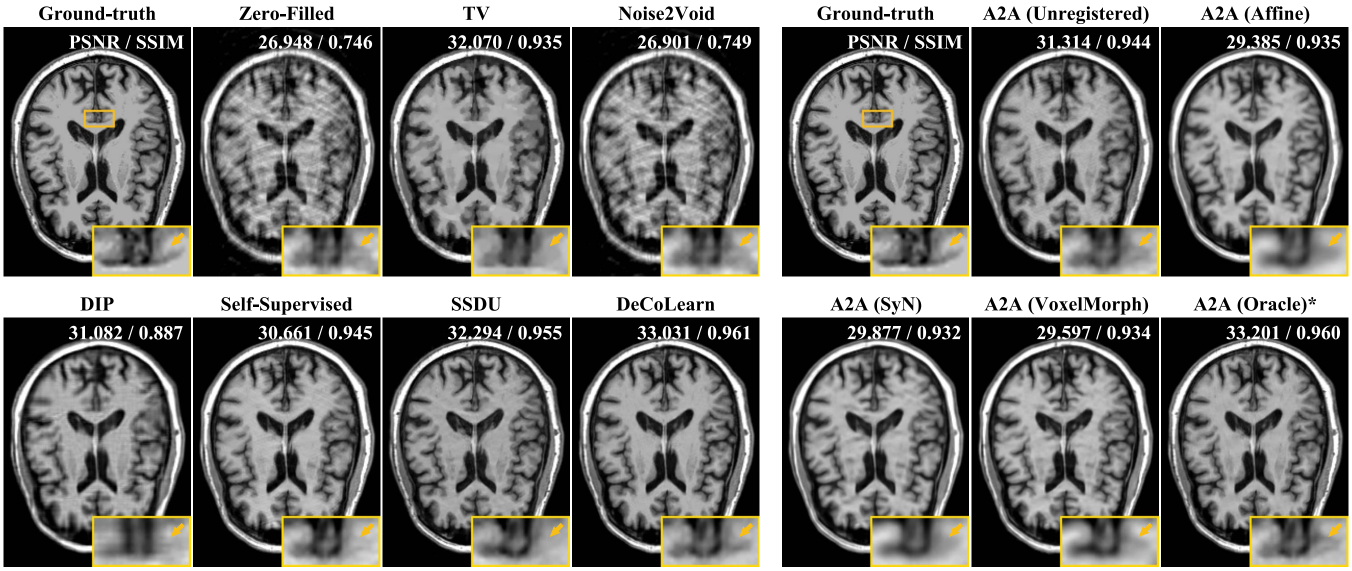
and Hochberg [78]. The statistical analysis was performed using GraphPad Prism 9 (Version 9.3.1 for macOS, GraphPad Software, San Diego, CA, USA). Statistical significance was defined as $P < 0.05$. Our evaluations on experimental data are qualitative due to the ground-truth being unavailable.

3) Implementation: We have experimented with several choices for the loss functions in eq. (10). The best empirical results were obtained when using the ℓ_1 loss for the experimentally collected measurements, and the Huber function (or smooth-ℓ_1 loss [79]) for the simulated measurements. We set the learning rates of Adam_{reg} and Adam_{rec} to 0.0005, and the mini-batch sizes to 4. We performed all our experiments on a machine equipped with an Intel Xeon Gold 6130 Processor and an NVIDIA GeForce RTX 2080 Ti GPU.

B. Simulated Measurements and Deformations

1) Dataset: We used the T1-weighted MR brain acquisitions of 60 subjects obtained from the open dataset OASIS-3 [80] as the raw ground-truth for simulating measurements. The raw ground-truth images are magnitude images. These 60 subjects were split into 48, 6, and 6 for training, validation, and testing, respectively. For each subject, we extracted the middle 50 to 70 (depending on the shape of the brain) out of the 256 slices on the transverse plane, containing the most relevant regions of the brain. Each slice corresponds to $\boldsymbol{x}^r$ in (8a). We synthesized motion fields ($\boldsymbol{\phi}^{r\rightarrow m}$ in (8b)) based on the method in [75] and used them to deform the ground-truth images, where the resulting images correspond to $\boldsymbol{x}^m$ in (8b). Three predefined parameters of the generation were the number of points randomly selected in the zero vector field $p = 2000$, the range of random values assigned to those points $\delta = [-10, 10]$, and the standard deviations of the smoothing Gaussian kernel for the vector field $\sigma \in \{10, 18, 24\}$. Thus, σ is inversely related to the strength of deformation in the image. Fig. 3 shows visual examples of the deformed images generated by synthetic registration fields with different values of σ. In order to obtain corrupted measurement pairs, we simulated a single-coil MRI setting with a Cartesian sampling pattern that sub-samples and fully-samples along ky and kx dimension in the k-space, respectively. We set the sampling rate to 25% and 33% (corresponding to 4× and 3× acceleration) of the full sampling rate for the complete k-space data and added measurement noise corresponding to an input SNR of 40dB.

2) Results: Table I summarizes quantitative results of all the evaluated methods. Note that the improvement of *SSDU*



(a) (b)

Fig. 4. Quantitative evaluation of DeCoLearn on simulated MRI measurements with *in-vivo* deformations and 33% sampling rate: *(a)* comparison against other methods and *(b)* results of an ablation study showing the influence of registration. The top-right corner of each image provides the PSNR and SSIM values with respect to the ground-truth. Yellow arrows in the highlight brain regions that were well reconstructed using DeCoLearn. Note that *A2A (Oracle)* is an idealized algorithm that requires perfectly registered measurements that are unavailable in practice. This figure highlights that DeCoLearn can achieve excellent quantitative and visual performance.

