Recurrent inference machine for medical image registration

Zhang, Yi; Zhao, Yidong; Xue, Hui; Kellman, Peter; Klein, Stefan; Tao, Qian

DOI
10.1016/j.media.2025.103748

Publication date
2025

Document Version
Final published version

Published in
Medical Image Analysis

Citation (APA)
Zhang, Y., Zhao, Y., Xue, H., Kellman, P., Klein, S., & Tao, Q. (2025). Recurrent inference machine for medical image registration. *Medical Image Analysis*, *106*, Article 103748. https://doi.org/10.1016/j.media.2025.103748

Medical Image Analysis 106 (2025) 103748

Contents lists available at ScienceDirect

Medical Image Analysis

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

Recurrent inference machine for medical image registration

Yi Zhang [a], Yidong Zhao [a], Hui Xue [b], Peter Kellman [c], Stefan Klein [d], Qian Tao [a],*

[a] *Delft University of Technology, Department of Imaging Physics, Delft, The Netherlands*
[b] *Microsoft Research, Health Futures, Redmond, WA, USA*
[c] *National Institutes of Health, National Heart, Lung, and Blood Institute, Bethesda, MD, USA*
[d] *Erasmus University Medical Center, Department of Radiology and Nuclear Medicine, Rotterdam, The Netherlands*

ARTICLE INFO

Keywords:
Image registration
Meta learning
Recurrent inference machine

ABSTRACT

Image registration is essential for medical image applications where alignment of voxels across multiple images is needed for qualitative or quantitative analysis. With recent advances in deep neural networks and parallel computing, deep learning-based medical image registration methods become competitive with their flexible modeling and fast inference capabilities. However, compared to traditional optimization-based registration methods, the speed advantage may come at the cost of registration performance at inference time. Besides, deep neural networks ideally demand large training datasets while optimization-based methods are training-free. To improve registration accuracy and data efficiency, we propose a novel image registration method, termed Recurrent Inference Image Registration (RIIR) network. RIIR is formulated as a meta-learning solver for the registration problem in an iterative manner. RIIR addresses the accuracy and data efficiency issues, by learning the update rule of optimization, with implicit regularization combined with explicit gradient input.

We extensively evaluated RIIR on brain MRI, lung CT, and quantitative cardiac MRI datasets, in terms of both registration accuracy and training data efficiency. Our experiments showed that RIIR outperformed a range of deep learning-based methods, even with only 5% of the training data, demonstrating high data efficiency. Key findings from our ablation studies highlighted the important added value of the hidden states introduced in the recurrent inference framework for meta-learning. Our proposed RIIR offers a highly data-efficient framework for deep learning-based medical image registration.

1. Introduction and related works

Medical image registration, the process of establishing anatomical correspondences between two or more medical images, finds wide applications in medical imaging research, including imaging feature fusion (Haskins et al., 2020; Oliveira and Tavares, 2014), treatment planning (Staring et al., 2009; King et al., 2010; Byrne et al., 2022), and longitudinal patient studies (Sotiras et al., 2013; Jin et al., 2021). Medical image registration is traditionally formulated as an optimization problem, which aims to solve a parameterized transformation in an iterative manner (Klein et al., 2007). Typically, the optimization objective consists of two parts: a similarity term that enforces the alignments between images, and a regularization term that imposes smoothness constraints. Due to the complexity of non-convex optimization, traditional methods often struggle with long run time, especially for large, high-resolution images. This hinders its practical use in clinical practice, e.g. surgery guidance (Sauer, 2006), where fast image registration is demanded (Avants et al., 2011; Balakrishnan et al., 2019).

With recent developments in machine learning, the data-driven deep-learning paradigm has gained popularity in medical image registration (Rueckert and Schnabel, 2019). Instead of iteratively updating the transformation parameters by a conventional optimization pipeline, deep learning-based methods make fast image-to-transformation predictions at inference time. Early works learned the transformation in a supervised manner (Miao et al., 2016; Yang et al., 2016), while unsupervised learning methods later became prevalent. They adopt similar loss functions as those in conventional methods but optimize them through amortized neural networks (Balakrishnan et al., 2019; De Vos et al., 2019). These works demonstrate the great potential of deep learning-based modes for medical image registration. Nonetheless, one-step inference of image transformation is in principle a difficult problem, compared to the iterative approach, especially when the deformation field is large. In practice, the one-step inference requires a relatively large amount of data to train the deep learning network for consistent prediction, and may still lead to unexpected transformations

* Corresponding author.
E-mail address: q.tao@tudelft.nl (Q. Tao).

https://doi.org/10.1016/j.media.2025.103748
Received 18 June 2024; Received in revised form 20 July 2025; Accepted 22 July 2025
Available online 5 August 2025

at inference time (Fechter and Baltas, 2020; Hering et al., 2019; Zhao et al., 2019).

In contrast to one-step inference, recent studies revisited iterative registration, using multi-step inference processes (Fechter and Baltas, 2020; Kanter and Lellmann, 2022; Qiu et al., 2022; Sandkühler et al., 2019; Zhao et al., 2019). Some of these iterative methods (Kanter and Lellmann, 2022; Qiu et al., 2022) fall within the realm of meta-learning. Instead of learning the optimized parameters, meta-learning focuses on learning the optimization process itself. The use of meta-learning in optimization, as explored by Andrychowicz et al. (2016) and Finn et al. (2017) for image classification tasks, has led to enhanced generalization and faster convergence. For medical imaging applications, a prominent example is the recurrent inference machine (RIM) by Putzky and Welling (2017), originally proposed to solve inverse problems with explicit forward physics models. RIM has demonstrated excellent performance in fast MRI reconstruction (Lønning et al., 2019) and MR relaxometry (Sabidussi et al., 2021).

In this study, we propose a novel meta-learning medical image registration method, named Recurrent Inference Image Registration (RIIR). RIIR is inspired by RIM, but significantly extends its concept to solve more generic optimization problems: different from inverse problems, medical image registration presents a high-dimensional optimization challenge with no closed-form forward model. Below we provide a detailed review to motivate our work.

1.1. Related works

In this section, we review deep learning-based medical image registration methods in more detail, categorizing them into one-step methods for direct image-to-transformation inference, and iterative methods for multi-step inference. Additionally, we provide a brief overview of meta-learning for medical imaging applications.

1.1.1. One-step deep learning-based registration

Early attempts of utilizing convolutional neural networks (CNNs) for medical image registration supported confined transformations, such as SVF-Net (Rohé et al., 2017), Quicksilver (Yang et al., 2017), and the work of Cao et al. (2017), which are mostly trained in a supervised manner. With the introduction of U-Net architecture (Ronneberger et al., 2015), which has excellent spatial expression capability thanks to its multi-resolution and skip connection, Balakrishnan et al. (2019), Dalca et al. (2019) and Hoopes et al. (2021) proposed unsupervised deformable registration frameworks. In the work of De Vos et al. (2019), a combination of affine and deformable transformations was further considered. More recent methods extended the framework by different neural network backbones such as transformers (Zhang et al., 2021) or implicit neural representations (Wolterink et al., 2022; van Harten et al., 2023).

1.1.2. Iterative deep learning-based registration

However, a one-step inference strategy may struggle when predicting large and complex transformations (Hering et al., 2019; Zhao et al., 2019). In contrast to one-step deep learning-based registration methods, recent work adopted iterative processes, reincarnating the conventional pipeline of optimization for medical image registration, either in terms of image resolution (Hering et al., 2019; Mok and Chung, 2020; Fechter and Baltas, 2020; Xu et al., 2021; Liu et al., 2021), multiple optimization steps (Zhao et al., 2019; Sandkühler et al., 2019; Falta et al., 2022; Kanter and Lellmann, 2022), or combined (Qiu et al., 2022). In Sandkühler et al. (2019), the use of RNN with gated recurrent unit (GRU) (Chung et al., 2014) was considered, where each step progressively updates the transformation by adding an independent parameterized transformation. Another multi-step method proposed in Zhao et al. (2019) uses recursive cascaded networks to generate a sequence of transformations, which is then composed to get the final transformation. However, the method requires independent modules for each step, which can be memory-inefficient. Hering et al. (2019) proposed a variational method on different levels of resolution, where the final transformation is the composition of the transformations from coarse- to fine-grained. Fechter and Baltas (2020) addresses the importance of data efficiency of deep learning-based models by evaluating the model performance when data availability is limited, and a large domain shift exists. Falta et al. (2022) proposed an iterative method named Learn-to-Optimize (L2O) to emulate the gradient-based optimization in lung CT registration. Unlike the fully unsupervised training scheme, the method utilizes a deep supervision strategy on the generated key points with a recurrent use of U-Net. Noticeably, the method uses additional input feature modalities including dynamically sampled coordinates and MIND features (Heinrich et al., 2012) to enhance the model. A more recent work proposed in Qiu et al. (2022), Gradient Descent Network for Image Registration (GraDIRN), integrates multi-step and multi-resolution for medical image registration. Specifically, the update rule follows the idea of conventional optimization by deriving the gradient of the similarity term *w.r.t.* the current transformation and using a CNN to estimate the gradient of the regularization term. Though the direct influence of the gradient term shows to be minor compared to the CNN output (Qiu et al., 2022), the method bridges gradient-based optimization and deep learning-based methods. The method proposed in Kanter and Lellmann (2022) used individual long short-term memory (LSTM) modules for implementing recurrent refinement of the transformation. However, the scope of the work is limited to affine transformation, which only serves as an initialization for the conventional medical image registration pipeline.

1.1.3. Meta-learning and recurrent inference machine

Meta-learning, also described as “learning to learn”, is a subfield of machine learning. In this approach, an outer algorithm updates an inner learning algorithm, enabling the model to adapt and optimize its learning strategy to achieve a broader objective. For example, in a meta-learning scenario, a model could be trained on a variety of tasks, such as different types of image recognition, with the goal of quickly adapting to unseen similar tasks, like recognizing new kinds of objects not included in the original training set, using a few training samples. Hospedales et al. (2021). An early approach in meta-learning is designing an architecture of networks that can update their parameters according to different tasks and data inputs (Schmidhuber, 1993). The work of Cotter and Conwell (1990) and Younger et al. (1999) further show that a fixed-weight RNN demonstrates flexibility in learning multiple tasks. More recently, methods learning an optimization process with RNNs were developed and studied in Andrychowicz et al. (2016), Chen et al. (2017) and Finn et al. (2017), demonstrating superior convergence speed and better generalization ability for unseen tasks.

In the spirit of meta-learning, RIM was developed by Putzky and Welling (2017) to solve inverse problems. RIM learns a single recurrent architecture that shares the parameters across all iterations, with internal states passing through iterations (Putzky and Welling, 2017). In the context of meta-learning, RIM distinguishes two tasks of different levels: the ‘inner task’, which focuses on solving a specific inverse problem (e.g., superresolution of an image), and the ‘outer task’, aimed at optimizing the optimization process itself. This setting enables RIM to efficiently learn and apply optimization strategies to complex problems. Therefore, RIM only has one neural network component which learns the outer task. RIM has shown robust and competitive performance across different application domains, from cosmology (Morningstar et al., 2019; Modi et al., 2021) to medical imaging (Karkalousos et al., 2022; Lønning et al., 2019; Putzky et al., 2019; Sabidussi et al., 2021, 2023). To the best of our knowledge, most applications of RIM aim to solve an inverse problem with a known differentiable forward model in closed form, such as Fourier transform with sensitivity map and sampling mask in MRI reconstruction (Lønning et al., 2019).

However, the definition of an explicit forward model does not exist for the medical image registration task. Although RIM does not require a forward model by its design, the absence of a concrete forward model makes the problem more complicated. In this case, our formulation is similar to a realization of iterative amortized inference (Marino et al., 2018). In Marino et al. (2018), a variational auto-encoder (VAE) framework is studied to learn the amortized optimization process given the input data and approximate posterior gradients where the likelihood under certain forward model could be absent. In this work, we sought to extend the framework of RIM, which demonstrated state-of-the-art performance in medical image reconstruction challenges (Muckley et al., 2021; Putzky et al., 2019; Zbontar et al., 2018), to the medical image registration problem which relaxes the need of gradient likelihood under specific forward model solely. The same formulation can be generalized to other high-dimensional optimization problems where explicit forward models are absent but differentiable evaluation metrics are available.

