Medical Image Analysis 110 (2026) 103987

Contents lists available at ScienceDirect

Medical Image Analysis

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

Deformation-Recovery diffusion model (DRDM): Instance deformation for image manipulation and synthesis

Jian-Qing Zheng [a,b,1,*], Yuanhan Mo [c,d,1], Yang Sun [c], Jiahua Li [c], Fuping Wu [c], Ziyang Wang [e], Tonia Vincent [a], Bartłomiej W Papież [c]

[a] The Kennedy Institute of Rheumatology, University of Oxford, UK
[b] Chinese Academy of Medical Sciences Oxford Institute, University of Oxford, UK
[c] Big Data Institute, University of Oxford, UK
[d] MRC Laboratory of Medical Sciences, Imperial College London, UK
[e] Department of Computer Science, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:
Image synthesis
Generative model
Data augmentation
Segmentation
Image registration

ABSTRACT

In medical imaging, diffusion models have shown great potential for synthetic image generation. However, these approaches often lack interpretable correspondence between generated and real images and can create anatomically implausible structures or illusions. To address these limitations, we propose the Deformation-Recovery Diffusion Model (DRDM), a novel diffusion-based generative model that emphasizes morphological transformation through deformation fields rather than direct image synthesis. DRDM introduces a topology-preserving deformation field generation strategy, which randomly samples and integrates multi-scale Deformation Velocity Fields (DVFs). DRDM is trained to learn to recover unrealistic deformation components, thus restoring randomly deformed images to a realistic distribution. This formulation enables the generation of diverse yet anatomically plausible deformations that preserve structural integrity, thereby improving data augmentation and synthesis for downstream tasks such as few-shot learning and image registration. Experiments on cardiac Magnetic Resonance Imaging and pulmonary Computed Tomography show that DRDM is capable of creating diverse, large-scale deformations, while maintaining anatomical plausibility of deformation fields. Additional evaluations on 2D image segmentation and 3D image registration tasks indicate notable performance gains, underscoring DRDM's potential to enhance both image manipulation and generative modeling in medical imaging applications.

The project page: https://jianqingzheng.github.io/def_diff_rec/.

1. Introduction

Image synthesis, a rapidly advancing domain within Artificial Intelligence (AI), has been revolutionized by the advent of deep learning technologies (Zhan et al., 2023). It involves generating new images either from existing data or from random noise, guided by specific structural, appearance, semantic, or statistical constraints. Deep learning, with its ability to learn hierarchical representations (LeCun et al., 2015), has become the cornerstone of advancements in image synthesis, enabling applications that range from artistic image generation to the creation of realistic training data for machine learning models.

The heart of image synthesis via deep learning stems from the neural networks' ability to model and manipulate complex data distributions. Generative models, particularly Variational Autoencoders (VAEs) (Kingma and Welling, 2013) and Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), have emerged as powerful tools for this purpose. VAEs learn a compact latent space representation, enabling the generation of new images by sampling from this space. GANs, in contrast, consist of a generator that synthesizes images and a discriminator that evaluates their realism. The training process continues until the system reaches a Nash equilibrium, where the generator produces realistic images that the discriminator can no longer distinguish from real ones (Goodfellow et al., 2014).

Recently, intensity-based diffusion models, specifically Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), have achieved state-of-the-art performance in image generation across

* Corresponding author.
E-mail addresses: jianqing.zheng@ndm.ox.ac.uk (J.-Q. Zheng), y.mo16@lms.mrc.ac.uk (Y. Mo), yang.sun@bdi.ox.ac.uk (Y. Sun), jiahua.li@reuben.ox.ac.uk (J. Li), fuping.wu@ndph.ox.ac.uk (F. Wu), ziyang.wang@cs.ox.ac.uk (Z. Wang), tonia.vincent@kennedy.ox.ac.uk (T. Vincent), bartlomiej.papiez@bdi.ox.ac.uk (B.W Papież).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.media.2026.103987
Received 10 July 2024; Received in revised form 24 November 2025; Accepted 7 February 2026
Available online 11 February 2026

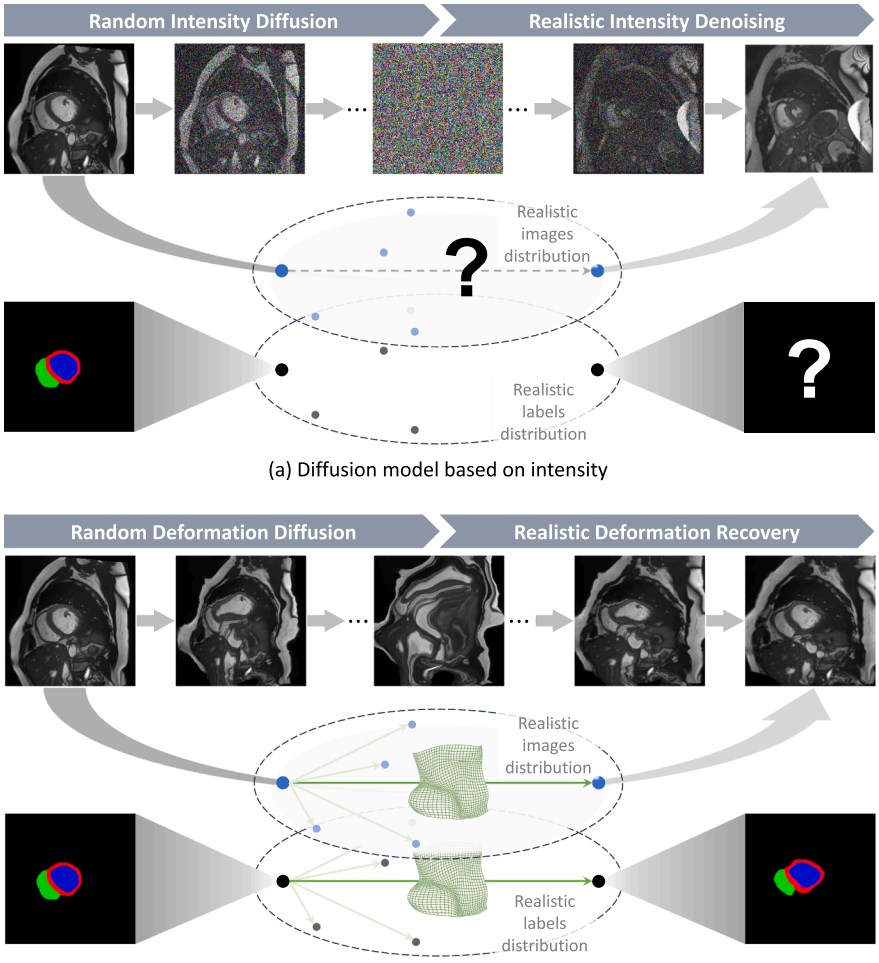

Fig. 1. (a) Intensity-based diffusion models can synthesize visually realistic images, but lack an explicit relationship with existing real subjects and thus unknown label relationship; (b) In contrast, the proposed Deformation-Recovery diffusion model (DRDM) applies generated deformation fields to real images, representing anatomical variations. These deformations can also be propagated to pixel-wise labels, thus enhancing the utility of the generated data for downstream tasks.

various computer-vision tasks. These models can generate high-fidelity data and exhibit desirable properties such as scalability and stable training. Furthermore, latent diffusion models (Rombach et al., 2022) extend DDPM by operating in feature space, enabling multimodal data conditioning e.g. text-guided image generation.

In medical imaging, diffusion models have been utilized for diverse tasks including synthetic medical image generation (Pinaya et al., 2022; Ju and Zhou, 2024; Du et al., 2023; Bradbury et al., 2024), biomarker quantification (Gong et al., 2023), anomaly detection (Li et al., 2023; Bercea et al., 2023; Liang et al., 2023), image segmentation (Graf et al., 2023; Liang et al., 2023; Zhang et al., 2024, 2025) and image registration (Qin and Li, 2023; Gao et al., 2023). These methods are capable of generating highly lifelike images but still suffer from potential issues such as producing visually plausible yet unrealistic artifacts and the inability of generated images to establish meaningful and interpretable relationships with pre-existing images, as illustrated in Fig. 1(a). This limitation hinders their applicability in tasks, such as image segmentation, that require precise understanding and correlation with real data (Kazerouni et al., 2023; Deo et al., 2025).

Generating deformation fields rather than image intensities can address this issue by focusing on anatomical changes. Several previous works (Kim et al., 2022; Kim and Ye, 2022; Starck et al., 2024; Wu et al., 2025; Wang et al., 2025) have attempted to generate deformation fields using image registration frameworks combined with diffusion models. However, they still employ diffusion-denoising approaches based on intensities (Kim et al., 2022; Kim and Ye, 2022) or latent-feature (Starck et al., 2024; Wu et al., 2025; Wang et al., 2025), depending on registration frameworks to guide and constrain the rationality of the generated deformations. Consequently, the generated deformations are often limited to interpolating transformations between image pairs (Kim et al., 2022; Kim and Ye, 2022; Wu et al., 2025) or atlas-based deformation (Starck et al., 2024; Wang et al., 2025), which restricts the ability to generate diverse and instance-specific morphological variations.

The noise added to image intensities in the existing diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020) is independently distributed across pixels and typically follows a normal distribution. However, our objective is to deform an existing image rather than generate a new one, and deformation vectors across anatomical regions are inherently correlated. Therefore, it is necessary to model deformation fields explicitly and design a diffusion process that reflects their realistic, spatially dependent distribution.

In this work, we propose a novel diffusion generative model, centred on deformation rather than intensity, termed the Deformation-Recovery Diffusion model (DRDM). DRDM represents a deformation-centric analogue of the conventional noising and denoising process, designed to achieve realistic and diverse anatomical variations as shown in Fig. 1(b). As illustrated in Fig. 2, the framework includes two stages: a random deformation diffusion process, followed by realistic deformation recovery, which collectively enable the generation of diverse deformations for individual images. Our main contributions are summarized as follows.

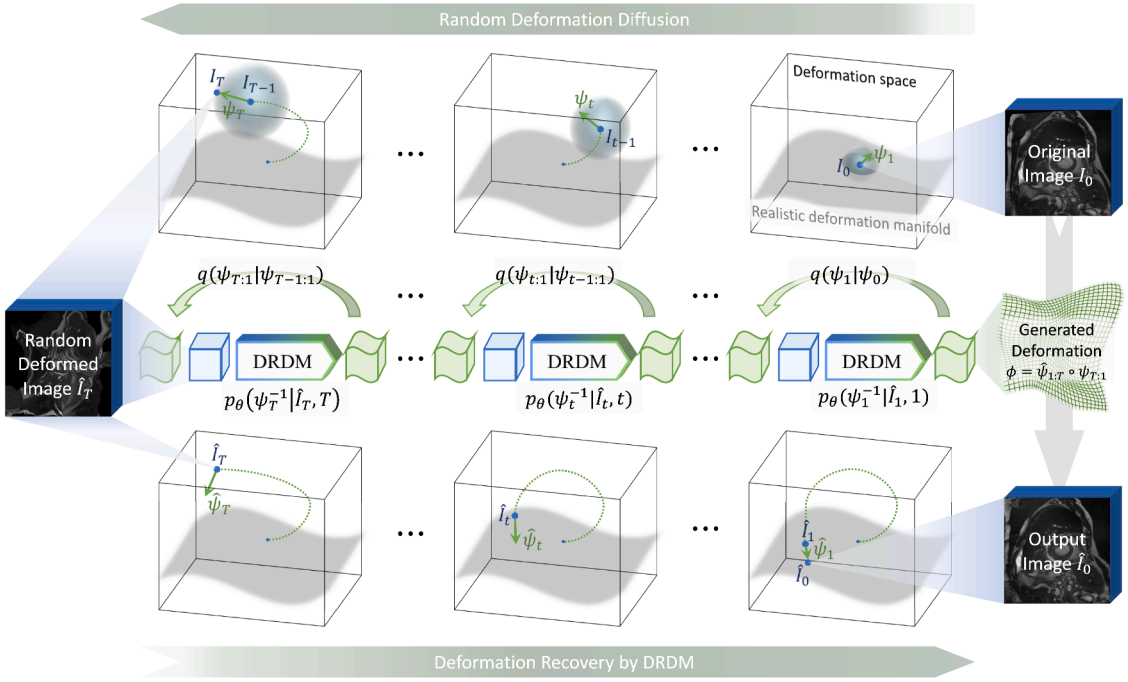

Fig. 2. Overview of the DRDM framework. The model comprises two stages: (i) a deformation diffusion process that applies random deformations within the deformation space, and (ii) a deformation recovery process that recursively estimates and refines deformation field to generate a anatomically realistic image within the learned deformation manifold.