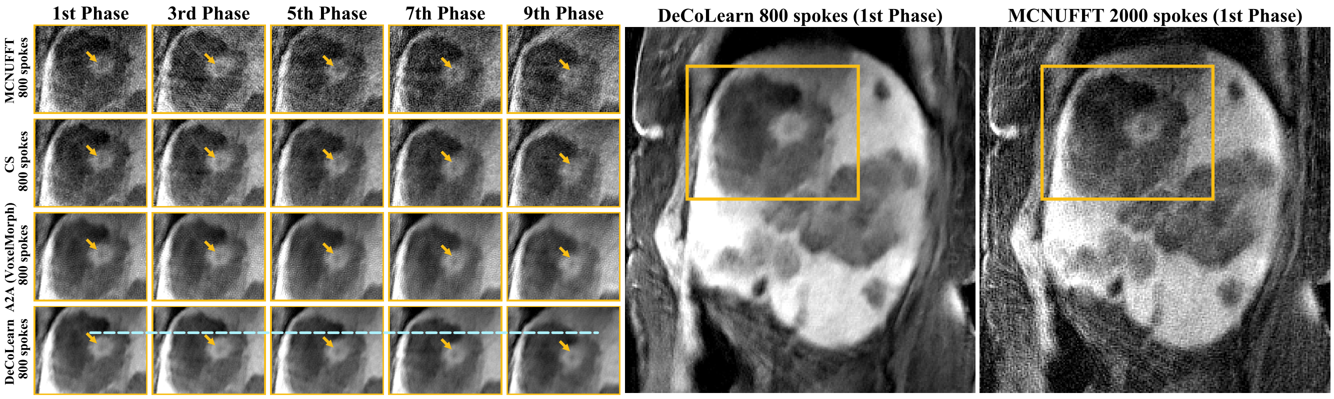


Fig. 5. Illustration of *in-vivo* respiratory deformation and several 3D reconstruction results from experimentally collected measurements corresponding to 800 spokes (about 2 minutes scan). The blue line provides a horizontal position reference of the tumor in the reconstruction result of DeCoLearn, demonstrating nonrigid deformations between images across different respiratory phases. Yellow arrows indicate areas that were well preserved by DeCoLearn. Note how DeCoLearn reconstructs higher quality images compared to both *CS* and *A2A (VoxelMorph)*.

over *Self-Supervised* is due to the *deep unrolling* architecture, that, in principle, can also be adopted in DeCoLearn to further improve its performance. Table I shows that DeCoLearn achieves the highest PSNR and SSIM values compared to other methods over all considered configurations of subsampling and deformation strengths. Statistical analysis of PSNR and SSIM values in Table I also highlights that DeCoLearn can achieve statistical significant results compared to the baseline image reconstruction methods. Table II shows the quantitative results of the ablation study evaluating the influence of the deep registration module. The results suggest that pre-registering images before training leads to sub-optimal performance, while DeCoLearn nearly matches the performance of the idealized *A2A (Oracle)* that uses the ground-truth deformations.

C. Simulated Measurements and Real Deformations

1) Dataset: We consider a data acquisition scheme that is similar to that described in Sec. IV-B, but differs in the approach to deform the ground-truth. Specifically, we used the second MR acquisitions of the 60 subjects from the OASIS-3 [80] dataset as the deformed images. The intervals between the two MR sessions of each subject range from one to ten years. Note that the deformations occurring in two different *in vivo* MR images of the same subject are due to normal aging and the potential effects of the Alzheimer disease. Fig. 3 visually illustrates the corresponding deformation.

2) Results: Fig. 4a summarizes the results from all the evaluated methods on this dataset. One can observe a significant reduction in imaging artifacts due to *TV* compared to the *Zero-Filled* reconstruction. However, *TV* also leads to a loss of

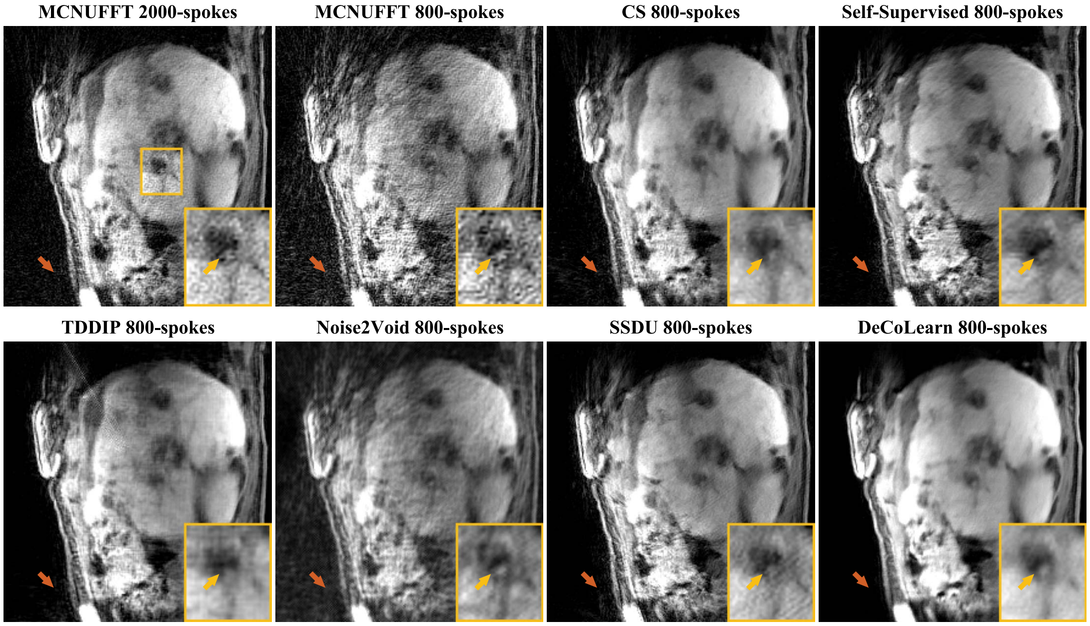


Fig. 6. Comparison of several reconstruction methods on experimentally collected data corresponding to 800 radial spokes (scans of about 2 minutes). N2V, SSDU, and Self-Supervised are all trained by using the available 800 spokes at each motion state. CS and TDDIP take advantage of the correlations in the respiratory motion dimension by imposing an additional regularizer and encoding the motion trajectory into input latent variables, respectively. DeCoLearn improves over A2A training by correcting for deformations in different motion states. The visually important differences are highlighted using arrows. Note how compared to other methods, DeCoLearn recovers sharper images (see yellow arrows in magnified regions) and reduces artifacts (see orange arrows in the background).