1.2. Contributions

The main contributions of our work are three-fold:

1. We propose a novel meta-learning framework, RIIR, for medical image registration. RIIR learns the optimization process, in the absence of explicit forward models. RIIR is flexible *w.r.t.* the input modality while demonstrating competitive accuracy in different medical image registration applications.

2. Unlike existing iterative deep learning-based methods, our method integrates the gradient information of input images into the prediction of dense incremental transformations. As such, RIIR largely simplifies the learning task compared to one-step inference, significantly enhancing the overall data efficiency, as demonstrated by our experiments.

3. Through in-depth ablation experiments, we not only showed the flexibility of our proposed method with varying input choices but also investigated how different architectural choices within the RIM framework impact its performance. In particular, we showed the added value of hidden states in solving complex optimization problems in the context of medical imaging, which was under-explored in existing literature.

2. Methods

2.1. Deformable image registration

Deformable image registration aims to align a moving image I_{mov} to a fixed image I_{fix} by determining a transformation $\boldsymbol{\phi}$ acting on the shared coordinates χ, such that the transformed image $I_{\text{mov}} \circ \boldsymbol{\phi}$ is similar enough to I_{fix}. The similarity is often evaluated by a scalar-valued metric. In deformable image registration, $\boldsymbol{\phi}$ is considered to be a relatively small displacement added to the original coordinate χ, expressed as $\boldsymbol{\phi} = \chi + u(\chi)$. Since the transformation $\boldsymbol{\phi}$ is calculated between the pair $(I_{\text{mov}}, I_{\text{fix}})$, the process is often referred to as *pairwise* registration (Balakrishnan et al., 2019). Finding such transformation $\boldsymbol{\phi}$ in pairwise registration can be viewed as the following optimization problem:

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\operatorname{argmin}}\ \mathcal{L}_{\text{sim}}\ (I_{\text{mov}} \circ \boldsymbol{\phi}, I_{\text{fix}}) + \lambda \mathcal{L}_{\text{reg}}\ (\boldsymbol{\phi}), \tag{1}$$

where $\mathcal{L}_{\text{sim}}$ is a similarity term between the deformed image $I_{\text{mov}} \circ \boldsymbol{\phi}$ and fixed image I_{fix}, $\mathcal{L}_{\text{reg}}$ is a regularization term constraining $\boldsymbol{\phi}$, and λ is a trade-off weight term.

2.2. Recurrent inference machine (RIM)

The idea of RIM originates from solving a closed-form inverse problem (Putzky and Welling, 2017):

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is a noisy measurement vector, $\boldsymbol{x} \in \mathbb{R}^d$ is the underlying noiseless signal, $A \in \mathbb{R}^{m \times d}$ is a measurement matrix, and $\boldsymbol{n}$ is a random noise vector. When $m \ll d$, the inverse problem is ill-posed. Thus, to constrain the solution space of x, a common practice is to solve a *maximum a posteriori* (MAP) problem:

$$\max_{\boldsymbol{x}} \log \mathcal{L}_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{x}) + \log p_{\text{prior}}(\boldsymbol{x}), \tag{3}$$

where $\mathcal{L}_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{x})$ is a likelihood term representing the noisy forward model, such as the Fourier transform with masks in MRI reconstruction (Putzky et al., 2019), and p_{prior} is the prior distribution of the underlying signal $\boldsymbol{x}$. A simple iterative scheme at step t for solving Eq. (3) is via gradient descent:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \gamma_t \nabla_{\boldsymbol{x}_t} \left(\log \mathcal{L}_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{x}) + \log p_{\text{prior}}(\boldsymbol{x}) \right), \tag{4}$$

where γ_t denotes a scalable step length and $\nabla_{\boldsymbol{x}_t}$ denotes the gradient *w.r.t.* $\boldsymbol{x}$, evaluated at $\boldsymbol{x}_t$. Then, in RIM implementation, Eq. (4) is represented as:

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + g_\theta \left(\nabla_{\boldsymbol{x}_t} \left(\log \mathcal{L}_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{x}) \right), \boldsymbol{x}_t \right), \tag{5}$$

where g_θ is a neural network parameterized by θ. In RIM, the prior distribution on the data prior (regularization) $p_{\text{prior}}(\boldsymbol{x})$ is implicitly integrated into the parameterized neural network g_θ which is trained with a weighted sum of the individual prediction losses between $\boldsymbol{x}$ and $\boldsymbol{x}_t$ (*e.g.*, the mean squared loss) at each time step t.

In the context of meta-learning, we regard the likelihood term $\mathcal{L}_{\text{likelihood}}$ guided by the forward model as the 'inner loss', denoted by $\mathcal{L}_{\text{inner}}$ as it is serving as the input of the neural network g_θ. For example, given the Gaussian assumption of the noise $\boldsymbol{n}$ with a known variance of σ^2 and linear forward model described in Eq. (2), the inner loss can be given as the logarithm of the maximum likelihood estimation (MLE) solution:

$$\mathcal{L}_{\text{inner}} = \frac{1}{\sigma^2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2. \tag{6}$$

In RIM, the gradient of $\mathcal{L}_{\text{inner}}$ is calculated explicitly with the (linear) forward operator A, which is free of the forward pass of a neural network. That means $\mathcal{L}_{\text{inner}}$ does not directly contribute to the update of the network parameters θ. The weighted loss for training the neural network g_θ for efficient solving the inverse problem can be regarded as the 'outer loss', denoted by $\mathcal{L}_{\text{outer}}$. In the form of the inverse problem shown in Eq. (2), the outer loss to update the network parameter θ across T time steps can be expressed as:

$$\mathcal{L}_{\text{outer}}(\theta) = \frac{1}{T} \sum_{i=1}^{T} \|\boldsymbol{x} - \boldsymbol{x}_t\|_2^2. \tag{7}$$

For clarity and consistency, these notations of $\mathcal{L}_{\text{inner}}$ and $\mathcal{L}_{\text{outer}}$ will be uniformly applied in the subsequent sections.

2.3. Recurrent Inference Image Registration Network (RIIR)

Inspired by the formulation of RIM and the optimization nature of medical image registration, we present a novel deep learning-based image registration framework, named the Recurrent Inference Image Registration Network (RIIR). The overview of our proposed framework can be found in Fig. 1.

Originally, RIM aimed to learn a recurrent solver for an inverse problem where the forward model from signal to measurement is known for inverse problems, such as quantitative mapping (Sabidussi et al., 2021) or MRI reconstruction (Lønning et al., 2019). Similarly to RIM in other medical image applications, the regularization is proposed

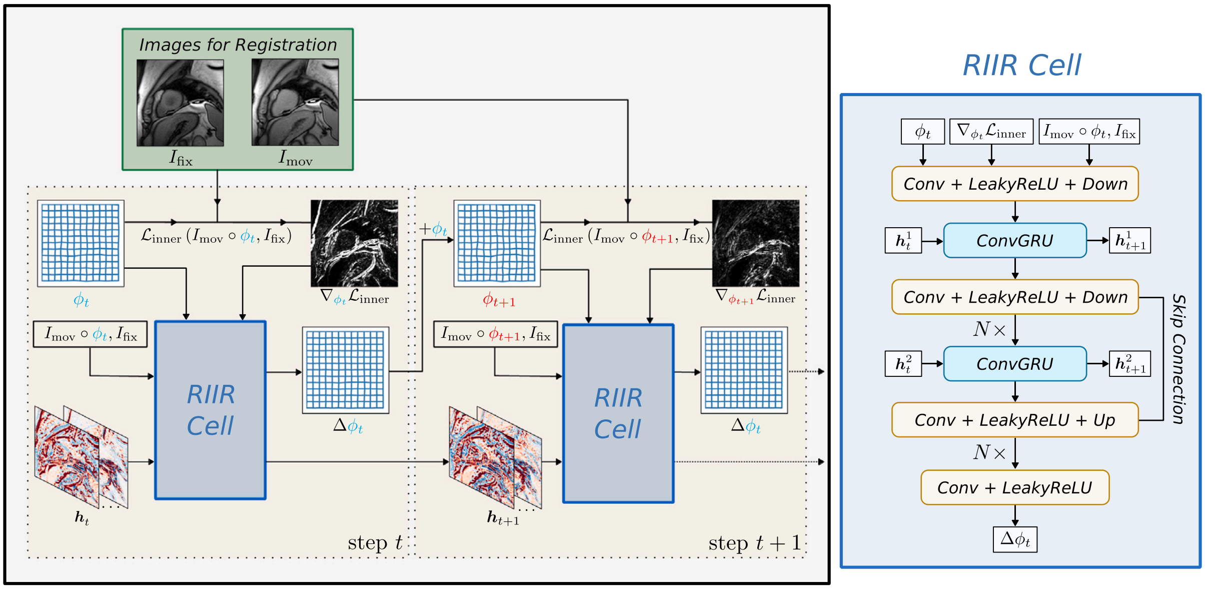
Images for Registration
I_{fix}
I_{mov}
ϕ_t
$\mathcal{L}_{\text{inner}}(I_{\text{mov}} \circ \phi_t, I_{\text{fix}})$
$\nabla_{\phi_t}\mathcal{L}_{\text{inner}}$
$I_{\text{mov}} \circ \phi_t, I_{\text{fix}}$
RIIR Cell
$\Delta\phi_t$
h_t
step t
$+\phi_t$
ϕ_{t+1}
$\mathcal{L}_{\text{inner}}(I_{\text{mov}} \circ \phi_{t+1}, I_{\text{fix}})$
$\nabla_{\phi_{t+1}}\mathcal{L}_{\text{inner}}$
$I_{\text{mov}} \circ \phi_{t+1}, I_{\text{fix}}$
RIIR Cell
$\Delta\phi_t$
h_{t+1}
step $t+1$
RIIR Cell
ϕ_t
$\nabla_{\phi_t}\mathcal{L}_{\text{inner}}$
$I_{\text{mov}} \circ \phi_t, I_{\text{fix}}$
Conv + LeakyReLU + Down
h_t^1
ConvGRU
h_{t+1}^1
Conv + LeakyReLU + Down
$N\times$
h_t^2
ConvGRU
h_{t+1}^2
Conv + LeakyReLU + Up
$N\times$
Conv + LeakyReLU
$\Delta\phi_t$
Skip Connection

Fig. 1. Overview of RIIR framework. Here, an illustrative cardiac image pair is shown as an example. The hidden states $\boldsymbol{h}_t = [\boldsymbol{h}_t^1, \boldsymbol{h}_t^2]$ are visualized in channel-wise fashion. The inner loss $\mathcal{L}_{\text{inner}}$ is calculated during each step of RIIR thus dynamically changing. When $t = 0$, the deformation field $\boldsymbol{\phi}_0$ is initialized as an identical transformation. In RIIR Cell, the dimensions of Conv and ConvGRU layer are dependent on the input (2D or 3D).

to be learned implicitly in the neural network. Therefore, we determine the inner loss $\mathcal{L}_{\text{inner}}$ by adapting the optimization objective in Eq. (1). Specifically, we use the similarity part $\mathcal{L}_{\text{sim}}$ in Eq. (1) as the inner loss at time step t:

$$\mathcal{L}_{\text{inner}}\left(I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}\right) = \mathcal{L}_{\text{sim}}\left(I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}\right). \tag{8}$$

The gradient of $\mathcal{L}_{\text{inner}}$ can be calculated using auto differentiation.

Nevertheless, in the standard RIM framework on the optimization objective in Eq. (1), the displacement fields serve as the primary optimization signal which may overlook the rich structural and contextual information present in the original and warped images. This information loss can be critical, particularly for capturing implicit regularization introduced by the structures in original images, which directly influences $\mathcal{L}_{\text{outer}}$ and the goal of optimizing the evaluation metric (*e.g.*, the overlapping of tissues or organs) which is usually not equivalent to $\mathcal{L}_{\text{outer}}$. We recognize such potential loss of information when relying solely on the gradient of the inner loss *w.r.t.* to $\boldsymbol{\phi}$. To address this, we extend the RIM framework, drawing inspiration from a more generalized formulation: the iterative amortized inference (Marino et al., 2018), ensuring that the model leverages the original images more effectively in the iterative optimization process. Such an extension takes the warped images as input for optimization, while not changing the optimization objective but enriching the information that could be passed into the neural network.