- **Instance-specific deformation synthesis**: To the best of our knowledge, this is the first study to explore diverse deformation generation for individual images without utilizing any atlas/reference image or relying on population-level structural distributions (He et al., 2024; Starck et al., 2024; Wu et al., 2025; Wang et al., 2025);
- **Deformation Diffusion framework**: We propose a novel diffusion model method based on deformation diffusion and recovery in contrast to intensity-based diffusion (Kim et al., 2022; Kim and Ye, 2022) or latent-feature diffusion (Starck et al., 2024; Wu et al., 2025; Wang et al., 2025) based on registration frameworks;
- **Multi-scale random Dense Velocity Field (DVF) sampling and integrating**: We propose a method for multi-scale random DVF sampling and integration to generate deformation fields with physically possible distributions for training DRDM;
- **Training from scratch without annotation**: DRDM is trained entirely from unlabelled images, without requiring human annotations or auxiliary (registration or optical/scene flow) model/framework;
- **Data augmentation for few-shot learning**: The deformation fields generated by DRDM can be applied on both image and corresponding pixel-level segmentation, to augment morphological information while preserving anatomical topology, thus enabling data augmentation for few-shot learning tasks;
- **Synthetic training for image registration**: The synthetic deformations created by DRDM can be used to train image registration models without the need for external annotations;
- **Improved performance on downstream tasks**: The experimental results show that data augmentation and synthesis by DRDM improve performance in the downstream tasks, including segmentation and registration. Specifically, the DRDM-augmented segmentation model outperforms the BigAug baselline (Zhang et al., 2020), and the DRDM-trained registration model surpasses prior synthetic training approaches (Eppenhof and Pluim, 2019), validating the anatomical plausibility and utility of the learned deformation fields.

The remainder of this paper is organized as follows. Section 2 presents the design of the DRDM framework. Section 3 details the experimental setup and describes the generation of images and deformation fields. The applications of DRDM to few-shot image segmentation and synthetic registration training are discussed in Sections 4 and 5, respectively. Related work is reviewed in Section 6, and concluding remarks are provided in Section 7.

2. Framework design of DRDM

The overall framework of the proposed DRDM is illustrated in Fig. 2. The deformation field generated by DRDM, represented as a dense displacement field Dense Displacement Field (DDF), is defined as a spatial transformation $\phi : \mathbb{R}^{H\times W\times D} \to \mathbb{R}^{H\times W\times D}$, where each element corresponds to a displacement vector denoted by $\phi[\boldsymbol{x}] \in \mathbb{R}^3$ at the voxel coordinate $\boldsymbol{x} \in \mathbb{Z}^3$ of an image $\boldsymbol{I} \in \mathbb{R}^{H\times W\times D}$, where $H, W, D \in \mathbb{Z}_+$ denote the image height, width, and thickness, respectively. To maintain notational simplicity, all formulations in the following sections are presented for the 3D case, though they can be readily adapted to 2D by removing the depth dimension.

The generation of plausible DDF ϕ by DRDM can be formulated as a composition of random deformation diffusion and deformation recovery processes:

$$\begin{aligned}\phi &= \hat{\phi}_{1:T} \circ \phi_{T:1} \\ &= \underbrace{\hat{\psi}_1 \circ \hat{\psi}_2 \cdots \hat{\psi}_T}_{\text{deformation recovery}} \circ \underbrace{\psi_T \circ \psi_{T-1} \cdots \psi_1}_{\text{deformation diffusion}}\end{aligned} \tag{1}$$

The first part of Eq. (1), random deformation diffusion, as described in Section 2.1, is to generate a DDF ($\phi_{t:1}$) through a fixed Markov process of random DVF generation and integration of the DVFs ($\psi_t, \cdots, \psi_1$):

$$\phi_{t:1} := \psi_t \circ \psi_{t-1} \cdots \psi_1 \sim q(\phi_{t:1} |\, \phi_{t-1:1}) \tag{2}$$

The second part of Eq. (1), deformation recovery, as described in Section 2.2, is to estimate the recovering DDF $\hat{\phi}_{t:T}$ with the inverse DVF for each step ψ_t^{-1} estimated as $\hat{\psi}_t$ recursively based on the input of the deformed image $\boldsymbol{I}_t$:

$$\begin{cases} \hat{\phi}_{t:T} := \hat{\psi}_t \circ \hat{\psi}_{t+1} \cdots \hat{\psi}_T \\ \hat{\psi}_t \sim p(\psi_t^{-1} |\, \hat{\boldsymbol{I}}_t, t) \end{cases} \tag{3}$$

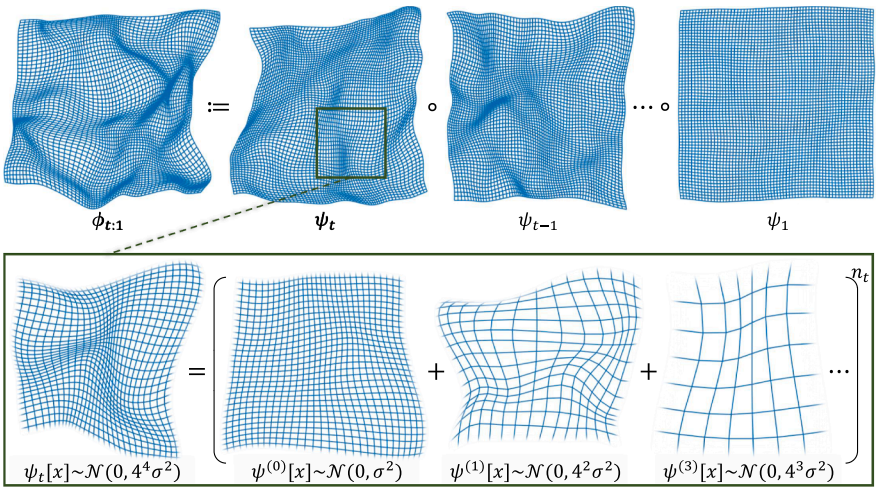

Fig. 3. Illustration of the principle underlying multi-scale random DVF generation and integration in the deformation diffusion process, as detailed in Eq. (2) and (6).

where $\boldsymbol{I}_0, \boldsymbol{I}_t$ denote the original image and the randomly deformed image respectively, $\hat{\boldsymbol{I}}_0$ and $\hat{\boldsymbol{I}}_t$ denote the synthesised image by DRDM and the intermediate recovered images, respectively. These are calculated as:

$$\begin{cases} \boldsymbol{I}_t := \phi_{t:1}(\boldsymbol{I}_0) \\ \hat{\boldsymbol{I}}_t := \langle \hat{\phi}_{t+1:T} \circ \phi_{T:1} \rangle (\boldsymbol{I}_0) \end{cases} \tag{4}$$

where $0 < t \leq T$ denotes the deformation step in diffusion or recovery processing, T denotes the total number of deformation steps for the diffusion and recovery process, and $\circ$ denotes the composition operation. This is calculated by:

$$\langle \phi_1 \circ \phi_2 \rangle[x] := \phi_2[\phi_1[x] + x] + \phi_1[x] \tag{5}$$

to simulate the process of gradual deformation (Arsigny et al., 2006; Vercauteren et al., 2009). Unlike intensity-based diffusion models (Ho et al., 2020; Song et al., 2020), which add and normalise independent noise components across time steps, the DRDM formulation explicitly models spatially correlated deformation compositions in the transformation manifold.

The transformation of images by a given deformation field and the composition between two deformation fields are implemented based on the Spatial Transformer Network (STN) (Jaderberg et al., 2015).

2.1. Forward process for random deformation diffusion

This section introduces the forward processing for random deformation diffusion as illustrated in Fig. 3. This process defines how physically plausible random deformations are generated within the DRDM framework for medical imaging applications. To ensure anatomical realism, the underlying assumptions and deformation constraints are introduced in Section 2.1.1. Based on these assumptions, Section 2.1.2 describes the random generation of noisy deformation velocity fields (DVFs). The procedure for determining the corresponding noise levels is presented in Section 2.1.3, and the integration of DVFs into dense displacement fields (DDFs) is detailed in Section 2.1.4.

2.1.1. The nature of deformation

As presented previously, the Gaussian noise applied to image intensities in conventional diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020) is independent for each pixel/voxel. In contrast, anatomical deformations exhibit spatial correlations, as neighboring regions of an organ typically move coherently. To ensure anatomically plausible transformations during the forward deformation diffusion process, a set of physical and topological constraints is defined to regularize the random deformation generation:

1. Randomness: The deformation vector of each position should yield a normal distribution $\psi_t[x] \sim \mathcal{N}(0, \sigma_t^2)$;
2. Local dependency: the deformation field of a continuum should be continuous and thus the stochastic regional discontinuity is bounded by $\Delta(\psi_t, \Delta x) := \psi_t[x + \Delta x] - \psi_t[x] \sim \mathcal{N}(0, \sigma_t'(\Delta x)^2)$, where $\sigma_t'(\Delta x_1) \geq \sigma_t'(\Delta x_2),\ \|\Delta x_1\|_\infty > \|\Delta x_2\|_\infty$;
3. Diffeomorphism: the generated deformation field of a continuum should preserve anatomical topology: $|\boldsymbol{J}|_{<0} < \epsilon$.

where Chebyshev distance $\|\cdot\|_\infty$ is used to measure spatial neighbourhood relationships, $|\boldsymbol{J}|_{<0}$ denotes the proportion of voxels with negative Jacobian determinant values of the deformation field, ϵ denotes a small positive value to constrain the unrealistic deformation, σ_t^2 denotes the deformation variance of DVF ψ_t at the t^{th} time step, and $\sigma_t'^2$ denotes the deformation discontinuity variance of DVF ψ_t.

These rules are primarily formulated for modeling the deformation of a single continuous structure. However, in cases involving discontinuous deformations across multiple organs or tissue interfaces, the situation becomes more complex, as previously discussed in (Papież et al., 2014; Zheng et al., 2024).

2.1.2. Multi-scale random DVF generation

Following the deformation constraints described in Section 2.1.1, a multi-scale random DVF is synthesized at each time step by sampling from multiple Gaussian distributions at different spatial scales:

$$\begin{cases} \psi = \psi^{(0)} + \text{intrp}(\psi^{(1)}) + \cdots + \text{intrp}(\psi^{(m)}) \\ \psi^{(0)} \in \mathbb{R}^{3\times h\times w\times d},\ \psi^{(0)}[x] \sim \mathcal{N}(0, \sigma^2) \\ \psi^{(1)} \in \mathbb{R}^{3\times (h/2)\times (w/2)\times (d/2)},\ \psi^{(1)}[x] \sim \mathcal{N}(0, (2\sigma)^2) \\ \cdots \\ \psi^{(m)} \in \mathbb{R}^{3\times (h/2^m)\times (w/2^m)\times (d/2^m)},\ \psi^{(m)}[x] \sim \mathcal{N}(0, (2^m\sigma)^2) \end{cases} \tag{6}$$

where $\text{intrp}(\cdot)$ denotes interpolation of the input image/DDF/DVF to the spatial resolution of $h \times w \times d$. The comoponents $\psi^{(0)}, \psi^{(1)}, \cdots \psi^{(m)}$ represent the independent DVF samples at different scales, namely the original scale, the first-order half-down-sampled scale, $\cdots$ and up to the m^{th}-order half-down-sampled scale. Under these definitions, the first constraint (randomness) described in Section 2.1.1 is satisfied when:

$$\sigma_t^2 \approx \frac{4^{m+1} - 1}{3} \sigma^2 n_t \tag{7}$$

and the second constraint (local dependency) is satisfied when:

$$\sigma_t'^2 \approx 2 n_t \sigma^2 \sum_{i=0}^{m} \min(\|\Delta x\|_\infty, 2^i)^2 \tag{8}$$

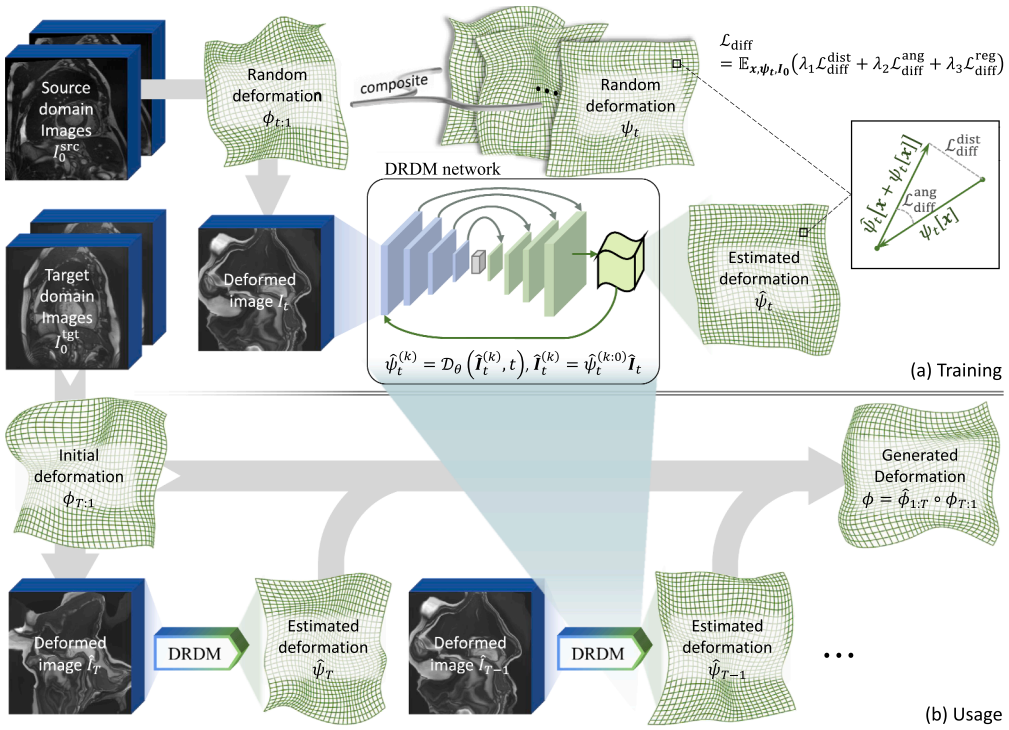

Fig. 4. (a) The DRDM is trained at each time step using distance- and angle-based loss function, as described in Algorithm 1. (b) During inference, deformation fields are generated by the DRDM with varying time step and integrated to produce the final deformation ϕ according to Algorithm 2.

where σ^2 denotes the minimum unit of DVF variance, n_t denotes the noise scale factor for each diffusion time step, as described in Section 2.1.3.