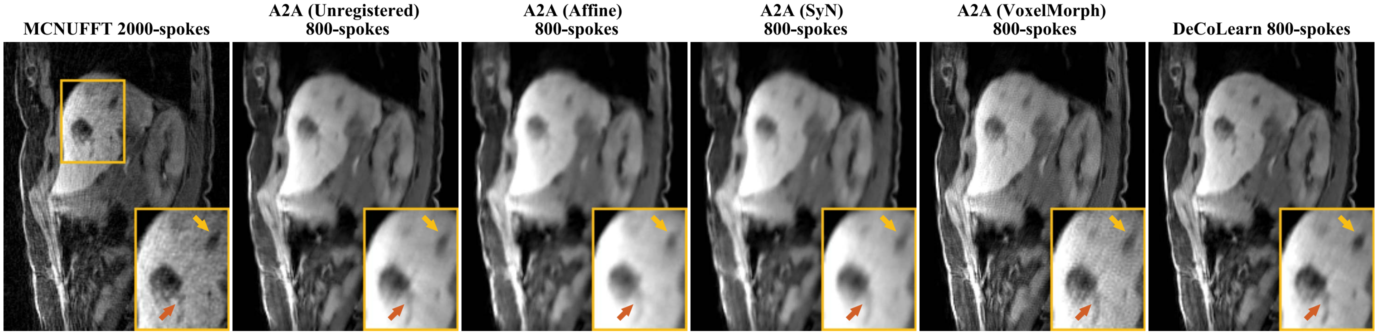


Fig. 7. Illustration of the results from the ablation study of DeCoLearn on experimentally-collected data corresponding to 800 radial spokes (scans of about 2 minutes). *A2A (Unregistered)* is directly trained on unregistered 3D measurement pairs, while *A2A (SyN)* and *A2A (VoxelMorph)* train CNNs on pre-registered but artifact-corrupted images. *MCNUFFT 2000-spokes* requires data corresponding to 2000 radial spokes (scans of about 5 minutes). The visual differences are highlighted using arrows in magnified regions. Note how DeCoLearn outperforms its ablated variants by jointly performing 3D image reconstruction and registration.

detail due to the well-known "staircase effect". While *N2V* can achieve good performance on removing unstructured artifacts, such as AWGN, it is suboptimal for the removal of structured MRI ghosting artifacts due to k-space undersampling. The yellow arrows in the magnified regions of Fig. 4a highlight brain tissue that was clearly reconstructed using only DeCoLearn.

Fig. 4b provides results from the ablation study. Pre-registration methods, such as *A2A (VoxelMorph)*, lead to a significant improvements over the registration-free methods by using pre-registered artifact-contaminated images, but they still suffer from smoothing in the region indicated by yellow arrows. DeCoLearn achieves better performance compared to all of these ablated methods in terms of sharpness, contrast, and artifact removal, due to its ability to correct for deformations during training. Note that although the measurements were simulated in this experiment for quantitative evaluation, the deformations in the data are *in vivo*.

D. Real Measurements and Real Deformations

1) Dataset: All acquisition processes were performed on a 3T PET/MRI scanner (Biograph mMR; Siemens Healthcare, Erlangen, Germany). We collected the data by using the CAPTURE method, a T1-weighted stack-of-stars 3D spoiled gradient-echo sequence with fat suppression that has consistently acquired projections for respiratory motion detection [76]. The acquisition parameters were as follows: TE/TR = 1.69ms/3.54ms, FOV = $360 \times 360 \times 288$ - $360 \times 360 \times 360$ mm^3, resolution = $1.125 \times 1.125 \times 6$ mm^3, partial Fourier factor = 6/8, number of radial spokes = 2000, slice