With this modification, our proposed framework performs an end-to-end iterative prediction of a dense transformation $\boldsymbol{\phi}$ in T steps for pairwise registration: Given the input image pair $\left(I_{\text{mov}}, I_{\text{fix}}\right)$, the optimization problem in Eq. (1) can be solved by the iterative update of $\boldsymbol{\phi}$. And the update rule at step $t \in \{0, 1, \ldots, T-1\}$ is:

$$\boldsymbol{\phi}_{t+1} = \boldsymbol{\phi}_t + \Delta\boldsymbol{\phi}_t, \tag{9}$$

where $\boldsymbol{\phi}_0$ is initialized as an identity mapping $\boldsymbol{\phi}_0(\chi) = \chi$. The update at step t, $\Delta\boldsymbol{\phi}_t$, is calculated by a recurrent update network g_θ by taking a channel-wise concatenation of

$$\{\boldsymbol{\phi}_t, \nabla_{\boldsymbol{\phi}_t}\mathcal{L}_{\text{inner}}\left(I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}\right), I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}\} \tag{10}$$

as input, where $\nabla_{\boldsymbol{\phi}_t}$ denotes the gradient *w.r.t.* $\boldsymbol{\phi}$ evaluated for $\boldsymbol{\phi} = \boldsymbol{\phi}_t$ and $\mathcal{L}_{\text{inner}}$ denotes the inner loss.

In the implementation of RIM, the iterative update Eq. (9) is achieved by a recurrent neural network (RNN) to generalize the update rule in Eq. (5) with hidden memory state variable $\boldsymbol{h}$ estimated for each time step t, which is the only neural network component in the whole pipeline. Unlike previous RIM-based works (Putzky et al., 2019; Sabidussi et al., 2021) which use two linear gated recurrent units (GRU) to calculate the hidden states $\boldsymbol{h}_t$, in RIIR, two convolutional gated recurrent units (ConvGRU) (Shi et al., 2015) are used to better preserve spatial correlation in the image. We further investigate the necessity of including such two-level recurrent structures in our experiment, particularly considering potential complexities in constructing computation graphs for neural networks. The iterative update equations of RIIR at step t have the following form, with the hidden memory states:

$$\left\{\Delta\boldsymbol{\phi}_t, \boldsymbol{h}_{t+1}\right\} = g_\theta(\boldsymbol{\phi}_t, \nabla_{\boldsymbol{\phi}_t}\mathcal{L}_{\text{inner}}\left(I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}\right), I_{\text{mov}}\circ\boldsymbol{\phi}_t, I_{\text{fix}}, \boldsymbol{h}_t), \tag{11}$$

$$\boldsymbol{\phi}_{t+1} = \boldsymbol{\phi}_t + \Delta\boldsymbol{\phi}_t, \tag{12}$$

where $\boldsymbol{h}_t = \{\boldsymbol{h}_t^1, \boldsymbol{h}_t^2\}$ denotes the two-level hidden memory states at step t. The size of $\boldsymbol{h}_t$ depends on the size of input image pair $(I_{\text{mov}}, I_{\text{fix}})$ with multiple channels. For $t = 1$, $\boldsymbol{h}_1$ is initialized to a zero input. We name our network g_θ as RIIR Cell, with its detailed architecture illustrated in Fig. 1. To address the difference between our RIIR from the existing gradient-based iterative algorithm (GraDIRN) (Qiu et al., 2022) under the same definition of $\mathcal{L}_{\text{inner}}$ as in Eq. (8), RIIR uses the gradient of inner loss as the neural network input to calculate the incremental update. On the other hand, GraDIRN takes the channel-wise warped image pair $(I_{\text{mov}}\circ\phi, I_{\text{fix}})$ and deformation field ϕ as the input to the network to output regularization update in Eq. (1), while the gradient of $\mathcal{L}_{\text{inner}}$ is added to the update of the deformation without any further processing thus does not directly affect the network part of the pipeline.

Unlike previous work in deep learning-based iterative deformable image registration methods which does not incorporate internal hidden states (Zhao et al., 2019; Fechter and Baltas, 2020; Qiu et al., 2022), we propose to combine the gradient information and hidden states as the network input. Our method also differs from Falta et al. (2022) in several aspects: in Falta et al. (2022), the input consists of a collection of images, displacement, sampled coordinates and MIND features; also, the part of the U-Net output channels serve as hidden states, instead of using dedicated ConvGRU units as in our design. Using $\boldsymbol{h}_t$ also suggests an analogy with gradient-based optimization methods such as the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS) to track and memorize progression (Putzky and Welling, 2017).

To substantiate this design, the input selections of RIIR will be further ablation-studied and discussed in our experiments.

Since the ground-truth deformation field is not known in deformable image registration, we use the optimization objective Eq. (1) as the proposed outer loss to optimize the parameters θ of RIIR Cell g_θ. We incorporate a weighted sum of losses for the outer loss $\mathcal{L}_{\text{outer}}$ to ensure that each step contributes to the final prediction:

$$\mathcal{L}_{\text{outer}}(\theta) = \sum_{t=1}^{T} w_t \left(\mathcal{L}_{\text{sim}} \left(I_{\text{mov}} \circ \boldsymbol{\phi}_t, I_{\text{fix}} \right) + \lambda \mathcal{L}_{\text{reg}} \left(\boldsymbol{\phi}_t \right) \right), \tag{13}$$

where w_t is a (positive) scalar indicating the weight of step t. In our experiment, both uniform ($w_t = \frac{1}{T}$) and exponential weights ($w_t = 10^{\frac{t-1}{T-1}}$) are considered and will be compared in the experiments. It is noticeable that the design of using a (weighted) average of the stepwise loss also makes our proposed RIIR different from other iterative deep learning-based methods (Qiu et al., 2022; Zhao et al., 2019) which use only the final output to calculate the loss, and Falta et al. (2022) use a uniform weight, addressing the fact that early steps in the prediction process were neglected before.

2.4. Metrics

Similarity Functions for Inner Loss $\mathcal{L}_{\text{inner}}$: In the context of image registration, unlike inverse problems with straightforward forward models, the problem is addressed as a broader optimization challenge. Therefore, it requires an investigation of choosing a (differentiable) function acting as the inner loss function evaluating the quality of estimation of $\boldsymbol{\phi}_t$ iteratively in RIIR. Furthermore, the gradient of $\mathcal{L}_{\text{inner}}$ as an input of a convolutional recurrent neural network has not been studied before for deformable image registration. These motivate the study on the different choices of $\mathcal{L}_{\text{inner}}$ under a fixed choice of outlet loss $\mathcal{L}_{\text{outer}}$. In this work, we evaluate three similarity functions: mean squared error (MSE), normalized cross-correlation (NCC) (Avants et al., 2008), and normalized mutual information (NMI) (Studholme et al., 1999).

The MSE between two 3D images $I_1, I_2 \in \mathbb{R}^{d_x \times d_y \times d_z}$ is defined as follows:

$$\text{MSE}\left(I_1, I_2\right) = \frac{1}{d_x d_y d_z} \|I_1 - I_2\|_2^2, \tag{14}$$

where $|\Omega_I| = d_x d_y d_z$ denotes the all possible coordinates. The MSE metric is minimized when pixels of I_1 and I_2 have the same intensities. Therefore, it is sensitive to the contrast change. In comparison, the NCC metric measures the difference between images with the image intensity normalized. The NCC difference between I_1 and I_2 is given by:

$$\text{NCC}(I_1, I_2) = \frac{1}{|\Omega_{I_1}|} \sum_{\chi \in \Omega_{I_1}} \frac{\sum_{\chi' \in \Omega_\chi} (I_1(\chi') - \bar{I}_1(\chi))(I_2(\chi') - \bar{I}_2(\chi))}{\sqrt{\hat{I}_1(\chi) \hat{I}_2(\chi)}}, \tag{15}$$

where Ω_{I_1} denotes all possible coordinates in I_1, Ω_χ represents a neighborhood of voxels around coordinate position χ and $\bar{I}(\chi)$ and $\hat{I}(\chi)$ denote the (local) mean and variance in Ω_χ.

Compared to MSE and NCC, NMI is shown to be more robust when the linear relation of signal intensities between two images does not hold (Studholme et al., 1999; de Vos et al., 2020), which is often the case in quantitative MRI as the signal models are mostly exponential (Messroghli et al., 2004; Chow et al., 2022). The NMI between two images can be written as:

$$\text{NMI}(I_1, I_2) = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)}, \tag{16}$$

where $H(I_1)$ and $H(I_2)$ are marginal entropies of I_1 and I_2, respectively, and $H(I_1, I_2)$ denotes the joint entropy of the two images. Since the gradient is both necessary for $\mathcal{L}_{\text{inner}}$ and $\mathcal{L}_{\text{outer}}$ we adopt a differentiable approximation of the joint distribution proposed in Qiu et al. (2021) based on Parzen window with Gaussian distributions (Thévenaz and Unser, 2000).

Regularization Metrics: To ensure a smooth and reasonable deformation field, we primarily use a diffusion regularization loss which penalizes large displacements in $\boldsymbol{\phi}$ acting on $I \in \mathbb{R}^{d_x \times d_y \times d_z}$ (Fischer and Modersitzki, 2002):

$$\mathcal{L}_{\text{diff}} = \frac{1}{|\Omega_I|} \sum_{\chi \in \Omega_I} \|\nabla \boldsymbol{\phi}(\chi)\|_2^2, \tag{17}$$

where $|\Omega_I| = d_x d_y d_z$, $\nabla \boldsymbol{\phi}(\chi)$ denotes the Jacobian of $\boldsymbol{\phi}$ at coordinate χ. It is noticeable that Eq. (17) and its gradient are not evaluated in each RIIR inference step as indicated in Eq. (8), the outer loss $\mathcal{L}_{\text{outer}}$ and the data-driven training process can guide the RIIR Cell g_θ to learn the regularization implicitly.

Since RIIR aims to learn implicit regularization from data with the outer loss, we also include two additional regularization metrics in our ablation study: curvature regularization and linear elastic regularization with fixed elasticity parameters (Fischer and Modersitzki, 2004). The curvature loss penalizes the second spatial derivatives of the displacement field $\boldsymbol{\phi}$, encouraging smoothness in the rate of change of $\boldsymbol{\phi}$:

$$\mathcal{L}_{\text{curv}} = \frac{1}{|\Omega_I|} \sum_{\chi \in \Omega_I} \sum_{i,j=1}^{3} \left\| \frac{\partial^2 \boldsymbol{\phi}(\chi)}{\partial \chi_i \partial \chi_j} \right\|_2^2, \tag{18}$$

where $\frac{\partial^2 \boldsymbol{\phi}(\chi)}{\partial \chi_i \partial \chi_j}$ denotes the second-order partial derivative of $\boldsymbol{\phi}$ *w.r.t.* dimensions χ_i and χ_j at coordinates χ.

The linear elastic regularization aims to regularize the displacement field by considering both the divergence and strain of the field. Its variational formulation can be described as follows (Fischer and Modersitzki, 2004):

$$\mathcal{L}_{\text{elas}} = \frac{1}{|\Omega_I|} \sum_{\chi \in \Omega_I} \left(\frac{\lambda_e}{2} (\text{div}\, \boldsymbol{\phi}(\chi))^2 + \frac{\mu}{4} \sum_{i,j=1}^{3} \left\| \frac{\partial \phi_i(\chi)}{\partial \chi_j} + \frac{\partial \phi_j(\chi)}{\partial \chi_i} \right\|_2^2 \right), \tag{19}$$

where $\text{div}\, \boldsymbol{\phi}(\chi) = \sum_{i=1}^{3} \frac{\partial \phi_i(\chi)}{\partial \chi_i}$ denotes the divergence of the displacement field, λ_e and μ_e are Lamé parameters controlling the strength of volumetric and shear deformation respectively. The first term penalizes volume changes, while the second term regularizes shear deformation of the displacement field $\boldsymbol{\phi}$.

3. Experiments

3.1. Dataset

We evaluated our proposed RIIR framework on two separate datasets: (1) A 3D brain MRI image dataset with inter-subject registration setup, OASIS (Marcus et al., 2007) with pre-processing from Hoopes et al. (2021), denoted as **OASIS**. (2) A 3D lung CT dataset with intra-subject registration setup, National Lung Screening Trial (NLST) (Aberle et al., 2011), provided and processed by the Learn2Reg challenge (Hering et al., 2022). This dataset is denoted by **NLST**. (3) A 2D quantitative cardiac MRI image datasets based on multiparametric SAturation-recovery single-SHot Acquisition (mSASHA) image time series (Chow et al., 2022), denoted as **mSASHA**. These datasets, each serving our interests in inter-subject tissue alignment and respiratory motion correction with and without contrast variation.