2.1.3. Noise scale of the random deformation field

To ensure the diffeomorphism of the generated deformation fields, the DDF is modeled as a pseudo flow, which can be differentiated into a DVF at each time step, following the continuum flow formulation (Christensen et al., 1996). For sampling of DVF with varying variance of magnitude at different time steps, an initial DVF is first sampled with a small fixed variance, and then integrated recursively to a larger DVF where the number of integration iterations n_t controls the overall deformation scale in the forward process. The integrating recursion number is used to control the magnitude of the random deformation field in the forward process. The noise scaling level is set to increase with the increasing time step t, expressed as:

$$n_t := t^{\alpha}/\beta \tag{9}$$

where α and β denote the parameters that control the rate of noise-level growth across diffusion steps.

2.1.4. Deformation diffusion by integrating DVF to DDF

As described in Eq. (2), the generated DVF in Section 2.1.2 is integrated to DDF $\phi_{t:1}$ by composing the sequence of deformation velocity fields $\psi_t,\ \psi_{t-1}, \cdots \psi_1$. Thus the random deformation field $\phi_{t:1}[x] \sim \mathcal{N}(0, \sigma^2_{t:1})$ can be sampled as:

$$\sigma^2_{t:1} := \sum_{i=1}^{t} \sigma_t^2 \approx \frac{4^{m+1}-1}{3} n_{t:1}\sigma^2 \tag{10}$$

where $n_{t:1}$ denotes the integrated noise scale factor, defined as:

$$n_{t:1} := \int_1^t n_\tau \mathrm{d}\tau \approx \frac{t^{\alpha+1}}{(\alpha+1)\beta} \tag{11}$$

2.2. Reverse process for deformation recovery

Unlike pixel-wise intensity prediction by DDPM or Denoising Diffusion Implicit Model (DDIM) (Ho et al., 2020; Song et al., 2020), the proposed DRDM is designed to estimate a deformation field. Fig. 4 illustrates the training and usage pipeline of the network for DRDM. The architecture of DRDM network is presented in Section 2.2.1, the training strategy is described in Section 2.2.2, and the application of the trained DRDM for instance-specific image deformation is described in Section 2.2.3.

2.2.1. Recursive network design for DRDM

As described in Eq. (3), the DVF used for deformation recovery is estimated and sampled by DRDM D_θ based on $\hat{\psi}_t \sim p(\psi_t^{-1} | \hat{\boldsymbol{I}}_t, t)$ with the input image $\hat{\boldsymbol{I}}_t$ at the time step t:

$$\begin{cases} \hat{\psi}_t^{(0)} : \hat{\boldsymbol{I}}_t \mapsto \hat{\boldsymbol{I}}_t \\ \hat{\psi}_t^{(k)} = \mathcal{D}_\theta(\hat{\boldsymbol{I}}_t^{(k)},\ t) \\ \hat{\boldsymbol{I}}_t^{(k)} = \langle \hat{\psi}_t^{(k-1)} \circ \hat{\psi}_t^{(k-2)} \cdots \hat{\psi}_t^{(0)} \rangle (\hat{\boldsymbol{I}}_t) \\ \hat{\psi}_t = \hat{\psi}_t^{(K)} \circ \hat{\psi}_t^{(K-1)} \cdots \hat{\psi}_t^{(1)} \end{cases} \tag{12}$$

where DRDM D_θ estimates a set of DVF $\hat{\psi}_t^{(k)}$ though the internal recursion $1 \le k \le K$ and integrates them to regress the inverse DVF ψ_t^{-1}.

The U-Net architecture (Ronneberger et al., 2015) is adapted into a recursive structure and combined with Atrous II blocks (Zhou et al., 2020) to enlarge the receptive field, thereby improving the network's ability to capture spatial context (Islam et al., 2019). The detailed network architecture is provided in A.

The internal recursion is designed to ensure that the network can adapt to each input deformed image in a single-step training strategy, and the number of internal iterations K is set to 2 following (Zheng et al., 2022).

Notably, in both Eqs. (4) and (12), multiple deformation fields are applied via compositional warping, rather repeatedly deforming the image ($\boldsymbol{I}_0$ and $\boldsymbol{I}_t^{(0)}$) itself, which prevents blurring and preserves fine anatomical structures.

Algorithm 1: Training DRDM.

Input: Training set of source domain images $\mathbf{D}^{\text{src}} \subset \mathbb{R}^{H\times W\times D}$
Output: DRDM weights θ

initialize the DRDM parameters θ;
while $\mathcal{L}_{\text{diff}}$ *not converge* **do**
  `// randomly sample the data`
  sample the original images: $\boldsymbol{I}_0 \in \mathbf{D}^{\text{src}}$;
  sample time steps: $t \sim \mathcal{U}(0, T) \cap \mathbb{Z}$;
  sample random DVFs ψ_t and DDFs $\phi_{t:1}$ according to (6), (7) and (10);
  `// compute the prediction and the loss`
  deform original images from $\boldsymbol{I}_0$ to $\boldsymbol{I}_t$ using (4);
  use DRDM D_θ to estimate recovering deformation $\hat{\psi}_t$ via (12);
  Update gradient descent step $\nabla_\theta \mathcal{L}_{\text{diff}}$ via (14);
end
return model weights θ.

Algorithm 2: Instance Deformation via DRDM.

Input: Images for deformation $\boldsymbol{I}_0 \in \mathbb{R}^{H\times W\times D}$
Output: Generated DDF ϕ

import the DRDM parameters θ from Algorithm 1;
set the deformation level $T' \le T$;
`// deformation diffusion process`
sample a random DDF $\phi_{T':1}$ using (6) and (10);
set the initial DDF for deformation recovery: $\phi \leftarrow \phi_{T':1}$;
deform original images from $\boldsymbol{I}_0$ to $\boldsymbol{I}_{T'}$ using (4);
set the initial image for deformation recovery: $\hat{\boldsymbol{I}}_{T'} \leftarrow \boldsymbol{I}_{T'}$;
`// deformation recovery process`
for $t = T', T'-1, \cdots, 1$ **do**
  use DRDM D_θ to estimate recovering deformation $\hat{\psi}_t$ via (12);
  update the deformation according to (1): $\phi \leftarrow \hat{\psi}_t \circ \phi$;
  deform original images from $\boldsymbol{I}_0$ to $\hat{\boldsymbol{I}}_{t-1}$ using (4);
end
return the generated deformation ϕ.

2.2.2. Network optimizing for DRDM

The DRDM D_θ is trained on randomly sampled time step $t \sim \mathcal{U}(0, T) \cap \mathbb{Z}$ where the trainable parameters θ are optimized by minimizing the following:

$$\min_\theta \{\mathcal{L}_{\text{diff}}(\psi_t, \hat{\psi}_t)\} \tag{13}$$

where the total loss function $\mathcal{L}_{\text{diff}}$ is defined by:

$$\begin{cases} \mathcal{L}_{\text{diff}} := \mathbb{E}_{x,\psi_t,\boldsymbol{I}_0}\left(\lambda_1 \mathcal{L}_{\text{diff}}^{\text{dist}} + \lambda_2 \mathcal{L}_{\text{diff}}^{\text{ang}} + \lambda_3 \mathcal{L}_{\text{diff}}^{\text{reg}}\right) \\ \mathcal{L}_{\text{diff}}^{\text{dist}} := \frac{\|(\psi_t \circ \hat{\psi}_t)[x]\|_2}{\|\psi_t[x]\|_2 + \epsilon} \\ \mathcal{L}_{\text{diff}}^{\text{ang}} := -\frac{\psi_t[x]^\top \hat{\psi}_t[x + \psi_t[x]]}{\|\psi_t[x]\|_2 \|\psi_t(\hat{\psi}_t[x])\|_2 + \epsilon} \\ \mathcal{L}_{\text{diff}}^{\text{reg}} := \|\nabla_x \hat{\psi}_t[x]\|_1 + \text{relu}(-\det(\nabla_x \hat{\psi}_t[x])) \end{cases} \tag{14}$$

Here, $\det(\cdot)$ denotes the determinant of a matrix, $\nabla\hat{\psi}_t$ denotes the Jacobian matrix of the estimated DVF. The total loss function for training the DRDM model consists of three terms as follows: (i) the distance error loss $\mathcal{L}_{\text{diff}}^{\text{dist}}$, which measures the magnitude difference between the true and estimated deformations; (ii) the angle error loss $\mathcal{L}_{\text{diff}}^{\text{ang}}$, which penalizes orientation discrepancies between the deformation vectors; and (iii) the regularization $\mathcal{L}_{\text{diff}}^{\text{reg}}$, which enforces spatial smoothness and penalizes non-diffeomorphic regions via the L1-norms and the negative determinant of the Jacobian. The relative importance of these components is controlled by the weighting factors $\lambda 1$, λ_2, and λ_3.

As shown in Algorithm 1, the weights of DRDM θ are optimized using a set of training images from the source domain ($\boldsymbol{I}_0 \in \mathbf{D}^{\text{src}} \subset \mathbb{R}^{H\times W\times D}$). The training procedure begins by initializing D_θ and iteratively sampling random time steps t from a uniform distribution. At each iteration, random DVFs (ψ_t) and DDFs ($\phi_{t:1}$) are generated according to the forward diffusion process. The original images $\boldsymbol{I}_0$ are then deformed into new states $\boldsymbol{I}_t$, and the network predicts the corresponding inverse deformation $\hat{\psi}_t$ required to recover the original image. The model parameters (θ) are updated through gradient descent to minimize the loss $\mathcal{L}_{\text{diff}}$, improving the model's deformation understanding and recovering capabilities. The training ends when the optimized weights (θ) are finalized upon convergence of the loss function.

2.2.3. Instance deformation synthesis by DRDM

After training the DRDM, the deformation field DDF ϕ is generated according to Algorithm 2, as shown in Fig. 4(b). The algorithm generates a DDF through composing a sequence of DVFs produced by the trained DRDM D_θ. Starting from an initial image from the target domain, represented as $\boldsymbol{I}_0 \in \mathbf{D}^{\text{tgt}}$, with the size of height H, width W, and depth D, the algorithm performs a series of steps to produce the final deformation.

Initially, a deformation step level T' is defined, not exceeding the maximum diffusion step level T. A random DDF $\phi_{T':1}$ is then sampled using the multi-scale DVF synthesis Eq. (6) and (10). This sampled DDF is set as the initial DDF ϕ for the subsequent deformation recovery process.

The input image $\boldsymbol{I}_0$ is first deformed to produce $\boldsymbol{I}_{T'}$, which serves as the initial state for deformation recovery, $\hat{\boldsymbol{I}}_{T'}$. The recovery process proceeds through reverse iteration from $t = T'$ down to 1. At each iteration, the DRDM network estimates a recovering deformation $\hat{\psi}_t$, which is used to update ϕ by integrating the current estimate with the accumulated deformations from previous steps.

Each iteration not only updates both the deformation field ϕ and the corresponding deformed image. Specifically, the deformation is applied to the original image $\boldsymbol{I}_0$, generating a progressively updated image state $\hat{\boldsymbol{I}}_{t-1}$ that transitions smoothly from $\boldsymbol{I}_{T'}$. The algorithm concludes by returning the fully integrated DDF ϕ, representing the cumulative deformation applied to the original image to reach its final deformed configuration.

This iterative estimation enables DRDM to model complex, nonlinear anatomical variations, with the total number of deformation steps T' controlling the overall deformation magnitude.

Table 1
Average ± standard deviation of the magnitude and the negative determinant ratio of Jacobian ($|\boldsymbol{J}|_{<0}$) for deformation fields genearted at varying deformation level T' in 2D cardiac MRI.