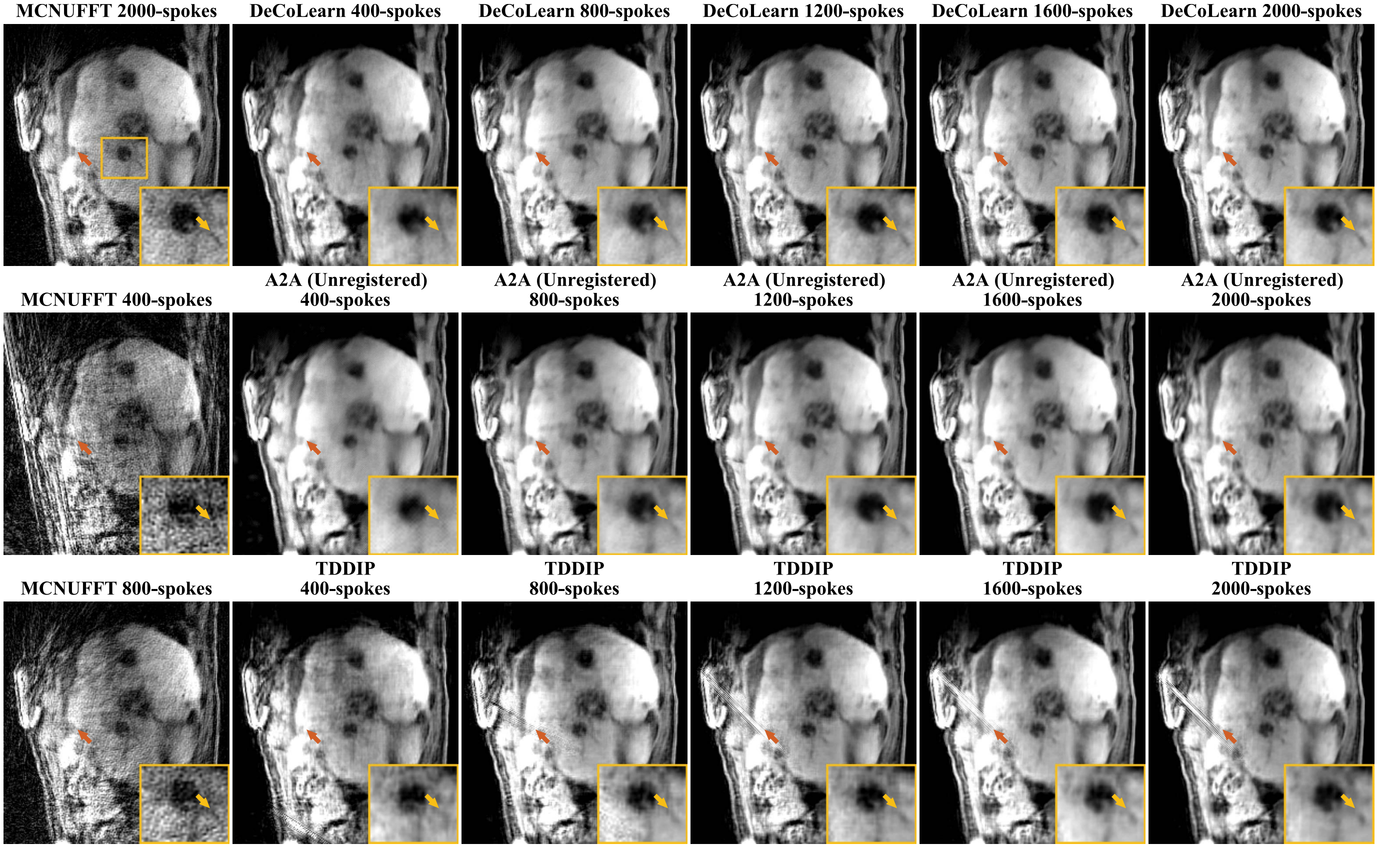

MCNUFFT 2000-spokes
DeCoLearn 400-spokes
DeCoLearn 800-spokes
DeCoLearn 1200-spokes
DeCoLearn 1600-spokes
DeCoLearn 2000-spokes
MCNUFFT 400-spokes
A2A (Unregistered) 400-spokes
A2A (Unregistered) 800-spokes
A2A (Unregistered) 1200-spokes
A2A (Unregistered) 1600-spokes
A2A (Unregistered) 2000-spokes
MCNUFFT 800-spokes
TDDIP 400-spokes
TDDIP 800-spokes
TDDIP 1200-spokes
TDDIP 1600-spokes
TDDIP 2000-spokes

Fig. 8. Illustration of reconstruction results of DeCoLearn, *A2A (Unregistered)*, and *TDDIP* from experimentally collected measurements using 400, 800, 1200, 1600, and 2000 spokes, corresponding to 1-, 2-, 3-, 4-, and 5- minute scans, respectively. *A2A (Unregistered)* trains CNNs on unregistered measurements. *TDDIP* is a variant of DIP that improves performance jointly reconstructing images of 10 respiratory phases. We highlighted visual differences by using arrows. Note how DeCoLearn reconstructs sharper edges (see liver tissues highlighted by yellow arrows in the magnified region) and better reduces artifacts (see image backgrounds highlighted by orange arrows). This figure shows that DeCoLearn can improve over these two methods at different acquisitions durations by integrating a deep image registration module.

resolution = 50%, slice per slab $N_z = \{96, 112, 120\}$ so as to cover the torso with an interpolated slice thickness of 3mm, total acquisition time was about 5 minutes (slightly longer for larger subjects). Note that the actual resolution in slice dimension is 6 mm, but being interpolated into 3 mm. We discarded the first ten spokes during reconstruction to ensure the acquired signal reached a steady state. Our free-breathing MRI data were subsequently binned into $N_p = 10$ respiratory phases, and thus each phase was reconstructed with $N_s = 199$ spokes. The dimension of raw measurement for each subject was $N_z \times N_c \times N_p \times N_s \times N_l$ with $N_c = \{5, 6\}$ being the number of coils and N_l being the length of radial spokes. The coil sensitivity maps were estimated from the central radial k-space spokes of each slice and were assumed to be known during experiments. Apodization was applied by using a Hamming window that covered the central k-space in order to avoid Gibbs ringing. We used inverse Multi-Coil Non-Uniform Fast Fourier Transform (MCNUFFT) [81] to map those measurements from k-space to the image domain, yielding 4D images $N_x \times N_y \times N_p \times N_z$ for each subject where $N_x \times N_y$ is the image domain matrix size.