OASIS: The dataset contains 414 subjects, where for each subject, the normalized T_1-weighted scan was acquired. The subjects are split into train/validation/test with counts of $[300, 30, 84]$. For training, images are randomly paired using an on-the-fly data loader, while in the validation and test sets, all images are paired with the next image in a fixed order. The dataset was preprocessed with FreeSurfer and SAMSEG by Hoopes et al. (2021), resulting in skull-stripped and bias-corrected 3D volumes with a size of 160 × 192 × 224. We further resampled the images into a size of 128 × 128 × 128 with intensity clipping between $(1\%, 99\%)$ percentiles. Fig. 2 illustrates an example

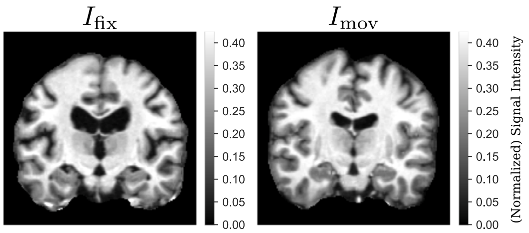

Fig. 2. An example of OASIS dataset for two subjects as I_{fix} and I_{mov}. The choices of I_{fix} and I_{mov} are random during training.

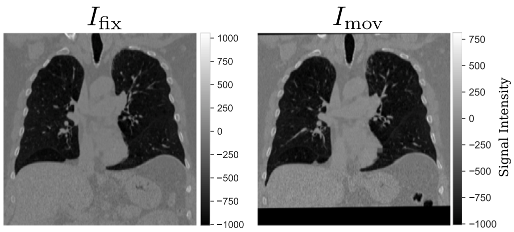

Fig. 3. An example of NLST dataset for a single subject with I_{fix} corresponding to the image captured at inspiratory phase and I_{mov} corresponding to the image captured at expiratory phase.

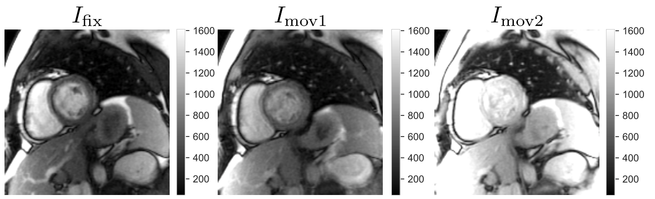

Fig. 4. An example of mSASHA dataset, from left to right: I_{fix}, I_{mov} (random sample 1), and I_{mov} (random sample 2). The three images were taken from the same image series, with different acquisition time points. To emphasize the difference in both signal intensity and contrast across images in a single series, the color ranges are set to be the same for the three images.

pair of OASIS images, showcasing the consistency in signal intensity and contrast.

NLST: We use the public NLST dataset from the Learn2Reg challenge. The dataset consists of 210 subjects with intra-subject inhale and exhale lung CT images that are affinely pre-aligned. The subjects are split into train/validation/test with counts of $[170, 10, 30]$. Following the convention, during training, only the inhale and exhale images from the same subject are paired, with the inhale image chosen as I_{fix} for all pairs. The images were preprocessed and resampled to a size of $224 \times 192 \times 224$. The keypoints and their correspondences in the lung lobe, along with the lung lobe mask, were provided by the organizers, and obtained by the corrField algorithm (Heinrich et al., 2015) and nnU-Net (Isensee et al., 2021) respectively. An example pair of NLST images is shown in Fig. 3, demonstrating the large deformation required for this task.

mSASHA: During an free-breathing mSASHA examination, a time series of $N = 30$ real-valued 2D images, denoted by $I = \{I_n |\ n = 1, 2, \dots, N\}$, are acquired for the same subject. In the setting of quantitative MRI, we aim to spatially align N images in a single sequence I into a common fixed template image I_{fix}, by individually performing N pairwise registration processes over (I_n, I_{fix}) where $n = 1, 2, \dots, N$.

The mSASHA acquisition technique (Chow et al., 2022) is a voxel-wise 3-parameter signal model based on the joint cardiac T_1-T_2 signal model:

$$\mathcal{S}\left(T_1, T_2, A\right) = A\left\{1 - \left[1 - \left(1 - e^{-TS/T_1}\right) e^{-TE/T_2}\right] e^{-TD/T_1}\right\}, \tag{20}$$

where (TS, TE, TD) denotes the set of three acquisition variables, and (T_1, T_2, A) is the set of parameters to be estimated for each voxel coordinate of the image series. The sequence for mSASHA consists of a reference image without magnetization preparation, a series of saturation recovery (SR) images, and a series of both SR and T_2-prepared images. We encourage interested readers to refer to Chow et al. (2022) for a more detailed explanation.

In our experiment, an in-house mSASHA dataset was used. This fully anonymized raw dataset was provided by NIH, and was considered "non-human subject data research" by the NIH Office of Human Subjects Research". The dataset comprises 120 subjects, with each subject having 3 slice positions, resulting in a total number of 360 slices. Each mSASHA time series consists of a fixed length of $N = 30$ images. We split **mSASHA** into train/validation/test with counts of $[84, 12, 24]$ by subjects to avoid data leakage across the three slices. Given variations in image sizes due to different acquisition conditions, we first center-cropped the images into the same size of 144×144. Subsequently, we applied intensity clipping between $(1\%, 95\%)$ percentiles to mitigate extreme signal intensities from the chest wall region. We selected the last image in the series, *i.e.,* in the T_2 preparation stage, as the template I_{fix}, which is then a T2-weighted image with the greatest contrast between the myocardium and adjacent blood pool. An illustrative example of mSASHA images can be found in Fig. 4, showing varying contrasts and non-rigid motion across frames.

3.2. Evaluation metrics

For evaluating the smoothness of the deformation field, we employ three complementary metrics based on the Jacobian of the deformation $J_{\boldsymbol{\phi}} = \nabla \boldsymbol{\phi}$: (1) the percentage of negative Jacobian determinant $|J_{\phi}|_{\leq 0}$, which quantifies the proportion of regions exhibiting folding or topology-breaking transformations, (2) the standard deviation of the log-Jacobian determinant $\text{std}(\log |\boldsymbol{J}_{\boldsymbol{\phi}}|)$, which characterizes the global variation of volume changes, and (3) the magnitude of the spatial gradient of the Jacobian determinant $\text{mag}(\nabla |\boldsymbol{J}_{\boldsymbol{\phi}}|)$, which measures the local rate of change in volume deformation. To allow fair comparisons between methods, we meticulously adjust the regularization parameter λ for each baseline to achieve comparable levels of smoothness of transformation. To assess registration accuracy, we utilize structural similarity metrics that are independent of optimization objectives $\mathcal{L}_{\text{sim}}$ and $\mathcal{L}_{\text{reg}}$, with dataset-specific metrics detailed in subsequent sections.

For OASIS, two metrics, Dice score and Hausdorff distance (HD) are considered to evaluate segmentation quality after registration. Given two sets $X \subset M$ and $Y \subset M$, the Dice score, is defined to measure the overlapping of X and Y:

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{21}$$

Similarly, the Hausdorff distance of two aforementioned sets X and Y is given by:

$$HD(X, Y) := \max\left\{\sup_{x \in X} d(x, Y), \sup_{y \in Y} d(X, y)\right\}, \tag{22}$$

where $d(\cdot, \cdot)$ is a metric (2-norm in this work) on M and $d(x, Y) := \inf_{y \in Y} d(x, y)$. As a remark, in this work, we consider the average across all segmentation labels to calculate the Dice score and HD in **OASIS** instead of only considering the major regions.

For the NLST dataset, we evaluate the registration performance using both lung lobe mask overlap and Target Registration Error (TRE) of the keypoints. For a quantitative evaluation of local registration accuracy, TRE is calculated based on the provided keypoint pairs. Given a set of corresponding keypoint pairs $\{(\boldsymbol{p}_i, \boldsymbol{q}_i)\}_{i=1}^{L}$ where $\boldsymbol{p}_i$ represents a keypoint in the moving image and $\boldsymbol{q}_i$ its corresponding point in the fixed image, TRE is defined as:

$$\text{TRE} = \frac{1}{L} \sum_{i=1}^{L} \|\boldsymbol{\phi}(\boldsymbol{p}_i) - \boldsymbol{q}_i\|_2, \tag{23}$$

where $\|\cdot\|_2$ denotes the Euclidean distance.

Furthermore, we also evaluate two more independent metrics for the mSASHA dataset proposed by Huizinga et al. (2016) isolated from training. The metrics are based on the principal component analysis (PCA) of images. Assume $M \in \mathbb{R}^{d_x d_y \times N}$ is the matrix representation of I, where a row of M represents a coordinate in the image space. The correlation matrix of M is then calculated by:

$$K = \frac{1}{d_x d_y - 1} \Sigma^{-1} (M - \overline{M})^{\intercal} (M - \overline{M}) \Sigma^{-1}, \tag{24}$$

where Σ is a diagonal matrix representing the standard deviation of each column, and $\overline{M}$ denotes the column-wise mean for each column entry. Since an ideal qMRI model assumes a voxel-wise tissue alignment, the actual underlying dimension of K can be characterized by a low-dimensional (linear) subspace driven by the signal model. In the mSASHA signal model, the dimension of such a subspace is assumed to be four according to Eq. (20), determined by the number of parameters to be estimated. With the fact that the trace of K, $\mathrm{tr}(K)$ is a constant, two PCA-based metrics were proposed as follows:

$$\mathcal{D}_{\mathrm{PCA1}} = \sum_{i=1}^{N} \sigma_i - \sum_{j=1}^{L} \sigma_j = \mathrm{tr}(K) - \sum_{j=1}^{L} \sigma_j, \tag{25}$$

$$\mathcal{D}_{\mathrm{PCA2}} = \sum_{j=1}^{N} j \sigma_j, \tag{26}$$

where σ_i denotes the ith largest eigenvalue of K. Both metrics were designed to penalize a long-tail distribution of the spectrum of K, and L is a hyperparameter regarding the number of parameters of the signal model. For $\mathcal{D}_{\mathrm{PCA1}}$, an ideal scenario would involve all images perfectly aligning with tissue anatomy and the signal model, resulting in a value of 0. Meanwhile, the interpretation of $\mathcal{D}_{\mathrm{PCA2}}$ further emphasizes the tail of the eigenvalues, thus enlarging the gaps across experiments.

To narrow the analysis to the region of interest to the heart region, the calculation is confined to this area by cropping the resulting images before computing the metric. This constraint ensures that the evaluation is focused on the relevant anatomical structures.

3.3. Experimental settings

We here summarize the main experiments for evaluation and further ablation experiments for RIIR. For all experiments, the main workflow is to register the image series I of length N in a pairwise manner: that is, we first choose a template I_{fix}, and then perform N registrations. When $N = 2$, the registration process simplifies to straightforward pairwise registration.

Experiment 1: Comparison Study with Varying Data Availability

We introduce five data-availability scenarios to evaluate the robustness of the models when data availability is limited, on both datasets, which often happens in both research settings and clinical practices as the number of subjects is heavily limited. The training data availability settings in this study were set to $[5\%, 10\%, 25\%, 50\%, 100\%]$ for all datasets. It is worth noticing that for limited data availability scenarios, the data used for training remained the same for all models in consideration, and the leave-out test split remained unchanged for all scenarios.

Experiment 2: Inclusion of Hidden States

Unlike most related works utilizing the original RIM framework (Lønning et al., 2019; Sabidussi et al., 2021) where two levels of hidden states are considered, we explored the impact of modifying or even turning off hidden states. In our implementation of convolutional GRU, at most two levels of hidden states $\boldsymbol{h}_t^1$ and $\boldsymbol{h}_t^2$ are considered, following recent works using RIM (Lønning et al., 2019; Sabidussi et al., 2021, 2023). Both hidden states were configured with 32 channels in their corresponding convolutional GRU layers. All experiments in this ablation study were performed in the validation split. We present the results for OASIS and NLST as they represent distinct imaging modalities (brain MRI and lung CT) and registration scenarios.