T' (-)	$\max_x \|\phi[x]\|_2$ (%img size)	$\text{avg}_x\|\phi[x]\|_2$ (%img size)	$\|\boldsymbol{J}\|_{<0}$ (‰)
30	7.8±1.3	2.4±0.6	0.7±1.5
45	10.4±2.4	3.0±0.8	0.9±1.2
55	11.8±3.2	3.3±0.8	2.4±3.2
65	12.7±4.0	3.6±0.9	6.6±5.5
70	14.8±3.5	4.3±1.4	9.8±7.2

3. Experiment of image deformation using DRDM

As shown in Fig. 5(a), a small number of 2D or 3D images are provided as inputs to the DRDM framework. The framework then generates corresponding deformed images, with or without labels, for downstream tasks as described in the following Section 4 and Section 5. For evaluation, the datasets are divided into a source domain and a target domain. The diffusion networks of DRDM are trained using the source domain data and then tested on the target-domain data for downstream tasks. The datasets used in the experimental implementation of DRDM

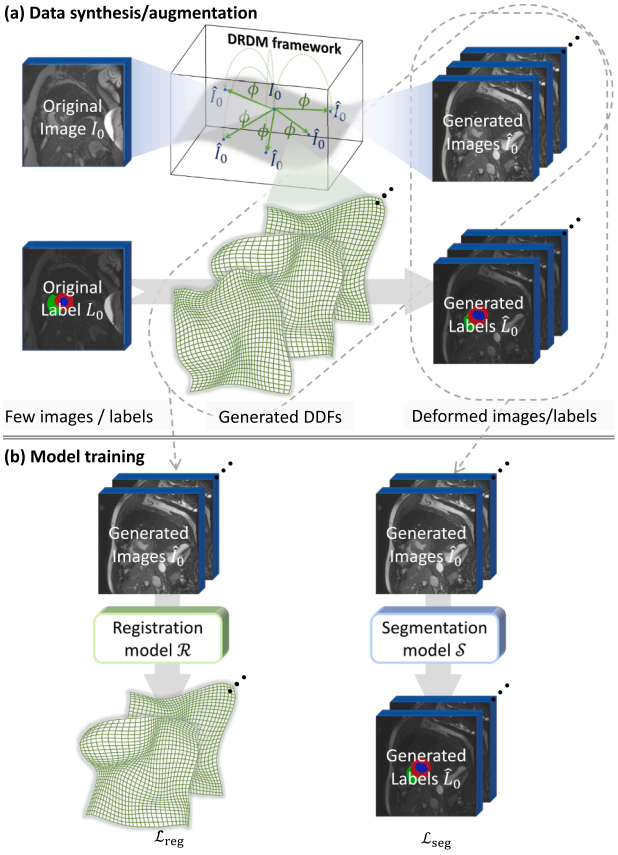

Fig. 5. Image and deformation synthesis using the proposed DRDM for few-shot-learning in image segmentation and image registration. (a) Diverse deformation fields, images, and corresponding labels are generated based on the input few images with labels, as described in Algorithm 3 and Algorithm 4; (b) The synthesized images and the corresponding labels are used to train a segmentation model, while the generated images and the corresponding DDFs are employed to train a registration model.

are described in Section 3.1, data preprocessing methods are explained in Section 3.2, the experimental setup is detailed in Section 3.3, and the evaluation of generated data in presented in Section 3.4.

3.1. Datasets

To evaluate the effectiveness of the proposed method, two imaging modalities were used: *cardiac MRI* and *thoracic CT*. The DRDM framework was trained independently on each modality and subsequently evaluated on both to assess its deformation generation and generalization performance.

3.1.1. Cardiac MRI

Four public cardiac Magnetic Resonance Imaging (MRI) datasets were used to construct the training set (source domain), and one additional dataset was employed for downstream evaluation (target domain) to assess the performance of the proposed DRDM framework. These datasets are as follows:

a. The Sunnybrook Cardiac Data (SCD) (Radau et al., 2009) comprises 45 cine-MRI images, representing a mix of subjects with various conditions including healthy individuals and patients with hypertrophy, heart failure with and without infarction.
b. Task-6 of the Medical Segmentation Decathlon released as part of the Left Atrial Segmentation Challenge (LASC) (Tobon-Gomez et al., 2015). It includes 30 3D MRI volumes.
c. The Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms dataset) (Campello et al., 2021) with 375 patients with hypertrophic and dilated cardiomyopathies, as well as healthy subjects.
d. Multi-Disease, Multi-View & Multi-Center Right Ventricular Segmentation in Cardiac MRI (M&Ms-2 dataset) (Martín-Isla et al., 2023) with 360 patients with various right ventricle and left ventricle pathologies, as well as healthy subjects.
e. Automated Cardiac Diagnosis Challenge (ACDC) dataset (Bernard et al., 2018), used for downstream evaluation in the whole-heart segmentation task, consists of 200 cardiac MRI cases, with 100 for training and 100 for testing.

The datasets (a)-(d) are used as source-domain data for training DRDM, while the dataset (e) served as the target-domain data for downstream validation in the segmentation task as described in Section 4.

3.1.2. Thoracic CT

Following a similar approach to Cardiac MRI, two publicly available Thoracic CT datasets from the *Cancer Imaging Archive*, were used to construct the training set (source domain), along with one additional dataset for downstream evaluation (target domain). These datasets are described as follows:

a. NSCLC-Radiomics (Version 4) (Aerts et al., 2015), which includes 422 patients diagnosed with non-small cell lung cancer (NSCLC).
b. QIN LUNG CT (Version 2) (Kalpathy-Cramer et al., 2015), which consists of 47 patients diagnosed with NSCLC at various stages and histologies.
c. The pulmonary Computer Tomography (CT) scans were provided by Hering et al. (2022) as part of the Learn2Reg 2021 challenge (task 2) dataset. These scans were consistently acquired at the same point within the breathing cycle to ensure uniformity. This dataset includes the inter-subject (exhale) registration task with 20 subjects for the training of a registration model and 10 for testing. Ground truth lung segmentations are also available for all scans.

The dataset (a) and (b) are used as source-domain data for training DRDM, and the dataset c is used as the target-domain data for downstream validation in the inter-subject registration task as described in Section 5.

3.2. Preprocessing and postprocessing

All images are first resized and padded to align with the isotropic resolution size of $H \times W \times D$. The images are then thresholded to remove irrelevant or non-anatomical regions, such as air cavity areas. Subsequently, image intensities were linearly normalized to the range [0, 1].

To enhance the robustness of the DRDM network, the images undergo several augmentations: rotation by a random angle $\sim \mathcal{U}(0°, 180°)$ around an arbitrary axis, translation by a random distance $\sim \mathcal{U}(-1/8, 1/8)$ of the image size along each of the three dimensions, randomly flipping with a probability of 0.5, and cropping with a ratio sampled from $\sim \mathcal{U}(0.6, 1.0)$.

After deformation fields ϕ were generated by DRDM, they were resized to the required resolution $\tilde{\phi} \in \mathbb{R}^{H' \times W' \times D'}$, ensuring compatibility with the image and label dimensions required for downstream tasks.

3.3. Experimental setting for DRDM

During the implementation of the random deformation diffusing process, thevectors in DDFs extending beyond the field boundary could introduce numerical friction that limits the increase in deformation magnitude. This occurs because the vectors outside the sampled region stop accumulating values when using the "zero" padding mode. To mitigate this issue, a larger deformation field was defined such that $h > H$, $w > W$, $d > D$, and then the desired deformation field was cropped

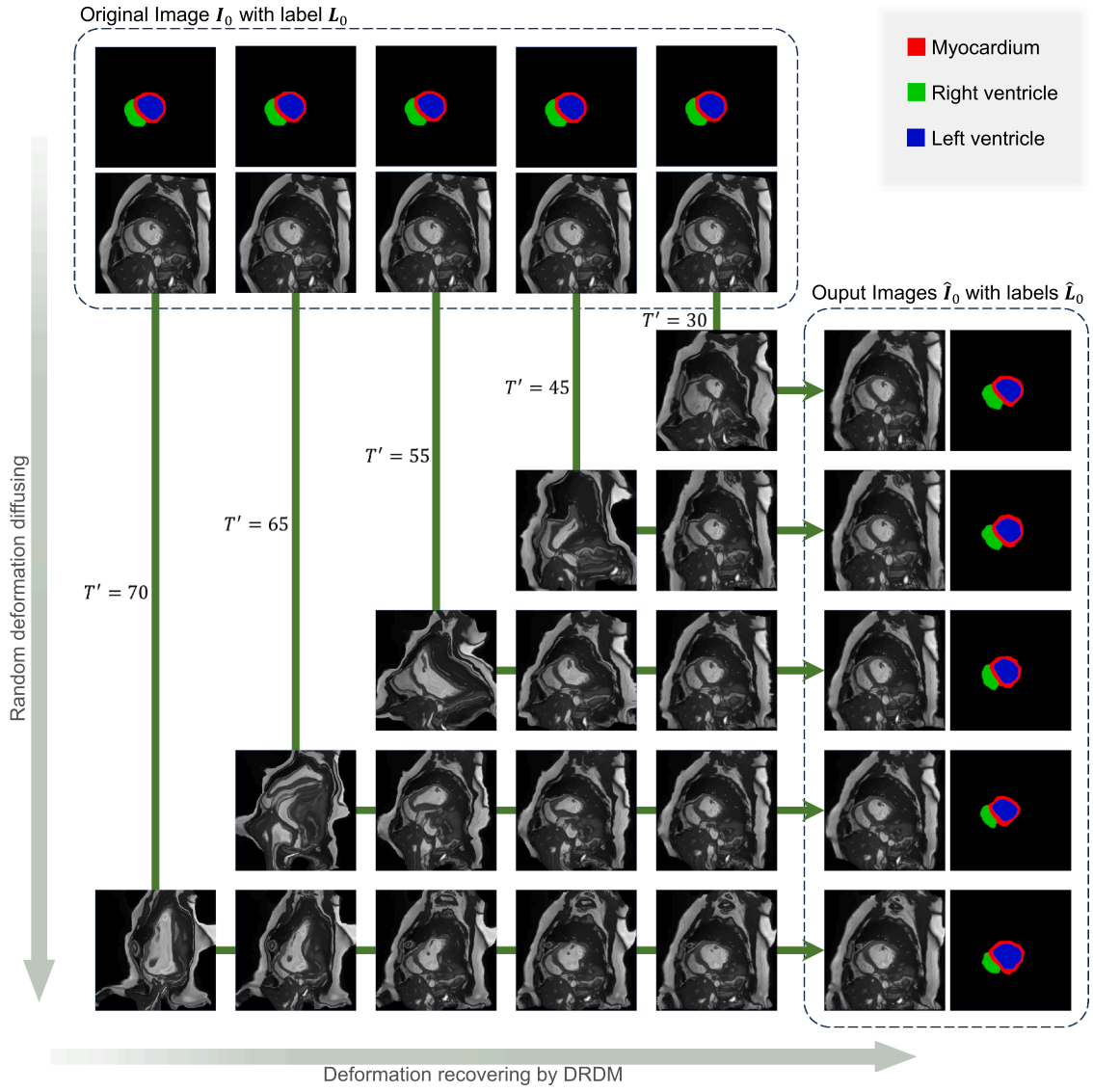

Fig. 6. Visualisation of the deformation diffusion and recovery processes for 2D cardiac MRI images using the proposed DRDM with varying deformation step levels T'.

from the centered region $H \times W \times D$ within the created deformation field.

For all experiments, H, W, D and H', W', D' are set the same values, 256 for 2D MRI scans and 128 for 3D CT scans. The parameters h, w, d are set to twice the dimensions of H, W, D, and T is set at 80. To improve the robustness of DRDM, a small amount of random noise was added during training, introducing a 5% perturbation to the generated DVFs.

The noise level at each time step is set as $n_t := \lfloor t^{0.6} \rfloor$ with $\alpha := 0.6$ and $\beta := 1$. As described in Section 2.1.4, the theoretical setting for the noise level for $\phi_{t:1}$ should be $n_{t:1} = \lfloor t^{1.6}/1.6 \rfloor$, however in practice, $n_{t:1}$ was set to $\lfloor t^{1.3}/1.5 \rfloor$ to mitigate the rounding effects and to increase the redundancy range of network's prediction capacity, thus enhancing its ability to recover from random deformation at each step.

The training process uses the Adam optimizer with $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = 10$. The initial learning rate of 0.0001 with batch sizes of 64 for 2D and 4 for 3D. The model was trained for 1000 epochs on the 2D dataset and 2000 on the 3D dataset to ensure the convergence. All experiments were performed on An Intel Xeon(R) Silver 4210R CPU @ 2.40 GHz Central Processing Unit (CPU) and an Nvidia Quadro RTX 8000 Graphics Processing Unit (GPU) with 48 GB of memory.