Upon the approval of our Institutional Review Board, multichannel liver data from ten healthy volunteers and six cancer patients were used in this paper, where eight healthy subjects were used for training, one healthy subject for validation, and the rest for testing. Raw measurements of each subject were first reformatted into N_z measurements, yielding $8N_z$ samples for training and N_z for validation. Each of the reformatted measurements is in three-dimensions with two spatial dimensions and one dimension corresponding to the respiratory phase. We then trained DeCoLearn on measurement pairs such that each pair contained the five odd respiratory phases and the five even respiratory phases of the same training sample. Fig. 5 shows examples of MCNUFFT images obtained from a training sample, demonstrating that DeCoLearn was trained on unregistered measurement pairs corresponding to images with nonrigid respiratory deformations. We used MCNUFFT images from the full acquisition duration (5 minutes) as the reference for qualitative evaluations. We conducted the experiments for various acquisition durations of 1, 2, 3, 4, and 5 minutes, corresponding to 400, 800, 1200, 1600, and 2000 radial spokes in k-space, respectively. The golden-angle acquisition scheme ensures approximately *uniform* coverage of k-space for any arbitrary number of consecutive spokes [82].

The original implementation[4] of SSDU [15], [46] is based on the *fast Fourier transform (FFT)*, which is not suitable

---

[4] Publicly available at github.com/byaman14/SSDU.

to the non-uniform sampling pattern used in our data. Therefore, we re-implemented SSDU by using a publicly available non-uniform FFT operator [81] and the *unrolled regularization by denoising* architecture [45]. Though *Self-Supervised* relies on the same 3D network as DeCoLearn, due to memory constraints, *SSDU* is implemented as a 2D architecture that processes each individual phase separately. Note that the original implementation of *SSDU* is also based on a 2D architecture.

2) Results: Fig. 6 shows reconstruction results of all the methods on 800 radial spokes (corresponding to about 2 minute acquisitions). The *MCNUFFT* image suffers from strong streaking artifacts. Note how even *MCNUFFT 2000-spokes*, which corresponds to about 5 minute acquisitions, leads to imaging artifacts. All other methods yield significant improvements over *MCNUFFT*. While the result of *CS* (which is similar to the well-known XD-GRASP method) shows a considerable reduction in the streaking artifacts, it also contains a noticeable amount of detail loss. *N2V* reduces the noise-like artifacts, but still preserves the structured streaking artifacts. The results of *SSDU* and *Self-Supervised* show the benefit of N2N-type of training over that of N2V for image reconstruction. Overall, DeCoLearn achieves the best qualitative performance. As highlighted in Fig. 6 using arrows, DeCoLearn reconstructs sharper edges (see yellow arrows) and reduces background imaging artifacts (see orange arrows).

Fig. 7 illustrates the results of the ablation experiments on the real data with 800 radial spokes. *A2A (Unregistered)* leads to a reasonable result even without registration in training, but it also contains a noticeable amount of blur, especially along the edges. *A2A (Affine)* and *A2A (SyN)* also suffer from smoothing in the region of interest even with the registration algorithms integrated to pre-align the samples. Note the reduction in blur in *A2A (VoxelMorph)* relative to the registration-free methods. However, a closer inspection indicates that the result of *A2A (VoxelMorph)* still suffers from artifacts, such as the noise-like artifacts around the spot highlighted by yellow and orange arrows. Fig. 7 depicts that DeCoLearn leads to improvements over several baseline methods, especially compared with *MCNUFFT 2000 spokes* with a longer acquisitions time (5 minutes). Fig. 5 also provides visual comparisons between DeCoLearn, *CS* and *A2A (VoxelMorph)*. Fig. 5 shows that DeCoLearn performs better across different respiratory phases, especially considering its ability to remove artifacts around the spot highlighted by yellow arrows. Note that both the measurements and the deformations in these results are from experimentally collected data, demonstrating the applicability of DeCoLearn in motion-resolved MRI.

Fig. 8 illustrates comparisons between *A2A (Unregistered)*, *TDDIP* and DeCoLearn for various acquisition durations. We annotated visual differences using yellow and red arrows. While *A2A (Unregistered)* trains CNNs directly on unregistered measurement pairs, DeCoLearn reconstructs sharper boundaries highlighted by yellow arrows due to its ability to take into account the deformation field during training. These results indicates the excellent performance of DeCoLearn across different acquisition durations.

V. Discussion and Conclusion

A. Benefits of DeCoLearn

DeCoLearn enables learning using information from multiple measurements of the same object undergoing nonrigid deformation. Unlike N2N/A2A, DeCoLearn relaxes the requirement on having registered measurements, making it more applicable in practice. DeCoLearn is fully complementary to existing self-supervised methods that use a single measurement, such as SSDU [15], [46] and N2V [14]. One can simply integrate DeCoLearn with these self-supervised schemes by imposing an additional self-supervision term. Note also that DeCoLearn is compatible with any deep unrolling architecture.

B. Limitations and Possible Extensions

1) Extension to Contrast-Variant Measurements: The current implementation of DeCoLearn can only compensate image deformations over different acquisitions of the same object. In some dynamic imaging scenarios, such as the dynamic contrast enhanced (DCE) imaging [83], different measurements acquired from the same object might also correspond to distinct image contrasts. DeCoLearn is not yet suitable for such imaging problems. Extension of DeCoLearn to this scenario would be an interesting direction of future research.

2) Extension to Sequential Image Reconstruction: The reconstruction of a sequence of images from the measurements of a dynamic object has many applications in medical imaging (e.g., cine dynamic imaging). The key concept behind dynamic imaging is to leverage the redundancies in the data across the motion dimension (see our discussion of *MoCo* reconstruction). Our experimental validation on free-breathing MRI has shown that DeCoLearn can be used to learn the redundancies over the respiratory dimension. However, DeCoLearn does not explicitly use properties specific to the motion dimension. Future work can address this by extending DeCoLearn to include an explicit motion regularization.

3) Availability of Training Data: DeCoLearn requires training datasets consisting of the unregistered measurements acquired from the same object. While DeCoLearn can relax the requirement on deformation-free measurements, there exist applications for which multiple measurements of the same object are not available. The availability of the training data can be a factor that could thus limit the usefulness of DeCoLearn for some applications. It is worth mentioning that the availability of multiple views of the same object also comes with the advantage that it can boost the imaging quality, as can be seen from the comparisons between DeCoLearn and N2V.