Experiment 3: Inclusion of Gradient of Inner Loss $\nabla_{\boldsymbol{\phi}_t} \mathcal{L}_{\mathrm{inner}}$ as RIIR Input

We performed an experimental study on the input composition for RIIR. As shown in Eq. (12), the goal was to study the data efficiency and the registration performance by incorporating the gradient of $\mathcal{L}_{\mathrm{inner}}$ in RIIR. We could achieve ablation by changing the input of g_θ. A comparison with other input modeling strategies seen in Qiu et al. (2022) was proposed against the gradient-based input for g_θ. Depending on whether the moving image is deformed (explicit) or not (implicit), as well as the original RIM formulation, we ended up with four input compositions:

1. Implicit Input without $\nabla \mathcal{L}_{\mathrm{inner}}$: $[\boldsymbol{\phi}_t, I_{\mathrm{mov}}, I_{\mathrm{fix}}]$;
2. Explicit Input without $\nabla \mathcal{L}_{\mathrm{inner}}$: $[\boldsymbol{\phi}_t, I_{\mathrm{mov}} \circ \boldsymbol{\phi}_t, I_{\mathrm{fix}}]$;
3. RIM Input: $[\boldsymbol{\phi}_t, \nabla_{\boldsymbol{\phi}_t} \mathcal{L}_{\mathrm{inner}}]$;
4. RIIR Input: $[\boldsymbol{\phi}_t, \nabla_{\boldsymbol{\phi}_t} \mathcal{L}_{\mathrm{inner}}, I_{\mathrm{mov}} \circ \boldsymbol{\phi}_t, I_{\mathrm{fix}}]$.

This study aimed to provide information on the impact of different input compositions on the efficiency of RIIR when data availability varies. We conducted the experiment with two data availability choices ($[5\%, 100\%]$) to examine the data efficiency and other potential influences induced by the gradient input.

Experiment 4: Regularization Analysis

It is known that the regularization metric and weight influence the registration performance. In this experiment, we want to investigate the change in the evaluation metrics. For comparison, we used diffusion, curvature, and elastic regularization functions. For elastic regularization, we set the elastic parameters to be $[\lambda_e = 540.8, \mu = 22.5]$ for OASIS and $[\lambda_e = 45.33, \mu = 8]$ for NLST according to the literature (Kumaresan and Radhakrishnan, 1996; Lai-Fook and Hyatt, 2000; Reithmeir et al., 2024).

Experiment 5: RIIR Architecture Ablation Since RIIR is the first attempt to formulate and implement the RIM framework for medical image registration, we performed an ablation study on the RIIR network architecture for the number of evaluation steps T.

3.4. Baseline methods and implementation details

We compared our proposed method to various registration methods that are closely related to our interest. We use the same choice of similarity and regularization functions for all deep learning-based methods in the comparative study for fair comparison, unless otherwise indicated. We used MSE loss for OASIS, NCC with window size $w = 5$ for NLST, and NMI with $n = 32$ bins for mSASHA. We used diffusion regularization for all three datasets in the comparative study. The methods and applicable hyperparameters are described as follows:

- Elastix (Klein et al., 2009): An iterative optimization-based registration toolbox. Specifically, we used ITK-Elastix (Ntatsis et al., 2023) in Python. Three resolution levels with third-order B-spline transformation and a grid spacing of four were applied for all datasets.
- VoxelMorph (Balakrishnan et al., 2019): We used $\lambda = 0.02$ for OASIS as in the original paper, $\lambda = 0.15$ for NLST, and $\lambda = 0.3$ for mSASHA. The channels used in each downsampling encoder block were $[16, 16, 32, 32, 32]$. Two layers of activation with channels $[16, 16]$ after the decoder was used.
- GraDIRN (Qiu et al., 2022): A multi-resolution multi-step deep learning method that uses explicit similarity loss gradient and dense CNN to produce incremental updates. We followed the original implementation with 3 resolutions and 3 steps per resolution and use the last-step output for loss calculation. The training losses were set to the same as VoxelMorph, with weight parameter $\lambda = 0.015$ for OASIS, $\lambda = 0.125$ for NLST, and $\lambda = 0.25$ for mSASHA.

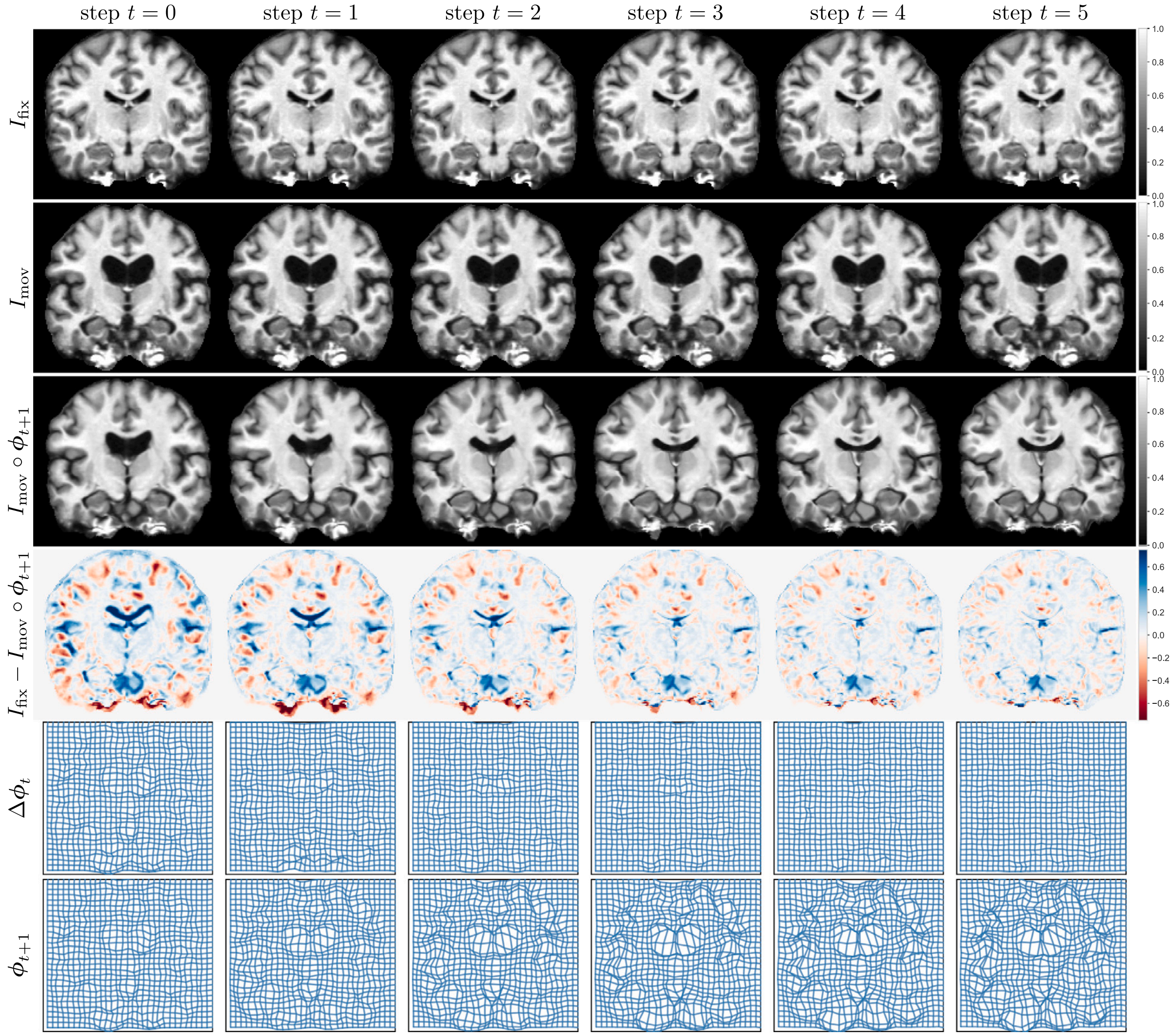

Fig. 5. A visualization of RIIR inference on OASIS test split, visualized with a 2D slice and in-plane deformation. The inference step was set to 6 in both training and inference. All images in the same row were plotted using the same color range for better consistency.

- LapIRN (Mok and Chung, 2020): A multi-resolution deep learning method with Laplacian pyramid networks. We followed the original implementation of the displacement version with three-stage training and used multi-level NCC in the paper as similarity loss for OASIS, with $\lambda = 0.6$. For NLST and mSASHA, we followed the same loss and weight as for VoxelMorph.
- Learn-to-Optimize (L2O) (Falta et al., 2022): A multi-step deep learning method that uses additional input modalities, including 3D MIND features (Heinrich et al., 2012) and sampled coordinates for 3D datasets. Constrained by the inherent 3D nature of MIND features, we evaluated L2O on the OASIS and NLST datasets. The original paper uses keypoint supervision only; therefore, we replaced the loss function by the same unsupervised loss as used by other baseline methods while keeping the uniform weight for each step, with $\lambda = 0.03$ for OASIS and $\lambda = 0.225$ for NLST.
- Recursive-cascaded VoxelMorph (RCVM) (Zhao et al., 2019):A single-resolution iterative deep learning method that uses cascaded U-Nets and composition of the deformation field to generate the final output. We used the same VoxelMorph backbone and loss functions with $\lambda = 0.015$ for OASIS, $\lambda = 0.25$ for mSASHA, and $\lambda = 0.125$ for NLST.

We implemented the RIIR in the following settings for experiment 1: The backbone network is the same as the VoxelMorph in baseline, with additional ConvGRU in the second level with 32 channels. We used the inference steps $T = 6$ and exponential weighting for $w_t = 10^{\frac{t-1}{T-1}}$. To ensure a similar level of smoothness, the trade-off parameter for RIIR was set to $\lambda = 0.0125$ for OASIS, $\lambda = 0.125$ for NLST, and $\lambda = 0.25$ for mSASHA. The optimizer of all methods remained the same using Adam (Kingma and Ba, 2015) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to 1×10^{-4} for all models. For all experiments, the maximum epochs was set to 100 epochs with early stopping if the evaluation metrics does not improve for 10 epochs. The experiments were performed on an NVIDIA RTX 4090 GPU with a VRAM of 24 GB. The source codes for RIIR, with the implementation of the baseline models and data processing are publicly available.[1]

4. Results

4.1. Experiment 1: Comparison study with varying data availability

An illustrative visualization of RIIR inference on an example test data, can be found in Fig. 5. The results for **OASIS** are presented in Fig. 6, as well as . LapIRN, leveraging its multi-resolution architecture,

[1] gitlab.tudelft.nl/ai4medicalimaging/riir-public.

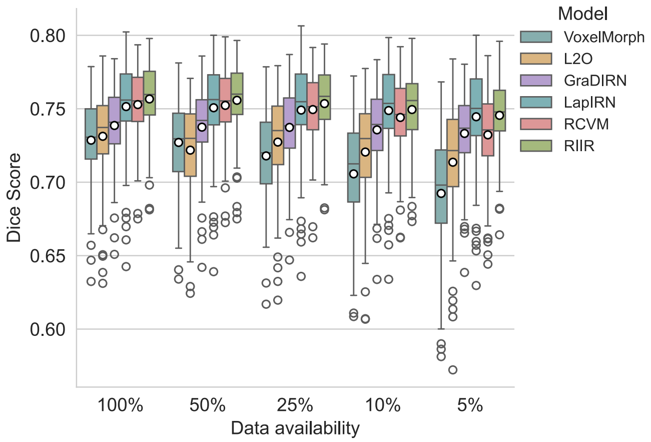

Fig. 6. Results of Experiment 1 with boxplots for Dice score on OASIS. The circle denotes the mean of the metric of interest. The segmentation metric Dice is calculated for all 35 segmentation labels and post-processed by taking the average.

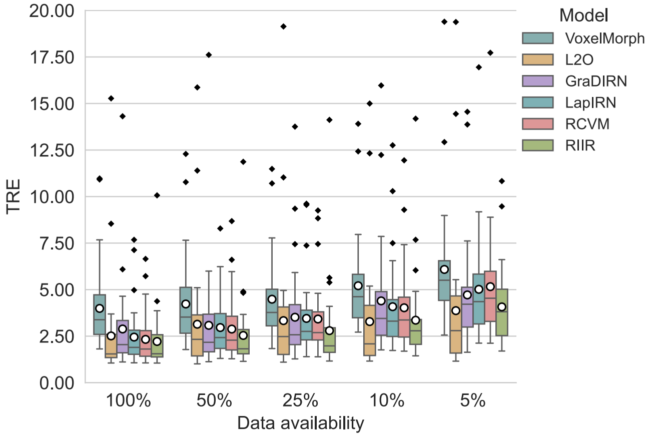

Fig. 7. Results of Experiment 1 with boxplots for TRE on NLST. The circle denotes the mean of the metric of interest. The TRE is calculated based on the anatomically meaningful keypoint pairs provided by corrField algorithm within lung lobes. For visualization, outliers over a TRE of 20 mm were excluded but used for statistical calculation.

also demonstrates robust performance. It is evident that RIIR outperforms most deep learning-based baselines when data availability is severely limited and maintains consistent performance across various data availability scenarios, showcasing its data efficiency and accuracy.