3.4. Image and deformation synthesis results

An example of the deformation diffusion and recovery process for cardiac MRI is shown in Fig. 6. It shows that the deformation magnitude becomes larger with increasing deformation level T'. Additional examples of image synthesis for both cardiac MRI and pulmonary CT are presented in Fig. 7, illustrating that diverse images can be generated from a limited number of input MRI and CT images in both 2D and 3D.

Further qualitative results are shown in Fig. 8 for 2D cardiac MRI and Fig. 9 for 3D pulmonary CT scans. These examples highlight the diversity and anatomical plausibility of the synthesized images. Comparisons with baseline methods are provided in Section B, including BigAug (Zhang et al., 2020) (Fig. A.14) and the synthetic training method by Eppenhof and Pluim (Eppenhof and Pluim, 2019) (Fig. A.15).

Quantitative evaluation of the generated deformation is summarized in Table 1. The maximum and average magnitudes of the deformation fields were measured as percentage of the image size and the ratio of

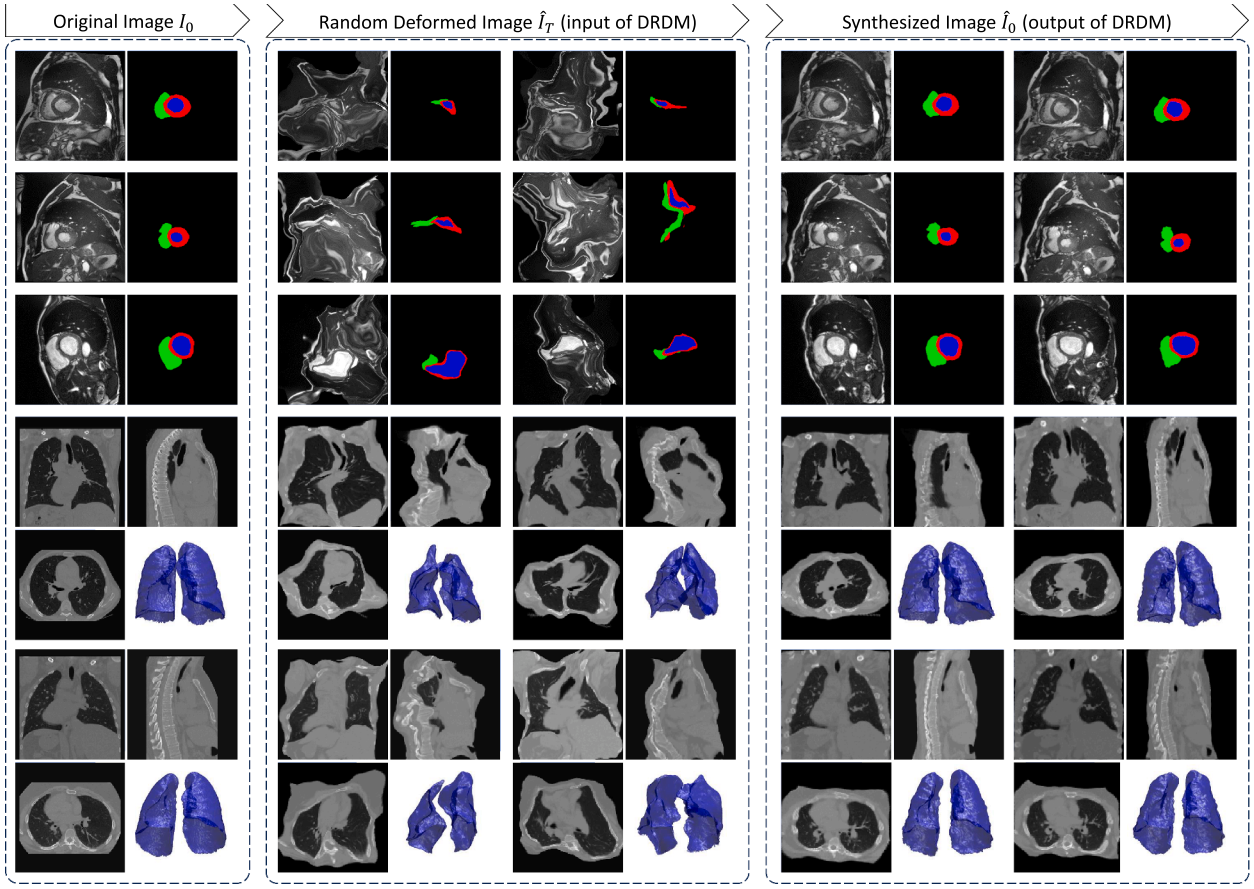

Fig. 7. Diverse image deformation for 2D cardiac MRI and 3D pulmonary CT images (shown as cross-sections through the image center in three orthogonal planes) using the proposed DRDM. Left: original images, middle: randomly deformed images as inputs of DRDM, and right: synthesized output images from DRDM.

voxels with a negative Jacobian determinant of the deformation is used to assess deformation validity of the generated DDF. The results in Table 1 show the ratio of the negative Jacobian determinant ($\mathbb{E}_x \text{detJ}_{<0}(\phi)$) and the magnitude ($\max_x \|\phi[x]\|_2$ and $\mathbb{E}_x \|\phi[x]\|_2$) of the generated deformation fields both increase with the larger deformation level T'. Nevertheless, the overall deformation quality remains high ($\mathbb{E}_x \text{detJ}_{<0}(\phi) < 1\%$) even with a large deformation magnitude ($\max_x \|\phi[x]\|_2 > 10\% \times H, W$).

These results confirm that the proposed DRDM framework can generate large, diverse, and anatomically plausible diffeomorphic deformations while maintaining image quality. The quantitative metrics also provide practical guidance for selecting appropriate deformation magnitudes in different application scenarios.

4. Downstream application in image segmentation

As illustrated in Fig. 5(b), the generated images and their corresponding labels can be used for training a segmentation model $\mathcal{S}$. In this section, DRDM is evaluated as a data augmentation framework for few-shot learning in medical image segmentation. The segmentation framework is described in Section 4.1, with the training process described in Section 4.2. The experimental setup and the corresponding results are explained in Section 4.3 and Section 4.4, respectively.

Algorithm 3: Data augmentation via DRDM.

Input: Images and labels for deformation
$\mathbf{D}^{\text{tgt}} \subset \mathbb{R}^{H\times W\times D} \times \mathbb{R}^{H\times W\times D\times C}$

Output: Deformed pairs of image and label
$\mathbf{D}^{\text{aug}} \subset \mathbb{R}^{H\times W\times D} \times \mathbb{R}^{H\times W\times D\times C}$

```
import the DRDM parameters θ from Algorithm 1;
set a set of deformation levels 𝒯 ⊂ ℤ+ ∩ [1, T];
initialise the output set D^aug ← ∅;
// sample a pair of image and label
foreach (I_0, L_0) ∈ D^tgt do
    // sample a deformation level number
    foreach T' ∈ 𝒯 do
        generate DDF φ using Algorithm 2;
        deform the sampled image: Î ← φ(I_0);
        deform the sampled label: L̂ ← φ(L_0);
        append the deformed image and label into the output
          set: D^aug ← D^aug ∪ {(Î, L̂)};
    end
end
return the output set D^aug.
```

4.1. Segmentation framework

In the segmentation framework, segmentation masks of specific regions $\boldsymbol{L}$ are predicted by a segmentation network $\mathcal{S}_\zeta$ from a given image $\boldsymbol{I}$:

$$\mathcal{S}_\zeta : \mathbb{R}^{H\times W\times D} \to \{0,1\}^{H\times W\times D\times C},\ \boldsymbol{I} \mapsto \tilde{\boldsymbol{L}} \tag{15}$$

with the trainable parameters ζ optimized by minimizing the following:

$$\min_\zeta \{\mathcal{L}_{\text{seg}}(\boldsymbol{L}, \tilde{\boldsymbol{L}})\} \tag{16}$$

where c denotes the number of output channels, $\tilde{\boldsymbol{L}}$ denotes the predicted segmentation mask.

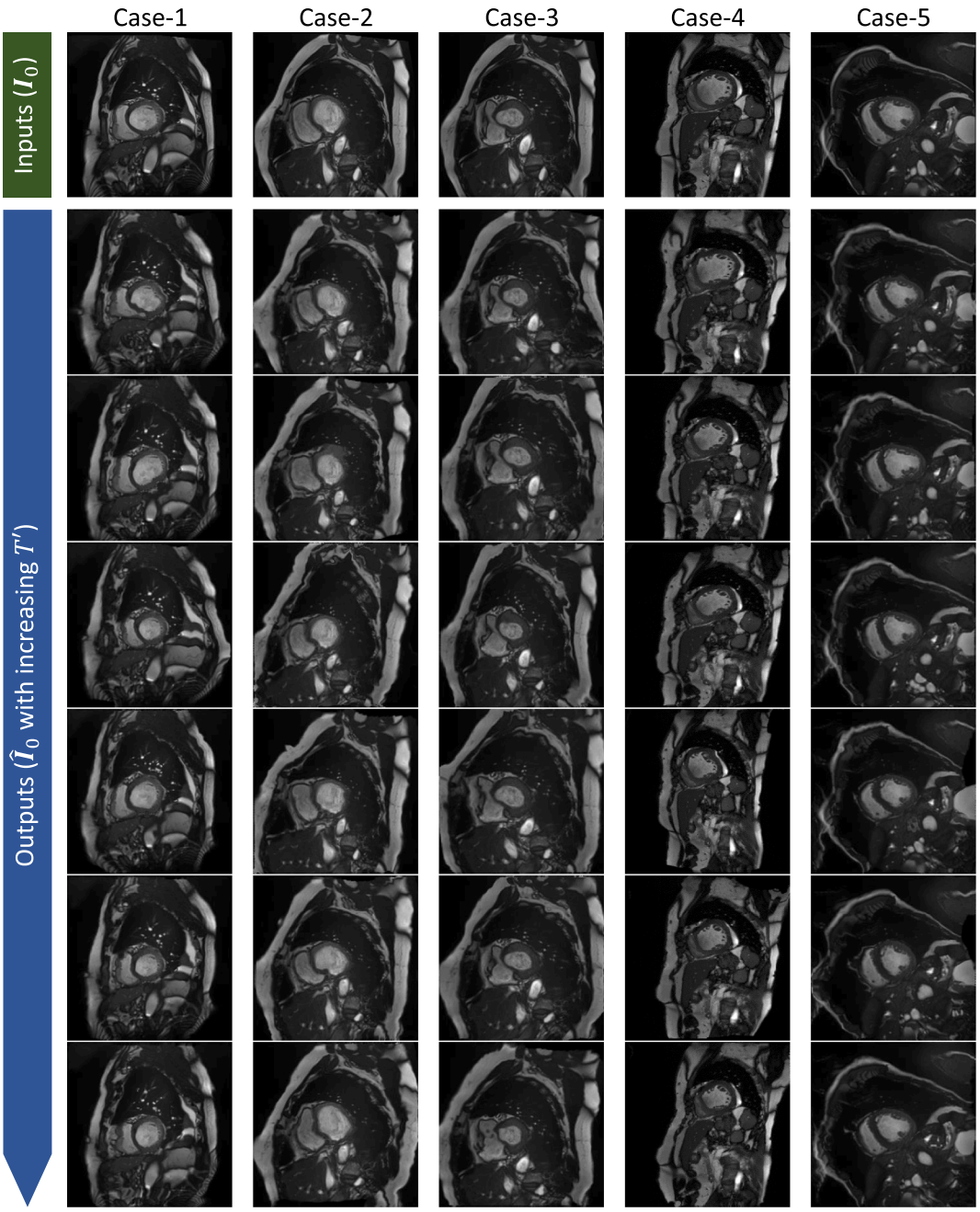

Fig. 8. Original and diversely deformed images of five subjects using the proposed DRDM for 2D cardiac MRI scans.

The U-Net (Ronneberger et al., 2015), is employed as the segmentation model in this experiment. Each dense block consists of two 3x3 convolutions, each followed by a Rectified Linear Unit (ReLU). The encoder path includes max-pooling operations with a stride of 2, while the decoder path employs up-sampling operations with the same stride. A Sigmoid activation function is used at the network output to produce probabilistic segmentation masks.

4.2. Segmentation network training

The segmentation loss function $\mathcal{L}_{\text{seg}}$ is based on Binary Cross Entropy (BCE):

$$\begin{cases} \mathcal{L}_{\text{seg}} := \mathbb{E}_x\left(\sum_{i=1}^{C} \mathcal{L}_{\text{seg}}^{\text{bce}}(\boldsymbol{L}[x,i], \tilde{\boldsymbol{L}}[x,i])\right) \\ \mathcal{L}_{\text{seg}}^{\text{bce}} := \boldsymbol{L}[x,i]\log(\tilde{\boldsymbol{L}}[x,i]) + (1-\boldsymbol{L}[x,i])\log(1-\tilde{\boldsymbol{L}}[x,i]) \end{cases} \tag{17}$$

The network is trained using the Adam optimizer with an initial learning rate of 0.001 and an exponential learning rate scheduling strategy. The batch size is set to 64 for 2D images. Training is conducted using the same computational hardware described in Section 3.3.