C. Conclusion

We proposed a new method for addressing an important issue in the context of training of deep neural networks for medical image reconstruction. Our proposed DeCoLearn method extends the influential Noise2Noise approach by working directly in the measurement domain and compensating for object motion in the data. We validated our method using simulated and experimentally collected MRI data. Our results demonstrated that DeCoLearn significantly improves image quality compared to several baseline methods. Though our experiments focused on MRI, the DeCoLearn method has the potential to be adopted in other imaging modalities as well, such as computerized tomography [27] and optical diffraction tomography [84]. In such imaging scenarios, it is often impossible to obtain fully-sampled measurements, but only several distinct views of the object where it is possible that these views are not registered onto each other.

APPENDIX

D. Network Architectures

We customized the residual CNN [85] into $\mathsf{h}_{\boldsymbol{\theta}}$. $\mathsf{h}_{\boldsymbol{\theta}}$ consists of three components. The first component is a convolution layer (Conv) that takes corrupted images as input. The second component is a sequence of residual blocks. Each block alternates between a Conv followed by a rectified linear unit (ReLU), a normal Conv, and an adding residual connection. The third component is a Conv followed by an adding skip connection. It processes the feature maps generated by the first component and produces a output with the same dimension as that of the network input. Kernel sizes of all Convs are set to 3, strides to 1, and filters to 64.

The architecture of $\mathsf{g}_{\boldsymbol{\varphi}}$ is similar to VoxelMorph [29]. $\mathsf{g}_{\boldsymbol{\varphi}}$ consists of five encoder blocks, four decoder blocks with skip connections, and an output block. Each encoder block sequentially has a Conv and a Parametric ReLu (PReLU). The Conv has a kernel size of 4 and a strides of 2 to reduce the feature maps by half in the spatial dimension. In the decoder pathway, the intermediate feature maps were first up-sampled to double the size through a bilinear interpolation kernel. They were then concatenated with the feature maps originated from the encoder block at the same level via the skip connection. The concatenated feature maps were used as inputs to a decoder block, which consists of a Conv with a kernel size of 3 and a stride of 1, and a PReLU. The output block has one normal Conv to generate a registration field. Filters of all Convs are set to 32.

E. Deep Registration Component of DeCoLearn

In this section, we evaluate the deep registration component of DeCoLearn, which was trained directly on subsampled and noisy data. We consider simulated measurements described in Sec. IV-B and Sec. IV-C.

1) Baseline Methods: We used several image registration methods as references that can be divided into three categories.

- **Reconstruction-free methods:** Both methods apply registration algorithms directly on zero-filled (ZF) images.
  1) *Affine (ZF)*: Uses the Affine algorithm implemented in *advanced normalization tools (ANTs)* [77].
  2) *SyN (ZF)*: Similar to *Affine (ZF)*, but uses the *SyN* [74] algorithm instead.
  3) *VoxelMorph (ZF)*: Trains a deep *VoxelMorph* [29] model.
- **Pre-reconstruction methods:**
  4) *SyN (TV)*: Uses SyN on images reconstructed using *TV*.
- **Oracle-reconstruction method:**
  5) *VoxelMorph (Oracle)*: Similar to *VoxelMorph (ZF)*, but uses clean images for training.

TABLE V

DICE SCORE OBTAINED OVER ON THE TESTING SET. NOTE THAT *VOXELMORPH (ORACLE)* USES GROUND-TRUTH IMAGES AS INPUTS, WHILE THE OTHERS RELY ON CORRUPTED COUNTERPARTS. THE HIGHEST DICE SCORES AMONG METHODS USING CORRUPTED IMAGES ARE IN BOLD. THIS TABLE ILLUSTRATES THAT DECOLEARN IS APPROPRIATE TO TRAIN AN END-TO-END REGISTRATION NETWORK ON CORRUPTED IMAGE PAIRS

Experiment of Deformable Corrupted Image Registration

Schemes	Dice Score							
Deformation	Synthetic with $\sigma =$ 10		18		24		Longitudinal	
Acceleration rate	x3	x4	x3	x4	x3	x4	x3	x4
Unregistered	0.681		0.786		0.837		0.739	
VoxelMorph (ZF)	0.848	0.809	0.903	0.873	0.906	0.891	0.831	0.814
Affine(ZF)	0.684	0.683	0.802	0.799	0.860	0.855	0.811	0.803
SyN (ZF)	0.866	0.815	0.897	0.851	0.850	0.852	0.839	0.801
SyN (TV)	**0.899**	0.859	0.939	0.898	0.943	0.902	0.860	0.835
DoDIR	0.895	**0.863**	**0.941**	**0.911**	**0.945**	**0.925**	**0.861**	**0.844**
VoxelMorph (Oracle)*	0.939		0.973		0.983		0.871	

*: unavailable without ground-truth images; *ZF*: zero-filled; *TV*: total variation.

2) Evaluation Metrics: We implemented Dice Score [86] as the evaluation metric that quantifies the overlap of anatomical segmentation maps generated through Freesurfer [87] between the registered images and their registration targets.

3) Results: The quantitative results over the test set are summarized in Table V. This table illustrates that DeCoLearn can achieve a significant quality improvement when performing deformable corrupted image registration by taking advantage of the concurrent deep image reconstruction.

REFERENCES

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

[2] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.

[3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[4] Y. Hu, S. G. Lingala, and M. Jacob, "A fast Majorize–Minimize algorithm for the recovery of sparse and low-rank matrices," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 742–753, Feb. 2012.

[5] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[6] F. Knoll *et al.*, "Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 128–140, Jan. 2020.

[7] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," *IEEE Trans. Signal Process. Mag.*, vol. 35, no. 1, pp. 20–36, Jan. 2018.