The results for **NLST** are shown in Fig. 7. An illustrative visualization of RIIR inference on an inhale-exhale lung CT pair, can be found in Fig. 20. The registration of lung CT presents unique challenges due to the large deformation between respiratory phases. RIIR demonstrates superior performance in capturing these large deformations, achieving lower TRE while maintaining anatomically plausible transformations.

The results of this experiment on mSASHA are shown in Fig. 8 using a composition of boxplots. Both LapIRN and RCVM achieved superior performance in group-wise registration, with RCVM showing slightly better results in terms of $\mathcal{D}_{\text{PCA1}}$. Our proposed RIIR demonstrated comparable performance levels in terms of $\mathcal{D}_{\text{PCA1}}$. The qualitative visualization of RIIR inference on mSASHA test split is shown in Fig. 21.

For a comprehensive quantitative comparison, Table 1 presents detailed statistics across all datasets under full data availability, including performance metrics, model parameters, memory consumption, and computational time.

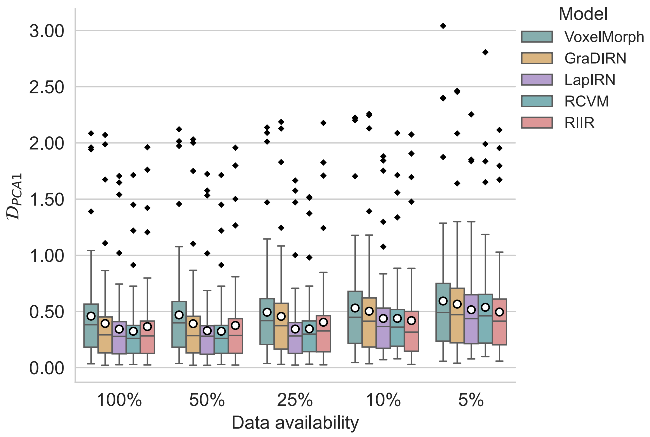

Fig. 8. Results of Experiment 1 with boxplots for $\mathcal{D}_{pca1}$ on mSASHA. The circle denotes the mean of the metric of interest. The group-wise metric $\mathcal{D}_{\text{PCA1}}$ was calculated based on further center-cropping at a ratio of 70% on the warped images.

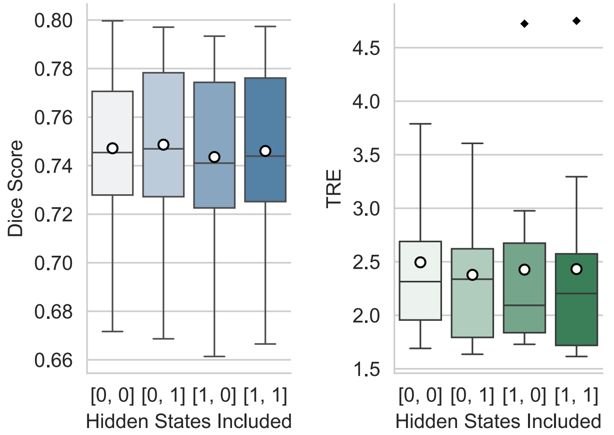

Fig. 9. Results of Experiment 2 evaluated on OASIS validation set (left) regarding Dice score and NLST validation set (right) regarding TRE. Here, for example, $[0, 0]$ denotes the case that no hidden states are considered, and $[1, 1]$ denotes both hidden states were considered in the pipeline. Two-sided Wilcoxon tests were conducted for $[0, 1]$ against other settings with statistical significance (p< 0.05), except for $[0, 0]$ ($p = 0.076$) in OASIS dataset.

4.2. Experiment 2: Inclusion of hidden states

The architectural settings were kept the same as in the aforementioned experiments, and the results are shown in Fig. 9. Although improved performance using hidden states ([0, 1]) over no hidden states is only observed in the NLST dataset, we empirically noticed that training was more stable when hidden states were enabled. It is also notable that the inclusion of hidden states in the pipeline does not incur significant computational overhead. For the OASIS dataset, the addition of hidden states only increases VRAM consumption by approximately 800 MB, while maintaining the model's computational speed.

4.3. Experiment 3: Inclusion of inner loss gradient as RIIR input

The results are shown in Fig. 10. It can be observed that the network struggled if the warped image $I_{\text{mov}} \circ \boldsymbol{\phi}$ is only implicitly fed to the RIIR cell, *i.e.,* there is an additional cost to learn the transformation. When training data is abundant (100% availability), RIIR and Explicit input achieve comparable performance, as evidenced by the non-significant statistical differences ($p = 0.17$ for OASIS and $p = 0.27$ for NLST). This suggests that with sufficient training data, both approaches can effectively learn the registration task, though RIIR maintains its advantage

Table 1
Quantitative comparison of different registration methods with 100% data availability. For each evaluation metric except Par. (number of parameters), Mem. (VRAM consumption), and T (training/inference time), we report mean ± std. VRAM consumption indicates the peak GPU memory consumption during training. The training/inference time is profiled for a batch on GPU (excluding data loading time), except for elastix where CPU time is reported. For LapIRN, metrics are reported at level 3. The best performance is shown in bold and marked with $*$ if there is a statistically significant difference ($p < 0.05$) from the second best by a two-sided Wilcoxon signed-rank test.

Dataset	Method	Dice	HD (mm)	$\lvert J_{\phi}\rvert_{\leq 0}$ (%)	std(log $\lvert J_{\phi}\rvert$)	mag($\nabla\lvert J_{\phi}\rvert$)	Par. (K)	Mem. (G)	T (s)
OASIS	Affine	0.563 ± 0.063	8.56 ± 1.19	–	–	–	–	–	–
	elastix	0.696 ± 0.040	3.89 ± 0.58	0.003 ± 0.002	0.254 ± 0.045	$\mathbf{0.015 \pm 0.002}^{*}$	–	–	17.45
	VM	0.729 ± 0.030	3.61 ± 0.49	0.016 ± 0.031	0.295 ± 0.115	0.029 ± 0.003	320	1.57	0.17/0.14
	L2O	0.733 ± 0.032	3.62 ± 1.01	0.021 ± 0.028	0.411 ± 0.132	0.035 ± 0.003	343	20.52	1.19/0.86
	GraDIRN	0.739 ± 0.027	3.54 ± 0.45	0.012 ± 0.003	0.206 ± 0.065	0.022 ± 0.002	269	16.13	0.57/0.25
	LapIRN	0.751 ± 0.026	3.50 ± 0.44	0.010 ± 0.002	0.260 ± 0.051	0.034 ± 0.002	923	3.78	0.16/0.10
	RCVM	0.753 ± 0.025	$\mathbf{3.46 \pm 0.43}^{*}$	$\mathbf{0.002 \pm 0.001}$	$\mathbf{0.189 \pm 0.043}^{*}$	0.027 ± 0.003	1920	10.20	0.64/0.54
	RIIR	$\mathbf{0.756 \pm 0.025}^{*}$	3.48 ± 0.41	0.011 ± 0.009	0.264 ± 0.073	0.029 ± 0.003	436	11.82	0.55/0.27
Dataset	Method	TRE (mm)	Dice	$\lvert J_{\phi}\rvert_{\leq 0}$ (%)	std(log $\lvert J_{\phi}\rvert$)	mag($\nabla\lvert J_{\phi}\rvert$)	Par. (K)	Mem. (G)	T (s)
NLST	Affine	8.43 ± 3.97	0.873 ± 0.041	–	–	–	–	–	–
	elastix	5.01 ± 2.92	0.946 ± 0.007	$\mathbf{0.000 \pm 0.000}^{*}$	$\mathbf{0.160 \pm 0.031}^{*}$	$\mathbf{0.011 \pm 0.003}^{*}$	–	–	14.21
	VM	3.99 ± 2.48	0.957 ± 0.012	0.120 ± 0.162	0.677 ± 0.399	0.057 ± 0.005	320	0.93	0.18/0.15
	L2O	2.51 ± 2.96	0.961 ± 0.019	0.451 ± 0.363	1.325 ± 0.554	0.068 ± 0.007	343	11.81	0.83/0.52
	GraDIRN	2.89 ± 2.49	0.960 ± 0.020	0.114 ± 0.068	0.454 ± 0.269	0.058 ± 0.005	279	9.32	0.37/0.15
	LapIRN	2.44 ± 1.60	0.967 ± 0.006	0.073 ± 0.065	0.555 ± 0.240	0.058 ± 0.004	923	2.18	0.14/0.05
	RCVM	2.32 ± 1.35	$\mathbf{0.968 \pm 0.005}^{*}$	0.229 ± 0.278	0.913 ± 0.511	0.063 ± 0.005	1920	5.89	0.38/0.27
	RIIR	$\mathbf{2.21 \pm 1.71}^{*}$	0.966 ± 0.012	0.108 ± 0.259	0.568 ± 0.481	0.048 ± 0.004	436	7.12	0.41/0.18
Dataset	Method	$\mathcal{D}_{\text{PCA1}}$	$\mathcal{D}_{\text{PCA2}}$	$\lvert J_{\phi}\rvert_{\leq 0}$ (%)	std(log $\lvert J_{\phi}\rvert$)	mag($\nabla\lvert J_{\phi}\rvert$)	Par. (K)	Mem. (G)	T (s)
mSASHA	Raw	1.28 ± 0.88	46.70 ± 10.40	–	–	–	–	–	–
	elastix	0.40 ± 0.38	35.81 ± 4.72	0.162 ± 0.005	0.551 ± 0.685	$\mathbf{0.022 \pm 0.009}^{*}$	–	–	2.46
	VM	0.46 ± 0.42	36.49 ± 5.13	$\mathbf{0.002 \pm 0.001}$	0.210 ± 0.082	0.053 ± 0.012	79	0.06	0.05/0.01
	GraDIRN	0.39 ± 0.38	35.85 ± 4.51	0.002 ± 0.002	$\mathbf{0.156 \pm 0.043}^{*}$	0.040 ± 0.012	89	0.21	0.09/0.04
	LapIRN	0.34 ± 0.32	35.11 ± 4.09	0.006 ± 0.002	0.248 ± 0.128	0.054 ± 0.015	309	0.09	0.04/0.01
	RCVM	$\mathbf{0.32 \pm 0.30}^{*}$	$\mathbf{34.79 \pm 3.83}^{*}$	0.005 ± 0.002	0.275 ± 0.081	0.073 ± 0.019	589	0.13	0.07/0.02
	RIIR	0.36 ± 0.36	35.46 ± 4.41	0.007 ± 0.002	0.312 ± 0.042	0.039 ± 0.008	148	0.24	0.13/0.05

Table 2
Comparison of Explicit Input and RIIR under 5% data availability in Experiment 3. For each metric, we report mean ± std.