4.3. Experimental setup for segmentation

Within the segmentation framework, the original images and corresponding labels from the target domain were augmented using DRDM with varying deformation levels T', as illustrated in Fig. 5(b). The augmented data were included in the training set for downstream segmentation tasks.

To assess the effectiveness of DRDM for data augmentation, several augmentation strategies were compared within the same segmentation framework. Among them, BigAug (Zhang et al., 2020) was selected as the primary baseline. BigAug applies nine stacked transformation

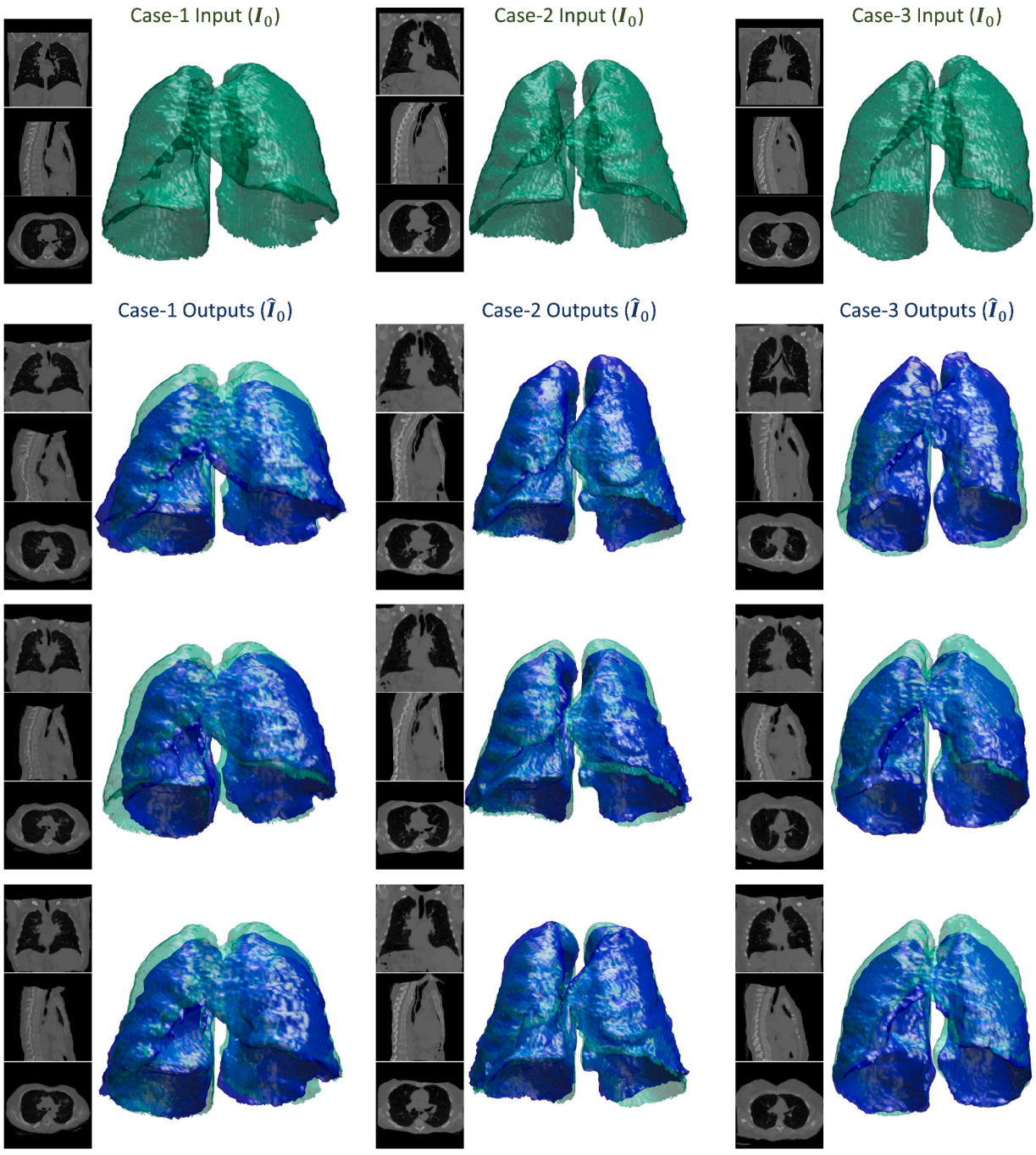

Fig. 9. Lung shapes and three orthogonal cross-sections (frontal, sagittal, and transverse) through the center of the image for original and deformed images (■ and ■) of three subjects generated using the proposed DRDM for 3D pulmonary CT scans.

modules that alter image quality, appearance, and spatial characteristics (including deformation) to improve domain generalization performance.

As described in Section 3.1, the ACDC data divided into 100 subjects for training and 100 for testing. In the segmentation experiments, each dataset was augmented 32 times using both DRDM, following Algorithm 3, and BigAug (Zhang et al., 2020). Comparisons were performed under different training conditions, with labeled data subsets containing 5 subjects (5%), 20 subjects (20%), 50 subjects (50%), and 100 subjects (100%).

Segmentation models trained with different augmentation strategies were evaluated using multiple performance metrics, including the Average Surface Distance (ASD), Dice Similarity Coefficient (DSC) (F1-score), precision (F0-score), and sensitivity (F∞-score) between the ground-truth segmentation masks $\boldsymbol{L}$ and the predicted results $\tilde{\boldsymbol{L}}$.

Ground-truth label maps were created where 0, 1, 2 and 3 represent voxels located in the background, in the RV cavity, in the myocardium, and in the LV cavity.

4.4. Image segmentation results

Representative segmentation results on cardiac MRI are shown in Fig. 10, for different ratios of labelled data at 5%, 20%, 50%, and 100%. These qualitative results demonstrate that the U-Net model augmented with the proposed DRDM framework outperforms the BigAug approach, particularly in the segmentation of the right ventricle.

Table 2
Segmentation results on cardiac MRI showing the average DSC (%), sensitivity (sns/%), and precision (prec/%) obtained using different data augmentation methods. Results are reported for a vanilla U-Net trained with varying subjects number (#subj) and ratio of the labelled images.

#subj/ratio	aug method	Left ventricle			Right ventricle			Myocardium			Average		
		dsc↑	sns↑	prec↑	dsc↑	sns↑	prec↑	dsc↑	sns↑	prec↑	dsc↑	sns↑	prec↑
5 /5%	N/A	55.0	**77.5**	45.5	39.5	48.9	35.6	46.3	63.9	39.0	46.9	63.4	40.0
	BigAug	75.8	73.6	83.2	37.6	29.6	65.2	65.6	66.0	69.1	59.7	57.7	72.5
	DRDM	**77.0**	73.7	**84.1**	**59.6**	**55.8**	**77.3**	**67.9**	**68.5**	**73.4**	**68.2**	**66.0**	**78.3**
20/20%	N/A	75.8	83.7	72.8	59.7	52.6	74.9	62.1	63.7	66.8	65.8	66.6	71.5
	BigAug	78.0	75.3	**85.5**	53.9	47.1	**79.9**	68.9	64.1	**81.6**	66.9	62.2	**82.3**
	DRDM	**83.0**	**86.8**	81.4	**75.0**	**75.6**	78.9	**74.6**	**82.5**	71.8	**77.5**	**82.6**	77.4
50/50%	N/A	82.9	80.6	83.0	70.2	65.1	83.3	74.6	77.9	72.7	75.9	74.5	79.7
	BigAug	84.8	83.1	**90.3**	61.1	54.5	**84.9**	78.5	77.1	82.8	74.8	71.6	86.0
	DRDM	**91.4**	**95.5**	88.3	**85.6**	**89.8**	83.5	**84.1**	**94.4**	**88.3**	**87.1**	**93.2**	**86.7**
100/100%	N/A	89.3	90.5	92.6	85.2	83.2	89.4	84.9	83.8	**86.9**	86.5	85.8	89.6
	BigAug	87.2	80.2	**97.6**	76.9	69.3	**92.7**	81.5	79.8	85.0	81.9	76.4	**91.8**
	DRDM	**92.5**	**96.5**	89.1	**87.9**	**93.2**	83.8	**85.4**	**95.1**	77.8	**88.6**	**94.9**	83.6

The distribution of DSC and ASD values are presented in Fig. 11 for further quantitative comparison. The results indicate that the proposed DRDM method outperforms BigAug across most label ratio settings. Specifically, the DSC and ASD metrics for the proposed DRDM are significantly higher ($p < .01$) than those for BigAug in the right ventricle and generally improved in the other cardiac structures.

Quantitative results for DSC, sensitivity, and precision are presented in Table 2. These results consistently show that DRDM outperforms BigAug in most settings for DSC and sensitivity (sns). Notably, BigAug tends to conservatively segment cardiac structures, as shown in Fig. 10, resulting in higher precision but an increased false-negative rate, and consequently lower sensitivity.

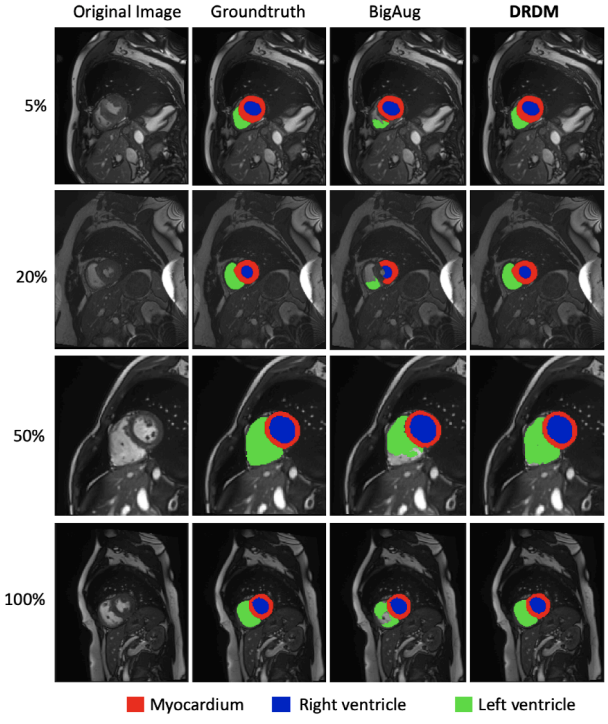

Fig. 10. Segmentation results of models trained with varying ratios of labeled data, comparing different augmentation methods based on BigAug and DRDM.

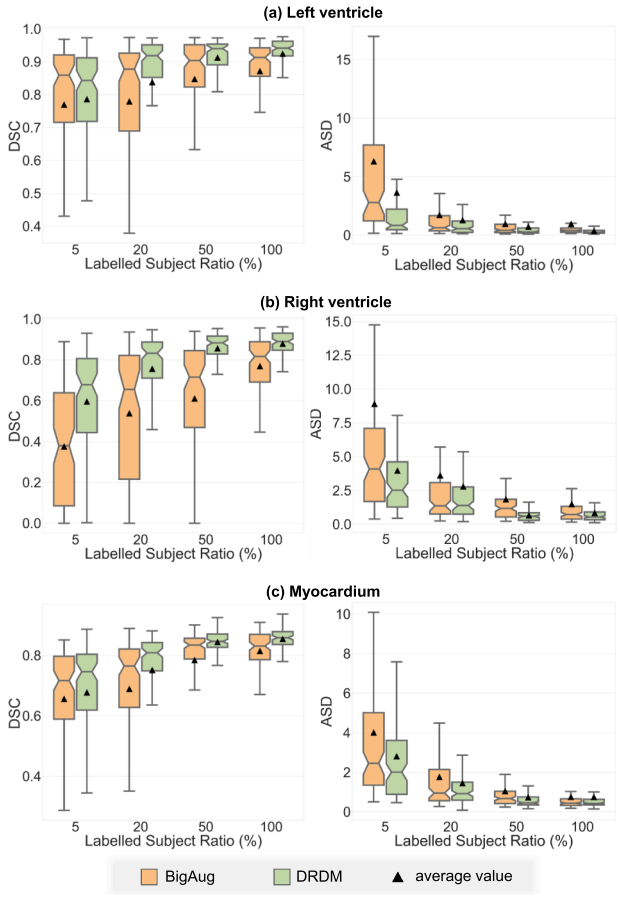

Fig. 11. Quantitative segmentation results of models trained with the varying ratios of labelled data, comparing the proposed DRDM framework with baseline methods across three cardiac structures in MRI. The results demonstrate that DRDM consistently outperforms the baseline under different levels of labelled data availability.

5. Downstream application in image registration

As illustrated in Fig. 5(b), the synthesised images and their corresponding DDFs generated by DRDM are used to pre-train a registration model $\mathcal{R}$. In this section, DRDM is evaluated as a data synthesis tool for

Algorithm 4: Data synthesis via DRDM.