[8] M. T. McCann, K. H. Jin, and M. Unser, “Convolutional neural networks for inverse problems in imaging: A review,” *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, Nov. 2017.

[9] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, “Deep learning techniques for inverse problems in imaging,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.

[10] G. Wang, J. C. Ye, and B. D. Man, “Deep learning for tomographic image reconstruction,” *Nature Mach. Intell.*, vol. 2, no. 12, pp. 737–748, Dec. 2020.

[11] M. Akçakaya, B. Yaman, H. Chung, and J. Chul Ye, “Unsupervised deep learning methods for biological image reconstruction and enhancement,” 2021, *arXiv:2105.08040*.

[12] J. Lehtinen *et al.*, “Noise2Noise: Learning image restoration without clean data,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965–2974.

[13] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, “RARE: Image reconstruction using deep priors learned without groundtruth,” *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1088–1099, Oct. 2020.

[14] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2 Void–Learning denoising from single noisy images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2129–2137.

[15] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Ugurbil, and M. Akcakaya, “Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data,” in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 921–925.

[16] S. Laine, T. Karras, J. Lehtinen, and T. Aila, “High-quality self-supervised deep image denoising,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 6970–6980.

[17] C. Eldeniz *et al.*, “Phase2Phase: Respiratory motion-resolved reconstruction of free-breathing magnetic resonance imaging using deep learning without a ground truth for improved liver imaging,” *Investigative Radiol.*, vol. 56, no. 12, pp. 809–819, 2021.

[18] M. Torop *et al.*, “Deep learning using a biophysical model for robust and accelerated reconstruction of quantitative, artifact-free and denoised images,” *Magn. Reson. Med.*, vol. 84, no. 6, pp. 2932–2942, Dec. 2020.

[19] T. Ehret, A. Davy, J. Morel, G. Facciolo, and P. Arias, “Model-blind video denoising via frame-to-frame training,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 11369–11378.

[20] S. Yu, B. Park, J. Park, and J. Jeong, “Joint learning of blind video denoising and optical flow estimation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 500–501.

[21] Z. Jiang *et al.*, “Weakly supervised deep learning-based optical coherence tomography angiography,” *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 688–698, Feb. 2021.

[22] T.-O. Buchholz, M. Jordan, G. Pigino, and F. Jug, “Cryo-CARE: Content-aware image restoration for cryo-transmission electron microscopy data,” in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 502–506.

[23] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, “Noise2inverse: Self-supervised deep convolutional denoising for tomography,” *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1320–1335, 2020.

[24] J. Xu and E. Adalsteinsson, “Deformed2Self: Self-supervised denoising for dynamic medical imaging,” 2021, *arXiv:2106.12175*.

[25] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “Deep learning in medical image registration: A review,” *Phys. Med. Biol.*, vol. 65, no. 20, Oct. 2020, Art. no. 20TR01.

[26] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, “A deep learning framework for unsupervised affine and deformable image registration,” *Med. Image Anal.*, vol. 52, pp. 128–143, Feb. 2018.

[27] Y. Lei *et al.*, “4D-CT deformable image registration using an unsupervised deep convolutional neural network,” in *Artificial Intelligence in Radiation Therapy*. Cham, Switzerland: Springer, 2019, pp. 26–33.

[28] I. Yoo, D. G. Hildebrand, W. F. Tobin, W.-C. A. Lee, and W.-K. Jeong, “ssEMNet: Serial-section electron microscopy image registration using a spatial transformer network with learned features,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 249–257.

[29] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A learning framework for deformable medical image registration,” *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[30] W. Gan, Y. Sun, C. Eldeniz, J. Liu, H. An, and U. S. Kamilov, “Deep image reconstruction using unregistered measurements without groundtruth,” 2020, *arXiv:2009.13986*.

[31] H. K. Aggarwal, M. P. Mani, and M. Jacob, “Modl: Model-based deep learning architecture for inverse problems,” *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.

[32] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic MR image reconstruction,” *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[33] Y. Yang, H. Li, J. Sun, and Z. Xu, “Deep ADMM-Net for compressive sensing MRI,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, p. 9.

[34] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241.

[35] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[36] Y. Romano, M. Elad, and P. Milanfar, “The little engine that could: Regularization by denoising (RED),” *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, Oct. 2017.

[37] S. Sreehari *et al.*, “Plug-and-play priors for bright field electron tomography and sparse interpolation,” *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[38] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

[39] Y. Sun, S. Xu, Y. Li, L. Tian, B. Wohlberg, and U. S. Kamilov, “Regularized Fourier ptychography using an online Plug-and-play algorithm,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7665–7669.

[40] K. Zhang, W. Zuo, and L. Zhang, “Deep plug-and-play super-resolution for arbitrary blur kernels,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1671–1681.

[41] R. Ahmad *et al.*, “Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery,” *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 105–116, Jan. 2020.

[42] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.

[43] J. Zhang and B. Ghanem, “ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1828–1837.

[44] Y. Chen, W. Yu, and T. Pock, “On learning optimized reaction diffusion processes for effective image restoration,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5261–5269.

[45] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov, “SGD-Net: Efficient model-based deep learning with theoretical guarantees,” *IEEE Trans. Comput. Imag.*, vol. 7, pp. 598–610, 2021.

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

[47] B. Yaman, S. A. H. Hosseini, and M. Akçakaya, “Zero-shot self-supervised learning for MRI reconstruction,” 2021, *arXiv:2102.07737*.

[48] J. Batson and L. Royer, “Noise2Self: Blind denoising by self-supervision,” in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 524–533.

[49] A. Krull, T. Vičar, M. Prakash, M. Lalit, and F. Jug, “Probabilistic Noise2Void: Unsupervised content-aware denoising,” *Frontiers Comput. Sci.*, vol. 2, p. 5, Feb. 2020.