Dataset	$\lvert J_{\phi}\rvert_{\leq 0}$ (%)	std(log $\lvert J_{\phi}\rvert$)	mag($\nabla\lvert J_{\phi}\rvert$)
OASIS			
Explicit	0.014 ± 0.015	0.299 ± 0.133	0.035 ± 0.003
RIIR	0.019 ± 0.012	0.350 ± 0.088	0.034 ± 0.003
NLST			
Explicit	0.092 ± 0.108	0.575 ± 0.372	0.057 ± 0.003
RIIR	0.085 ± 0.084	0.581 ± 0.309	0.055 ± 0.004

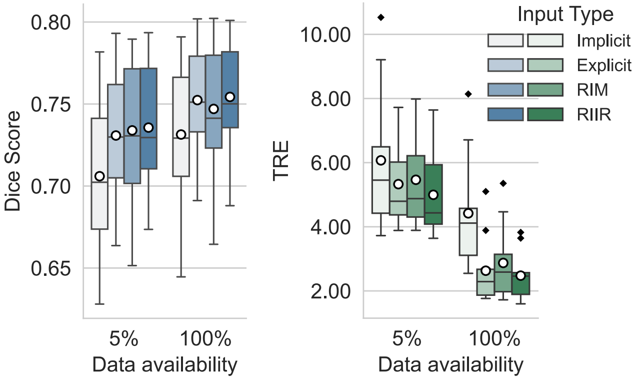

Fig. 10. Results of Experiment 3 evaluated on OASIS validation set (left) and NLST validation set (right). For OASIS with 5% data availability, RIIR input shows significance over all types except RIM ($p = 0.81$). At 100%, significance remains except for Explicit input ($p = 0.17$). For NLST with 5% data availability, RIIR input shows significance over all types. At 100%, significance remains except for Explicit input ($p = 0.27$).

over other input types. However, RIIR input demonstrates more robust performance in data-limited scenarios, particularly when only 5% of the training data is available. We also found that for both datasets when data availability is 5%, the deformation-related metrics have smaller variation for RIIR input compared with Explicit inputs as shown in Table 2.

4.4. Experiment 4: Regularization analysis

The results of regularization analysis on OASIS and NLST datasets are shown in Figs. 11 and 12, respectively. For OASIS, diffusion regularization achieves better Dice scores compared to affine registration across all weights. For NLST, while all three regularization terms improve the TRE over affine registration, curvature regularization exhibits a higher percentage of negative Jacobian determinants. For elastic regularization, it is worth noting that different parameter settings for λ_e and μ were used across datasets to reflect tissue differences, which may lead to varying regularization effects.

4.5. Experiment 5: RIIR architecture ablation

Here we demonstrate the model architecture ablation by showing the corresponding boxplots in Fig. 13. The increasing number of steps leads to proportionally higher VRAM consumption and inference time. Specifically, VRAM usage ranges from approximately 8 GB (4 steps) to 24 GB (12 steps), with corresponding inference times varying from 0.40 s to 1.12 s. Apart from the main experiments shown previously, this experiment can be regarded as a minor ablation study, aiming to strike a balance between computational precision and inference speed.

5. Discussion

In this study, we introduced RIIR, a deep learning-based medical image registration method that leverages recurrent inferences as a meta-learning strategy. We extended Recurrent Inference Machines (RIMs) to the image registration problem, which has no explicit forward models. Given the absence of explicit forward models, our approach can be viewed as a case of amortized iterative inference, where the network learns to progressively refine the registration. RIIR was extensively evaluated on public brain MR, lung CT, and in-house quantitative

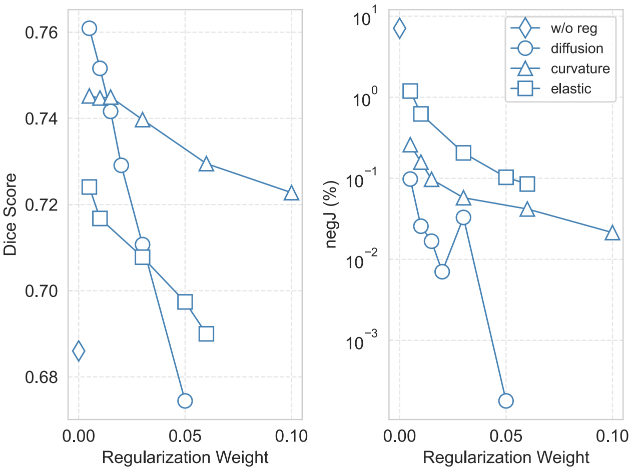

Fig. 11. Results of regularization analysis on OASIS validation set. Different regularization weights are compared in terms of Dice score. Here w/o reg denotes the results without regularization and negJ denotes the percentage of negative Jacobian determinant.

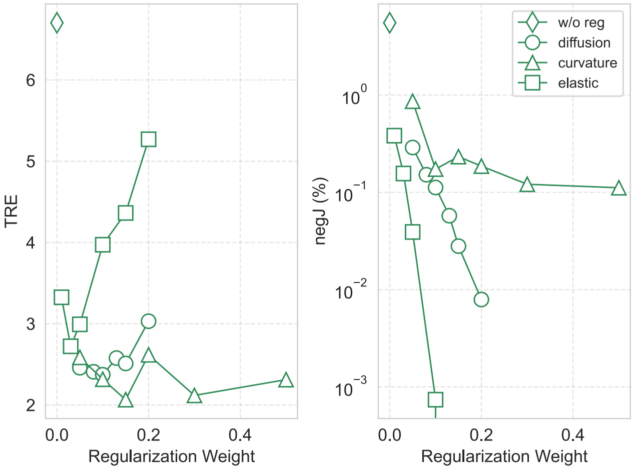

Fig. 12. Results of regularization analysis on NLST validation set. Different regularization weights are compared in terms of TRE. Here w/o reg denotes the results without regularization and negJ denotes the percentage of negative Jacobian determinant.

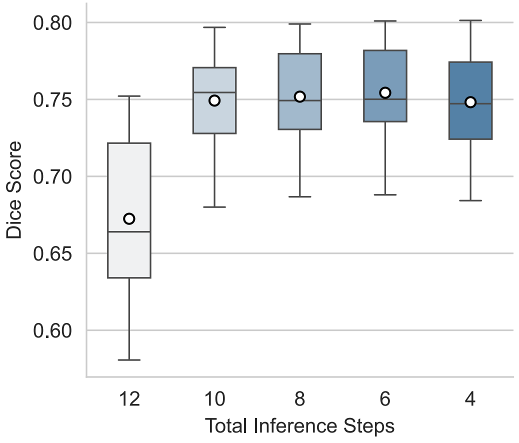

Fig. 13. Results of Experiment 5. For the ablation study on network steps on OASIS, two-sided Wilcoxon tests suggest significant difference is found for t = 6 against all other scenarios.

cardiac MR datasets, and demonstrated consistently improved performance over established deep learning models, both in one-step and iterative settings. Additionally, our ablation study confirmed the importance of incorporating hidden states within the RIM-based framework.

The acclaimed improvement in registration performance is especially pronounced in scenarios with limited training data, as demonstrated in Figs. 6, 7, and 8. In particular, RIIR achieved superior average evaluation metrics with lower variance. This performance advantage was particularly pronounced in the NLST experiments, where large deformations need to be estimated with very limited training data. Most baseline models required substantially more training data to achieve comparable performance levels, with the exception of L2O, which benefits from additional input modalities such as dynamically sampled coordinates and MIND features. Though both GraDIRN and RIIR are iterative methods that use similarity gradients, GraDIRN isolates the update of explicit $\mathcal{L}_{\text{sim}}$ and deep learning-based $\mathcal{L}_{\text{reg}}$ with no internal states, potentially resulting in worse generalization and slower convergence compared to RIIR. In particular, GraDIRN initialized the deformation field randomly by default, which could lead to optimization difficulties when the data were extremely limited during training. RCVM achieves lower folding rates ($|J_{\phi}|_{\leq 0}$) through its composed deformation strategy, where the final transformation is obtained by composing multiple smaller deformations. Although this approach effectively prevents topology-breaking deformations, it comes at the cost of an increased number of parameters. While the training and inference times on the OASIS dataset are the second longest among learning-based methods (after L2O), RCVM exhibits comparable runtime with RIIR for both 3D datasets, and shows faster inference time on the 2D dataset (mSASHA). Notably, LapIRN, as a single-pass multi-resolution method, shows remarkable advantages in both VRAM efficiency and inference speed.

Although hidden states were used in the original RIM and later work (Putzky and Welling, 2017; Putzky et al., 2019; Sabidussi et al., 2021, 2023), their impact on the optimization of RIM-based methods has not been investigated in detail. Our second experiment investigates the impact of these hidden states within RIIR. Our findings reveal that the presence of hidden states, as proposed in the original RIM work (Putzky and Welling, 2017), contributes positively to the performance of our model, as shown by the quantitative results in Fig. 9. Unlike L2O which operates on full resolution features at the output, our implementation of hidden states at downsampled resolution leads to minimal additional memory overhead.

The ablation study in input combinations (Experiment 3) demonstrates that, in scenarios with limited data, RIIR with images gradient input achieves superior registration performance in anatomical evaluation metrics, as shown in Fig. 10. This improvement is particularly evident in the NLST dataset, where RIIR achieves lower TRE and smoother deformation fields. Although including gradient input can be considered to offer additional information, its impact on regularization varies between datasets (Table 2). This could possibly be due to the different similarity losses and registration objectives (intra-subject for NLST versus inter-subject for OASIS).

The regularization analysis reveals that RIIR's performance remains dependent on both the choice and weight of regularization terms. As shown in Figs. 11 and 12, different regularization strategies lead to varying trade-offs between registration accuracy and deformation regularity. This suggests potential future improvements by incorporating adaptive regularization schemes (Hoopes et al., 2021; Mok and Chung, 2021; Reithmeir et al., 2024). Such extensions could enhance RIIR's robustness across different clinical scenarios without manual parameter tuning.

The superior registration performance and data efficiency of RIIR suggest its potential for applications in medical image registration. However, it is necessary to acknowledge the current limitations, to further enhance the framework in future work. From an architectural perspective, RIIR employs a relatively conventional design that lacks multi-resolution capability, which has proven effective in methods like LapIRN. The simple convolutional GRU structure could also be enhanced with modern components such as dilated convolutions for larger receptive fields or attention mechanisms for better feature extraction. Another significant limitation lies in GPU memory consumption. Despite having fewer parameters than LapIRN and RCVM, RIIR requires more VRAM due to its recurrent nature, and this memory usage increases linearly with the number of inference steps, as demonstrated in Experiment 5. To better adapt to downstream tasks, potential improvements could include semi-supervised strategy, instance optimization and adaptive regularization strategies, which could enhance the flexibility of the model in different clinical scenarios.

6. Conclusion

In conclusion, we present RIIR, a novel recurrent deep-learning framework for medical image registration. RIIR significantly extends the concept of recurrent inference machines for inverse problem solving, to high-dimensional optimization challenges with no closed-form forward models. Meanwhile, RIIR distinguishes itself from previous iterative methods by integrating implicit regularization with explicit loss gradients. Our experiments across diverse medical image datasets demonstrated RIIR's superior accuracy and data efficiency. We also empirically demonstrated the effectiveness of its architectural design and the value of hidden states, significantly enhancing both registration accuracy and data efficiency. RIIR is shown to be an effective and generalizable tool for medical image registration, and potentially extends to other high-dimensional optimization problems.

CRediT authorship contribution statement

Yi Zhang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yidong Zhao:** Writing – review & editing, Visualization, Validation, Methodology, Conceptualization. **Hui Xue:** Writing – review & editing, Validation, Data curation. **Peter Kellman:** Writing – review & editing, Supervision, Resources, Funding acquisition, Data curation. **Stefan Klein:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Qian Tao:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Acknowledgments

This work was supported by the TU Delft AI Initiative, Amazon Research Awards, the Dutch Research Council (NWO), and the National Heart, Lung, and Blood Institute, National Institutes of Health by the Division of Intramural Research (Z1A-HL006214).

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103748.

References

Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M., Sicks, J., et al., 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. New Engl. J. Med. 365, 395– 409.

Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., De Freitas, N., 2016. Learning to learn by gradient descent by gradient descent. Adv. Neural Inf. Process. Syst. 29.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ants similarity metric performance in brain image registration. NeuroImage 54, 2033–2044.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38, 1788–1800.

Byrne, M., Archibald-Heeren, B., Hu, Y., Teh, A., Beserminji, R., Cai, E., Liu, G., Yates, A., Rijken, J., Collett, N., Aland, T., 2022. Varian ethos online adaptive radiotherapy for prostate cancer: Early results of contouring accuracy, treatment plan quality, and treatment time. J. Appl. Clin. Med. Phys. 23, e13479. http://dx.doi.org/10.1002/acm2.13479.

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., Shen, D., 2017. Deformable image registration based on similarity-steered cnn regression. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September (2017) 11-13, Proceedings, Part I 20. Springer, pp. 300–308.

Chen, Y., Hoffman, M.W., Colmenarejo, S.G., Denil, M., Lillicrap, T.P., Botvinick, M., Freitas, N., 2017. Learning to learn without gradient descent by gradient descent. In: International Conference on Machine Learning. PMLR, pp. 748–756.