Input: Images for deformation $\mathbf{D}^{\text{tgt}} \subset \mathbb{R}^{H\times W\times D}$
Output: Pairs of deformed images and the DDF $\mathbf{D}^{\text{syn}} \subset \mathbb{R}^{H\times W\times D} \times \mathbb{R}^{H\times W\times D} \times \mathbb{R}^{H\times W\times D\times 3}$

import the DRDM parameters θ from Algorithm 1;
set a set of deformation levels $\mathcal{T} \subset \mathbb{Z}_+ \times \mathbb{Z}_+$;
initialise the output set $\mathbf{D}^{\text{syn}} \leftarrow \emptyset$;
// sample an image
foreach $\boldsymbol{I}_0 \in \mathbf{D}^{\text{tgt}}$ **do**
  // sample deformation level numbers
  foreach $(T'_{\text{aug}}, T'_{\text{syn}}) \in \mathcal{T}$ **do**
    // create the moving image
    generate DDF ϕ_{aug} based on $(\boldsymbol{I}_0, T'_{\text{aug}})$ using Algorithm 2;
    deform the sampled image: $\boldsymbol{I}^{\text{mv}} \leftarrow \phi_{\text{aug}}(\boldsymbol{I}_0)$;
    // create the fixed image and DDF
    generate DDF ϕ_{syn} based on $(\boldsymbol{I}^{\text{mv}}, T'_{\text{syn}})$ using Algorithm 2;
    deform the sampled image: $\boldsymbol{I}^{\text{fx}} \leftarrow \langle \phi_{\text{syn}} \circ \phi_{\text{aug}} \rangle(\boldsymbol{I}_0)$;
    append the deformed images and the DDF: $\mathbf{D}^{\text{aug}} \leftarrow \mathbf{D}^{\text{syn}} \cup \{(\boldsymbol{I}^{\text{mv}}, \boldsymbol{I}^{\text{fx}}, \phi_{\text{syn}})\}$;
  end
end
return the output set $\mathbf{D}^{\text{syn}}$.

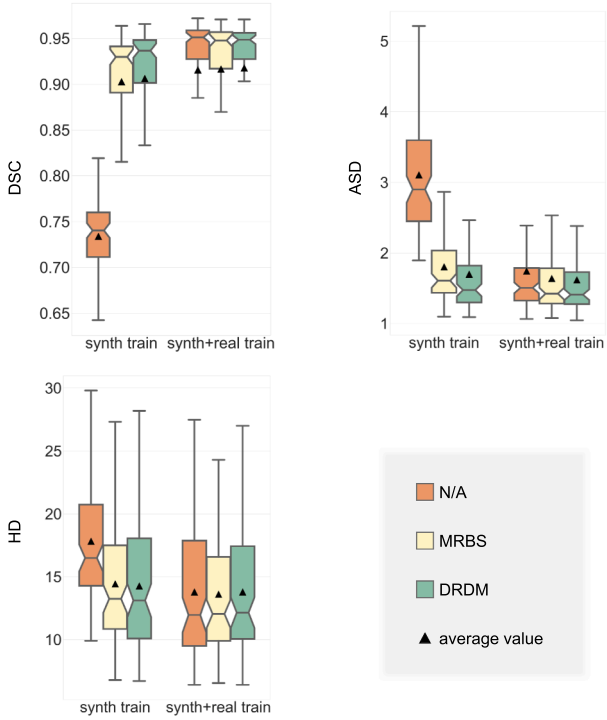

Fig. 12. Quantitative results of registration models trained with synthetic and real data, demonstrating the performance improvement achieved through pretraining using the proposed DRDM framework.

synthetic training in the image registration task. The registration framework is described in Section 5.1, with the training process described in Section 5.2. The experimental setup and the corresponding results are explained in Section 5.3 and Section 5.4, respectively.

5.1. Registration framework

In the registration framework, the deformation field between a pair of images is estimated by a registration network $\mathcal{R}_\eta$ defined as:

$$\mathcal{R}_\eta : (\mathbb{R}^{H\times W\times D}, \mathbb{R}^{H\times W\times D}) \to \mathbb{R}^{H\times W\times D\times 3},\ (\boldsymbol{I}^{\text{mv}}, \boldsymbol{I}^{\text{fx}}) \mapsto \tilde{\phi} \tag{18}$$

with the trainable parameters η optimized by minimizing the following:

$$\min_\eta \{\mathcal{L}_{\text{reg}}(\boldsymbol{I}^{\text{mv}}, \boldsymbol{I}^{\text{fx}}, \tilde{\phi})\} \tag{19}$$

where $\boldsymbol{I}^{\text{mv}}, \boldsymbol{I}^{\text{fx}}$ denotes the moving and the fixed images, respectively, and $\tilde{\phi}$ denotes the estimated deformation field.

In this study, the VoxelMorph architecture (Balakrishnan et al., 2019) is employed as the registration network, providing a robust and widely adopted baseline for learning-based image registration.

5.2. Registration network training

The registration network is pretrained using synthetic image pairs and their corresponding deformation fields, which are generated using DRDM, with varying deformation levels T'. The pretraining registration loss $\mathcal{L}_{\text{reg}}$ consists of two components, Mean Square Error (MSE) and a regularization term:

$$\begin{cases} \mathcal{L}_{\text{reg}} := \lambda_4 \mathcal{L}_{\text{reg}}^{\text{mse}} + \lambda_5 \mathcal{L}_{\text{reg}}^{\text{grad}} \\ \mathcal{L}_{\text{reg}}^{\text{mse}} := \mathbb{E}_x(\|\phi - \tilde{\phi}\|_2) \\ \mathcal{L}_{\text{reg}}^{\text{grad}} := \mathbb{E}_x(\|\nabla \tilde{\phi}[x]\|_1) \end{cases} \tag{20}$$

Following pretraining, the registration model is fine-tuned using the optimization strategy proposed by (Balakrishnan et al., 2019).

The synthetic training process uses $\lambda_4 = 1$ and $\lambda_5 = 1$, with the Adam optimizer. The initial learning rate is set to 0.0001 and the batch sizes of 12. Training is performed on the same computational hardware described in Section 3.3.

5.3. Experimental setup for registration

As described in Section 3.1, the pulmonary CT data provided by (Hering et al., 2022) is split into 20 for training and 10 for testing.

Following Algorithm 4, the original images are first augmented using DRDM to generate moving images $\boldsymbol{I}^{\text{mv}}$, and then further deformed by DRDM to produce fixed images $\boldsymbol{I}^{\text{fx}}$ with corresponding deformation fields ϕ. These image pairs, along with their associated deformation fields, were used as synthetic training data for the registration model.

To evaluate the effectiveness of DRDM in image registration, the proposed framework was compared with the synthetic training approach of (Eppenhof and Pluim, 2019), which is based on a model-registered B-spline (multiple-resolution B-spline (MRBS)) deformation generation method. The same experimental configuration as described in (Eppenhof and Pluim, 2019) was adopted for a fair comparison.

In accordance with that setting, 20 CT scans were each augmented 32 times, and the deformation fields to be learned were synthesized using either DRDM or the B-spline transformer. Following synthetic training, all registration models were further fine-tuned in an unsupervised manner using the real pulmonary CT data, following the optimization strategy of (Balakrishnan et al., 2019).

Registration performance was evaluated using multiple quantitative metrics, including DSC (F1), ASD, and Hausdorff Distance (HD) computed between the ground-truth lung masks in the fixed image $\boldsymbol{L}$ and the corresponding deformed masks $\tilde{\boldsymbol{L}}$ obtained via the estimated deformation field.

5.4. Image registration results

The distribution of DSC, ASD, and HD values evaluated for the proposed model and the baseline models are shown in Fig. 12. The registration model synthetically trained by the proposed DRDM method outperforms the model trained with the MRBS-based method in ASD ($p < .05$).

Table 3
Inter-subject registration results on pulmonary CT showing the average DSC (%), ASD (voxel), and HD (voxel) using a vanilla VoxelMorph (Balakrishnan et al., 2019). Results are reported for different synthetic training strategies, including the Multi-Resolution B-Spline (MRBS) method (Eppenhof and Pluim, 2019) and the proposed DRDM framework, followed by unsupervised fine-tuning on real training data.

synth method	real train	DSC↑ (%)	ASD↓ (vox)	HD↓ (vox)	$\lvert J\rvert_{<0}$ ↓ (‰)
N/A	×	73.39	3.11	17.82	–
MRBS	×	90.29	1.80	14.45	**3.25**
DRDM	×	**90.64**	**1.71**	**14.27**	4.42
N/A	✓	91.57	1.74	13.80	5.38
MRBS	✓	91.66	1.64	**13.62**	4.96
DRDM	✓	**91.79**	**1.62**	13.71	**4.95**

Furthermore, the model pretrained synthetically with DRDM achieves performance comparable to that of the model trained directly on real data.

Quantitative results, including the average values of DSC, ASD, HD, and negative Jacobian determinant ratio are summarised in Table 3. These results consistently demonstrate that the proposed DRDM framework significantly ($p < .05$) outperforms MRBS in synthetic training of the registration model, achieving registration accuracy comparable to real-data training.

Overall, these findings further confirm the effectiveness of the deformed images generated by DRDM for enhancing synthetic training in image registration tasks.

6. Related works

6.1. Diffusion models in medical image analysis

Several recent studies have investigated the application of diffusion models to various medical image analysis tasks, including anomaly detection (Wolleb et al., 2022; Bercea et al., 2023; Liang et al., 2023) and image registration (Qin and Li, 2023; Gao et al., 2023).

Wolleb et al. (2022) proposed a method combining an intensity noising-denoising scheme (Ho et al., 2020; Song et al., 2020) with classifier guidance for 2D image-to-image translation. This technique transforms diseased subjects' images into their healthy counterparts while preserving anatomical information, allowing the difference between the original and translated images to highlight anomaly regions in brain MRI. Similarly, Bercea et al. (2023) introduced an AutoDDPM method, based on DDPM (Ho et al., 2020), for anomaly detection in brain MRI, incorporating an iterative process of stitching-and-resampling to generate pseudo-healthy images. In another study, Liang et al. (2023) proposed MMCCD for multimodal brain MRI anomaly segmentation, utilizing an intensity-based diffusion model (Ho et al., 2020; Song et al., 2020).

Qin and Li (2023) integrated DDPM (Ho et al., 2020) into a registration framework, introducing two complementary diffusion modules: feature-wise diffusion-guided module to enhance feature processing during the registration process, and a score-wise diffusion module to guide the optimisation process while preserving topology in 3D cardiac image registration tasks. Similarly, Gao et al. (2023) employed a diffusion model (Ho et al., 2020) to facilitate multimodal brain MRI registration, combining DDPM with a discrete cosine transform module to disentangle structural information, thus simplifying the multimodal problem to a quasi-monomodal registration task.

Overall, existing diffusion-based methods in medical imaging predominantly operate as converter models, translating images from diseased to healthy states or across imaging modalities, rather than as generative models designed to produce anatomically diverse and physically meaningful deformations.

6.2. Diffusion model for medical image synthesis and manipulation

Pinaya et al. (2022) proposed a 3D T1-weighted brain MRI synthesis framework based on an Latent Diffusion Model (LDM) (Rombach et al., 2022), incorporating a DDIM sampler (Song et al., 2020) to condition the generated images on the subject's age, sex, ventricular volume, and brain volume. Similarly, Ju and Zhou (2024) combined the advanced Mamba network (Gu and Dao, 2023) with a cross-scan module into the DDPM framework (Ho et al., 2020) to generate medical images, validated on chest X-rays, brain MRI, and cardiac MRI.

Recent advances have further expanded the field of diffusion-based medical image generation. Fast-DDPM (Jiang et al., 2025) accelerates diffusion-based medical image generation by reducing denoising steps from thousands to only a few while maintaining high image quality and anatomical fidelity. DiffBoost (Zhang et al., 2024) leverages text-guided diffusion models to synthesize structure-aware medical images, enhancing realism and improving the performance of downstream segmentation tasks.

While these appearance-generation methods produce photorealistic images, they face fundamental limitations such as hallucinations and a lack of interpretable correspondence with real anatomical structures (Deo et al., 2025). While some works e.g. Med-DDPM (Dorjsembe et al., 2024) introduced a conditional diffusion model to generate anatomically consistent 3D MRI synthesis, the loss of correspondence with source images remains unavoidable. Consequently, such generated images cannot be efficiently used for annotation-consistent augmentation for downstream tasks (Sections 4 and 5).

An alternative strategy is to generate deformation fields rather than image intensities using diffusion modelling. DiffuseMorph (Kim et al., 2022) uses DDPM (Ho et al., 2020) to estimate the conditional score function for deformation, combined with a deformation module to estimate deformation between image pairs for registration tasks, including 4D temporal medical image generation of cardiac MRI (Kim et al., 2022). Starck et al. (2024) introduced a conditional atlas generation framework based on LDM (Rombach et al., 2022), generating deformation fields conditioned on specific parameters, with a registration network guiding the optimization of atlas deformation processes.