[50] S. Soltanayev and S. Y. Chun, “Training deep learning based denoisers without ground truth data,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, p. 11.

[51] Y. Quan, M. Chen, T. Pang, and H. Ji, “Self2Self with dropout: Learning self-supervised denoising from single image,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1887–1895.

[52] V. Lempitsky, A. Vedaldi, and D. Ulyanov, “Deep image prior,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[53] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov, “Image restoration using total variation regularized deep image prior,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7715–7719.

[54] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, “Time-dependent deep image prior for dynamic MRI,” *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3337–3348, Dec. 2021.

[55] G. Mataev, P. Milanfar, and M. Elad, “DeepRED: Deep image prior powered by RED,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Oct. 2019, pp. 1–10.

[56] X. Yang, P. Ghafourian, P. Sharma, K. Salman, D. Martin, and B. Fei, "Nonrigid registration and classification of the kidneys in 3D dynamic contrast enhanced (DCE) MR images," *Proc. SPIE*, vol. 8314, Feb. 2012, Art. no. 83140B.

[57] X. Han *et al.*, "Atlas-based auto-segmentation of head and neck CT images," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2008, pp. 434–441.

[58] Y. Fu, S. Liu, H. H. Li, and D. Yang, "Automatic and hierarchical segmentation of the human skeleton in CT images," *Phys. Med. Biol.*, vol. 62, no. 7, p. 2812, 2017.

[59] R. Bajcsy and S. Kovačič, "Multiresolution elastic matching," *Comput. Vis., Graph., Image Process.*, vol. 46, no. 1, pp. 1–21, 1989.

[60] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, Dec. 2015, pp. 2017–2025.

[61] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, 2016.

[62] L. Feng *et al.*, "Highly accelerated real-time cardiac cine MRI using $k-t$ SPARSE-SENSE," *Magn. Reson. Med.*, vol. 70, no. 1, pp. 64–74, Jul. 2013.

[63] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magn. Reson. Med.*, vol. 64, no. 3, pp. 767–776, 2010.

[64] M. Usman *et al.*, "Motion corrected compressed sensing for free-breathing dynamic cardiac MRI: Motion corrected compressed sensing," *Magn. Reson. Med.*, vol. 70, no. 2, pp. 504–516, Aug. 2013.

[65] G. Cruz, D. Atkinson, M. Henningsson, R. M. Botnar, and C. Prieto, "Highly efficient nonrigid motion-corrected 3D whole-heart coronary vessel wall imaging," *Magn. Reson. Med.*, vol. 77, no. 5, pp. 1894–1908, May 2017.

[66] A. Bustin *et al.*, "3D whole-heart isotropic sub-millimeter resolution coronary magnetic resonance angiography with non-rigid motion-compensated PROST," *J. Cardiovascular Magn. Reson.*, vol. 22, no. 1, p. 24, Dec. 2020.

[67] M. Blume, A. Martinez-Moller, A. Keil, N. Navab, and M. Rafecas, "Joint reconstruction of image and motion in gated positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1892–1906, Nov. 2010.

[68] F. Odille *et al.*, "Joint reconstruction of multiple images and motion in MRI: Application to free-breathing myocardial T2 quantification," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 197–207, Jan. 2016.

[69] V. Corona *et al.*, "Multi-tasking to correct: Motion-compensated MRI via joint reconstruction and registration," in *Scale Space and Variational Methods in Computer Vision*, vol. 11603, J. Lellmann, M. Burger, and J. Modersitzki, Eds. Cham, Switzerland: Springer, 2019, pp. 263–274.

[70] C. Munoz, H. Qi, G. Cruz, T. Küstner, R. M. Botnar, and C. Prieto, "Self-supervised learning-based diffeomorphic non-rigid motion estimation for fast motion-compensated coronary MR angiography," *Magn. Reson. Imag.*, vol. 85, pp. 10–18, Jan. 2022.

[71] H. Qi *et al.*, "End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary MRA," *Magn. Reson. Med.*, vol. 86, no. 4, pp. 1983–1996, Oct. 2021.

[72] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[73] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[74] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008.

[75] H. Sokooti, B. De Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. Med. Image Comput. Comput.-Assisted Intervent*, 2017, pp. 232–239.

[76] C. Eldeniz *et al.*, "Consistently-acquired projections for tuned and robust estimation–A self-navigated respiratory motion correction approach," *Invest. Radiol.*, vol. 53, no. 5, p. 293, 2018.

[77] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ANTs)," *Insight J.*, vol. 2, pp. 1–35, Jun. 2009.

[78] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. B, Methodol.*, vol. 57, no. 1, pp. 289–300, Aug. 1995.

[79] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[80] P. J. LaMontagne *et al.*, "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *MedRxiv*, Dec. 2019. [Online]. Available: https://www.medrxiv.org/content/10.1101/2019.12.13.19014902v1.article-metrics

[81] M. J. Muckley, R. Stern, T. Murrell, and F. Knoll, "TorchKbNufft: A high-level, hardware-agnostic non-uniform fast Fourier transform," in *Proc. ISMRM Workshop Data Sampling Image Reconstruction*, 2020, p. 1.

[82] L. Feng *et al.*, "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn. Reson. Med.*, vol. 72, no. 3, pp. 707–717, 2014.

[83] Y. Gordon *et al.*, "Dynamic contrast-enhanced magnetic resonance imaging: Fundamentals and application to the evaluation of the peripheral perfusion," *Cardiovascular Diagnosis Therapy*, vol. 4, no. 2, p. 147, 2014.

[84] U. S. Kamilov *et al.*, "Optical tomographic image reconstruction based on beam propagation and sparse regularization," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 59–70, Mar. 2016.

[85] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[86] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[87] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.