Chow, K., Hayes, G., Flewitt, J.A., Feuchter, P., Lydell, C., Howarth, A., Pagano, J.J., Thompson, R.B., Kellman, P., White, J.A., 2022. Improved accuracy and precision with three-parameter simultaneous myocardial t1 and t2 mapping using multiparametric sasha. Magn. Reson. Med. 87, 2775–2791.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Cotter, N.E., Conwell, P.R., 1990. Fixed-weight networks can learn. In: International Joint Conference on Neural Networks. IEEE, pp. 553–559.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2019. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med. Image Anal. 57, 226–236.

De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal. 52, 128–143.

de Vos, B.D., van der Velden, B.H., Sander, J., Gilhuijs, K.G., Staring, M., Išgum, I., 2020. Mutual information for unsupervised deep learning image registration. In: Medical Imaging 2020: Image Processing. SPIE, pp. 155–161.

Falta, F., Hansen, L., Heinrich, M.P., 2022. Learning iterative optimisation for deformable image registration of lung ct with recurrent convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 301–309.

Fechter, T., Baltas, D., 2020. One-shot learning for deformable medical image registration and periodic motion tracking. IEEE Trans. Med. Imaging 39, 2506–2517.

Finn, C., Abbeel, P., Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning. PMLR, pp. 1126–1135.

Fischer, B., Modersitzki, J., 2002. Fast diffusion registration. Contemp. Math. 313, 117–128.

Fischer, B., Modersitzki, J., 2004. A unified approach to fast image registration and a new curvature based registration technique. Linear Algebra Appl. 380, 107–124.

Haskins, G., Kruger, U., Yan, P., 2020. Deep learning in medical image registration: a survey. Mach. Vis. Appl. 31, 1–18.

Heinrich, M.P., Handels, H., Simpson, I.J., 2015. Estimating large lung motion in copd patients by symmetric regularised correspondence fields. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October (2015) 5-9, Proceedings, Part II 18. Springer, pp. 338–345.

Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A., 2012. Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. 16, 1423–1435.

Hering, A., van Ginneken, B., Heldmann, S., 2019. Mlvirnet: Multilevel variational image registration network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October (2019) 13–17, Proceedings, Part VI 22. Springer, pp. 257–265.

Hering, A., Hansen, L., Mok, T.C., Chung, A.C., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al., 2022. Learn2reg: comprehensive multitask medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Trans. Med. Imaging 42, 697–712.

Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V., 2021. Hypermorph: Amortized hyperparameter learning for image registration. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30 2021, Proceedings 27. Springer, pp. 3–17.

Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A., 2021. Meta-learning in neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44, 5149–5169.

Huizinga, W., Poot, D., Guyader, J.M., Klaassen, R., Coolen, B., van Kranenburg, M., van Geuns, R., Uitterdijk, A., Polfliet, M., Vandemeulebroucke, J., Leemans, A., Niessen, W., Klein, S., 2016. Pca-based groupwise image registration for quantitative mri. Med. Image Anal. 29, 65–78. http://dx.doi.org/10.1016/j.media.2015.12.004, URL https://www.sciencedirect.com/science/article/pii/S1361841515001851.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 203–211.

Jin, C., Yu, H., Ke, J., Ding, P., Yi, Y., Jiang, X., Duan, X., Tang, J., Chang, D.T., Wu, X., Gao, F., Li, R., 2021. Predicting treatment response from longitudinal images using multi-task deep learning. Nat. Commun. 12, 1851.

Kanter, F., Lellmann, J., 2022. A flexible meta learning model for image registration. In: Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., Albarqouni, S. (Eds.), Proceedings of the 5th Medical Imaging with Deep Learning. PMLR, pp. 638–652, URL https://proceedings.mlr.press/v172/kanter22a.html.

Karkalousos, D., Noteboom, S., Hulst, H.E., Vos, F.M., Caan, M.W., 2022. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated mri reconstruction. Phys. Med. Biol. 67, 124001.

King, A.P., Rhode, K.S., Ma, Y., Yao, C., Jansen, C., Razavi, R., Penney, G.P., 2010. Registering preprocedure volumetric images with intraprocedure 3-d ultrasound using an ultrasound imaging model. IEEE Trans. Med. Imaging 29, 924–937.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May (2015) 7-9, Conference Track Proceedings. URL http://arxiv.org/abs/1412.6980.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205.

Klein, S., Staring, M., Pluim, J.P., 2007. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. IEEE Trans. Image Process. 16, 2879–2890.

Kumaresan, S., Radhakrishnan, S., 1996. Importance of partitioning membranes of the brain and the influence of the neck in head injury modelling. Med. Biol. Eng. Comput. 34, 27–32.

Lai-Fook, S.J., Hyatt, R.E., 2000. Effects of age on elastic moduli of human lungs. J. Appl. Physiol. 89, 163–168.

Liu, R., Li, Z., Fan, X., Zhao, C., Huang, H., Luo, Z., 2021. Learning deformable image registration from optimization: perspective, modules, bilevel training and beyond. IEEE Trans. Pattern Anal. Mach. Intell. 44, 7688–7704.

Lønning, K., Putzky, P., Sonke, J.J., Reneman, L., Caan, M.W., Welling, M., 2019. Recurrent inference machines for reconstructing heterogeneous mri data. Med. Image Anal. 53, 64–78.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19, 1498–1507.

Marino, J., Yue, Y., Mandt, S., 2018. Iterative amortized inference. In: International Conference on Machine Learning. PMLR, pp. 3403–3412.

Messroghli, D.R., Radjenovic, A., Kozerke, S., Higgins, D.M., Sivananthan, M.U., Ridgway, J.P., 2004. Modified look-locker inversion recovery (molli) for high-resolution t1 mapping of the heart. Magn. Reson. Med. 52, 141–146.

Miao, S., Wang, Z.J., Liao, R., 2016. A cnn regression approach for real-time 2d/3d registration. IEEE Trans. Med. Imaging 35, 1352–1363.

Modi, C., Lanusse, F., Seljak, U., Spergel, D.N., Perreault-Levasseur, L., 2021. Cosmicrim: reconstructing early universe by combining differentiable simulations with recurrent inference machines. arXiv preprint arXiv:2104.12864.

Mok, T.C., Chung, A.C., 2020. Large deformation diffeomorphic image registration with laplacian pyramid networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October (2020) 4–8, Proceedings, Part III 23. Springer, pp. 211–221.

Mok, T.C., Chung, A.C., 2021. Conditional deformable image registration with convolutional neural network. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part IV 24. Springer, pp. 35–45.

Morningstar, W.R., Levasseur, L.P., Hezaveh, Y.D., Blandford, R., Marshall, P., Putzky, P., Rueter, T.D., Wechsler, R., Welling, M., 2019. Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines. Astrophys. J. 883, 14. http://dx.doi.org/10.1109/TIP.2007.909412.

Muckley, M.J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Shin, H., Jun, Y., Hwang, D., Mostapha, M., Arberet, S., Nickel, D., Ramzi, Z., Ciuciu, P., Starck, J.L., Teuwen, J., Karkalousos, D., Zhang, C., Sriram, A., Huang, Z., Yakubova, N., Lui, Y.W., Knoll, F., 2021. Results of the 2020 fastmri challenge for machine learning mr image reconstruction. IEEE Trans. Med. Imaging 40, 2306–2317. http://dx.doi.org/10.1109/TMI.2021.3075856.

Ntatsis, K., Dekker, N., Valk, V.Van.Der., Birdsong, T., Zukiczukic, D., Klein, S., Staring, M., Mccormick, M., 2023. Itk-elastix: Medical image registration in python. In: Proceedings of the 22nd Python in Science Conference. pp. 101–105.

Oliveira, F.P., Tavares, J.M.R., 2014. Medical image registration: a review. Comput. Methods Biomech. Biomed. Eng. 17, 73–93.

Putzky, P., Karkalousos, D., Teuwen, J., Miriakov, N., Bakker, B., Caan, M., Welling, M., 2019. I-rim applied to the fastmri challenge. arXiv preprint arXiv:1910.08952.

Putzky, P., Welling, M., 2017. Recurrent inference machines for solving inverse problems. arXiv preprint arXiv:1706.04008.

Qiu, H., Hammernik, K., Qin, C., Chen, C., Rueckert, D., 2022. Embedding gradient-based optimization in image registration networks. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2022. Springer, pp. 56–65.

Qiu, H., Qin, C., Hammernik, K., Schuh, A., Rueckert, D., 2021. Learning diffeomorphic and modality-invariant registration using b-splines. In: Medical Imaging with Deep Learning.

Reithmeir, A., Schnabel, J.A., Zimmer, V.A., 2024. Learning physics-inspired regularization for medical image registration with hypernetworks. In: Medical Imaging 2024: Image Processing. SPIE, pp. 625–635.

Rohé, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. Svf-net: learning deformable image registration using shape matching. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September (2017) 11-13, Proceedings, Part I 20. Springer, pp. 266–274.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October (2015) 5-9, Proceedings, Part III 18. Springer, pp. 234–241.

Rueckert, D., Schnabel, J.A., 2019. Model-based and data-driven strategies in medical image computing. Proc. IEEE 108, 110–124.

Sabidussi, E.R., Klein, S., Caan, M.W., Bazrafkan, S., den Dekker, A.J., Sijbers, J., Niessen, W.J., Poot, D.H., 2021. Recurrent inference machines as inverse problem solvers for mr relaxometry. Med. Image Anal. 74, 102220.

Sabidussi, E., Klein, S., Jeurissen, B., Poot, D., 2023. DtiRIM: A generalisable deep learning method for diffusion tensor imaging. NeuroImage 269, 119900.

Sandkühler, R., Andermatt, S., Bauman, G., Nyilas, S., Jud, C., Cattin, P.C., 2019. Recurrent registration neural networks for deformable image registration. Adv. Neural Inf. Process. Syst. 32.

Sauer, F., 2006. Image registration: enabling technology for image guided surgery and therapy. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, pp. 7242–7245.

Schmidhuber, J., 1993. A neural network that embeds its own meta-levels. In: IEEE International Conference on Neural Networks. IEEE, pp. 407–412.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 28.

Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable medical image registration: A survey. IEEE Trans. Med. Imaging 32, 1153–1190.

Staring, M., van der Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P., 2009. Registration of cervical mri using multifeature mutual information. IEEE Trans. Med. Imaging 28, 1412–1421.

Studholme, C., Hill, D.L., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3d medical image alignment. Pattern Recognit. 32, 71–86.

Thévenaz, P., Unser, M., 2000. Optimization of mutual information for multiresolution image registration. IEEE Trans. Image Process. 9, 2083–2099.

van Harten, L., Van Herten, R.L.M., Stoker, J., Isgum, I., 2023. Deformable image registration with geometry-informed implicit neural representations. In: Medical Imaging with Deep Learning.

Wolterink, J.M., Zwienenberg, J.C., Brune, C., 2022. Implicit neural representations for deformable image registration. In: Medical Imaging with Deep Learning. PMLR, pp. 1349–1359.

Xu, J., Chen, E.Z., Chen, X., Chen, T., Sun, S., 2021. Multi-scale neural odes for 3d medical image registration. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part IV 24. Springer, pp. 213–223.

Yang, X., Kwitt, R., Niethammer, M., 2016. Fast predictive image registration. In: Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21 2016, Proceedings 1. Springer, pp. 48–57.

Yang, X., Kwitt, R., Styner, M., Niethammer, M., 2017. Quicksilver: Fast predictive image registration – a deep learning approach. NeuroImage 158, 378–396. http://dx.doi.org/10.1016/j.NeuroImage.2017.07.008.

Younger, A.S., Conwell, P.R., Cotter, N.E., 1999. Fixed-weight on-line learning. IEEE Trans. Neural Netw. 10, 272–283.

Zbontar, J., Knoll, F., Sriram, A., Muckley, M.J., Bruno, M., Defazio, A., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M.G., Vincent, P., Pinkerton, J., Wang, D., Yakubova, N., Owens, E., Zitnick, C.L., Recht, M.P., Sodickson, D.K., Lui, Y.W., 2018. Fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839.

Zhang, Y., Pei, Y., Zha, H., 2021. Learning dual transformer network for diffeomorphic registration. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1 2021, Proceedings, Part IV 24. Springer, pp. 129–138.

Zhao, S., Dong, Y., Chang, E.I., Xu, Y., 2019. Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10600–10610.