Wu et al. (2025) formulated image generation as geodesic trajectories within a learned deformation space, enabling anatomically consistent synthesis through continuous shape transformations. Wang et al. (2025) proposed a diffusion-based model that learns a population template and generates subject-specific anatomy via deformation-field-driven synthesis of novel 3D brain MRIs.

Despite these advances, existing deformation-based diffusion methods still depend on intensity- (Kim et al., 2022; Kim and Ye, 2022) or latent-feature-level denoising (Starck et al., 2024; Wu et al., 2025; Wang et al., 2025), typically within registration-guided frameworks that constrain deformation plausibility. Consequently, the diversity of the generated deformations is largely limited to interpolation between image pairs (Kim et al., 2022; Kim and Ye, 2022; Wu et al., 2025) or the deformation of atlas images (Starck et al., 2024; Wang et al., 2025). These constraints restrict the diversity of instance-specific deformations and limit their potential for data augmentation or the creation of anatomically diverse deformation fields for downstream tasks.

6.3. Data augmentation for few-shot image segmentation

Several methods have been proposed for few-shot image segmentation, addressing the challenge of limited annotations in medical image analysis tasks. These approaches can be broadly categorised into pseudo-label-based strategies and data augmentation techniques.

Dang et al. (2022) introduced a few-shot learning framework for vessel segmentation, utilising weak and patch-wise annotations. This approach includes synthesising pseudo-labels for a segmentation network and utilising a classifier network to generate additional labels and assess low-quality images. Similarly, an uncertainty estimation-based

mean teacher segmentation method was proposed to enhance the reliable training of a student model in cardiac MRI segmentation (Wang et al., 2022b). Another semi-supervised method was introduced based on mutual learning between two Vision Transformers and one Convolutional Network, utilizing a dual feature-learning module and a robust guidance module designed for consistency (Wang et al., 2022a).

However, pseudo-label-based methods require a sufficient number of annotated labels to ensure the accuracy of an additional model for pseudo-label creation, which limits their applicability in tasks with extremely low-annotation scenarios.

To overcome this limitation, an atlas-based data augmentation technique was introduced in (Zhao et al., 2019) to create labelled medical images for brain MRI segmentation by spatially and cosmetically aligning an annotated atlas with other images. However, this method's diversity is constrained by the variability of the available atlases, and new registration and appearance transformation models must be trained for each atlas. Bortsova et al. (2019) proposed a few-shot segmentation method based on data augmentation through elastic deformation transform (Davis et al., 1997), with a segmentation consistency loss across labelled and unlabelled images. Furthermore, Zhang et al. (2020) proposed a combination of nine different types of cascaded augmentation methods, named BigAug. These methods vary in image quality (sharpness, blurriness, and intensity noise level), image appearance (brightness, contrast, and intensity perturbation), and spatial configuration (rotation, scaling, and deformation), validated on cardiac MRI/ultrasound and prostate MRI. This method, BigAug, has also been compared in Section 4. A statistical deformation model combining registration-learned transformations with principal component analysis was proposed by He et al. (2024) for volumetric medical image segmentation. While effective for data augmentation, this approach requires paired images for registration training and generates deformations through interpolation in a learned population space, limiting its capacity for instance-specific deformation synthesis (He et al., 2024).

Additionally, Mo et al. (2024) proposed a data-efficient approach that learns a vector field from cropped image patches to trace tissue boundaries, achieving strong performance with very limited labeled data on chest X-ray and dermoscopy segmentation. However, it struggles with complex tissue topologies and is difficult to extend to 3D imaging.

6.4. Synthetic training for image registration

Synthetic spatial transformations have been widely applied to train image registration models, particularly when annotated data are limited or unavailable. For example, random rigid transformations can be easily synthesised to train a model for rigid registration, aimed at aligning micro-CT scans of murine knees with and without contrast enhancement (Zheng et al., 2023).

Rohé et al. (2017) proposed a training strategy for cardiac MRI deformable registration, based on synthetically deforming the segmented mask of the target tissues via elastic body splines (Davis et al., 1997). Similarly, Uzunova et al. (2017) adopted a locality-based multi-object statistical shape model method (Wilms et al., 2017) for statistical appearance modelling, to synthesise training data for medical image registration (Uzunova et al., 2017). However, these methods rely on segmentation masks or statistical shapes of the target tissues prior to training the registration model, making them unsuitable for unsupervised training approaches.

eliminate the need for annotations, random deformations based on Gaussian smoothing sampling (Sokooti et al., 2017; Gonzales et al., 2021) have been used for registration model training. Recently, a mixture of Gaussian and thin-plate splines (Zheng et al., 2022, 2024) have also been used for pretraining registration models. For pulmonary CT registration tasks, Eppenhof and Pluim (2019) applied a multi-resolution B-spline (Lee et al., 1997) model to generate random deformations for data synthesis and model pretraining. This B-spline-based method serves as a key comparison baseline for synthetic registration experiments in Section 5.

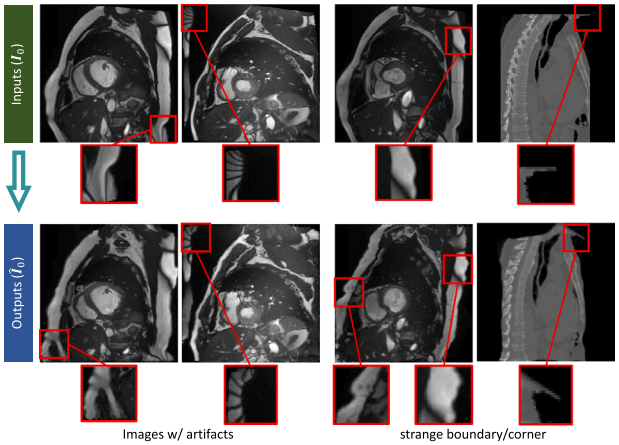

Fig. 13. Artifacts in the original MRI and occasional cropping of unexpected tissue in CT images can result in atypical or distorted structures in the generated outputs.

7. Discussion and conclusion

7.1. Plausible and diverse deformation synthesis

The experiments conducted on cardiac MRI and pulmonary CT, as presented in Section 3, demonstrate that the proposed method, DRDM, is capable of generating anatomically plausible and diverse deformations for instance-specific images. Unlike previous deformation methods (Kim et al., 2022; Kim and Ye, 2022; Starck et al., 2024), DRDM does not rely on a registration framework to guide the deformation process. This independence enables the model to produce a broader range of deformation patterns while maintaining anatomical coherence. In comparison to earlier deformation-based augmentation methods (Zhang et al., 2020; Eppenhof and Pluim, 2019), DRDM generates more customized and anatomically consistent deformations for each individual image.

The apparent diversity of deformations in 2D cardiac images is inherently more limited than in 3D volumes. This restriction arises because 2D slices capture fixed cross-sections of a 3D anatomy, with fewer degrees of freedom for spatial variation. Simulating inter-slice transformations that mimic positional changes across the 3D anatomy would introduce discontinuities, causing tissue to appear or disappear between slices, which could compromise anatomical correspondence and degrade downstream task performance.

7.2. Generated artifacts and unreasonable structures

We observed artifacts characterized by sharp corners and irregular boundaries in some generated images (Fig. 13). Further investigation identified two primary sources of these anomalies: inherent artifacts present in the original MRI acquisitions, and instances where anatomical boundaries were inadvertently cropped in the original CT data.

Although such imperfections may be undesirable when artifact-free images are required, they can enhance the realism for data augmentation by preserving the noise characteristics of clinical acquisitions. Importantly, the negative Jacobian determinant ratio (Table 1) remains below 1% even for large deformations, confirming that the generated deformation fields are smooth and topologically valid.

7.3. Improvement of downstream task

As described in Section 4 and Section 5, the additional experiments on segmentation and registration tasks confirm the efficacy and

applicability of the diverse and instance-specific deformations generated by DRDM. Previous augmentation methods typically rely on fully random transformations in image quality, image appearance, and spatial features, without adapting to the characteristics of each individual image. In contrast, DRDM synthesises more realistic images and thus improves the downstream-task models in learning the realistic distribution of images by balancing diversity and plausibility. It is also noteworthy that DRDM can be combined with other data augmentation methods to further enhance downstream tasks. The observed improvements in segmentation and registration tasks collectively validate the reliability, generalizability, and effectiveness of the deformations generated by DRDM.

7.4. Limitations of this research

The objective of this study is to generate diverse, high-quality, and realistic image deformations. Although experimental results in Section 3 show that the generated deformations are diverse and reasonable, the perceptual realism of these deformations can only be evaluated qualitatively. Quantitative assessment of visual realism remains challenging, a limitation common to image generation research (Deo et al., 2025).

The Fréchet Inception Distance (FID) is widely used to measure the similarity between generated and real image distributions in natural image synthesis (Rombach et al., 2022; Ho et al., 2020; Song et al., 2020; Qiao et al., 2019). However, applying FID to medical imaging tasks is problematic due to domain-specific factors, including misaligned data distributions (Jayasumana et al., 2024) and the lack of suitable pretrained feature extractors for medical images.

Therefore, in this study, the evaluation of DRDM relies on its effectiveness in downstream tasks, such as segmentation and registration, to demonstrate the quality and clinical utility of the synthesised medical images. The improvement observed in these tasks can prove that the synthetic images and deformation generated by DRDM align to the data distributions learned by the segmentation and registration models in the downstream tasks. Nevertheless, it should be noted that the distributions captured by these task-specific models are only approximations of real-world data distributions, and further investigation is required to establish a direct quantitative link.

7.5. Prospective applications in future

This paper demonstrates the application value of DRDM for data augmentation in few-shot segmentation and data synthesis for registration. There is considerable potential for exploring other directions. The DRDM can be modified to accept conditional inputs that regulate the generated deformation fields and their corresponding deformed images for specific applications, including conditional image registration and text-guided image synthesis. A segmentation module can be employed to decompose different regions of images, enabling DRDM to generate more complex deformation fields with multiple continuums. An image modality converter module can be combined to generate deformed images in another modality. Furthermore, DRDM can be combined with conventional intensity-based or latent-based diffusion models to address a broader range of variations. Such hybrid approaches would use DRDM for anatomical deformations while intensity-based models handle appearance changes, textural variations, and pathological features. This combination could be particularly valuable for generating pathological variations, addressing data imbalance in rare conditions, and creating realistic temporal sequences with both anatomical motion and appearance changes.

7.6. Conclusion

In this study, we proposed a novel diffusion-based deformation generative model, termed DRDM, for image manipulation and synthesis in medical imaging. The experimental results demonstrate that DRDM achieves both anatomical plausibility and diversity in the generated deformations and substantially enhances the performance of downstream tasks, including cardiac MRI segmentation and pulmonary CT registration. These findings highlight the potential of DRDM as a general framework for anatomically consistent image synthesis, deformation modelling, and data augmentation in a wide range of medical imaging applications.

8. Declaration of AI technologies used in writing

During the preparation of this work the authors used ChatGPT[1] in order to proofread the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

CRediT authorship contribution statement

Jian-Qing Zheng: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization; **Yuanhan Mo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation; **Yang Sun:** Writing – review & editing, Visualization, Validation, Software, Investigation, Formal analysis, Data curation; **Jiahua Li:** Software, Resources; **Fuping Wu:** Writing – review & editing, Investigation, Formal analysis; **Ziyang Wang:** Writing – review & editing, Resources; **Tonia Vincent:** Supervision; **Bartłomiej W Papież:** Writing – review & editing, Supervision.

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.

Acknowledgment

J.-Q. Z. acknowledges the Kennedy Trust Prize Studentship (AZT00050-AZ04) and the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (CIFMS), China (grant number: 2018-I2M-2-002). B.W.P. acknowledges the Rutherford Fund at Health Data Research UK (grant no. MR/S004092/1).

Appendix A. Network architecture for DRDM

The network structure detail for the DRDM is shown in Table B.4, where "embed" denotes feature embedding, "fc" denotes a fully connected layer, "act" denotes the ReLU activation function with 0.01 negative slope, "#chnl" denotes channel number for input or output, "conv" denotes the convolution with kernel size of 3, stride of 1 and padding size of 1, "stride conv" denotes convolution with stride of 2, "trans conv" denotes transpose convolution, and "ACNN" denotes the ACNN-II block (Zhou et al., 2020) as described in Table B.5. As described in Eq. (12), one image $\hat{\boldsymbol{I}}_t$ is fed into DRDM network and a DVF $\hat{\psi}_t^{(k)}$ is predicted.

As shown in Table B.5, the network structure detail for the i^{th} ACNN-II block, where "conv" denotes convolution, "ker param" denotes kernel parameters, with "dila" as dilation rate, "str" as the stride rate, and "pad" as the padding size, "norm" denotes the instance normalization, and "act" denotes the leaky ReLU activation function with 10^{-6} negative slope. The input of ACNN is a feature map with c_i and the output is the feature map with c_{i+1} processed by three convolution or dilated convolution and activation function.

[1] OpenAI. (2024). ChatGPT (4o) [Large language model]. https://chatgpt.com

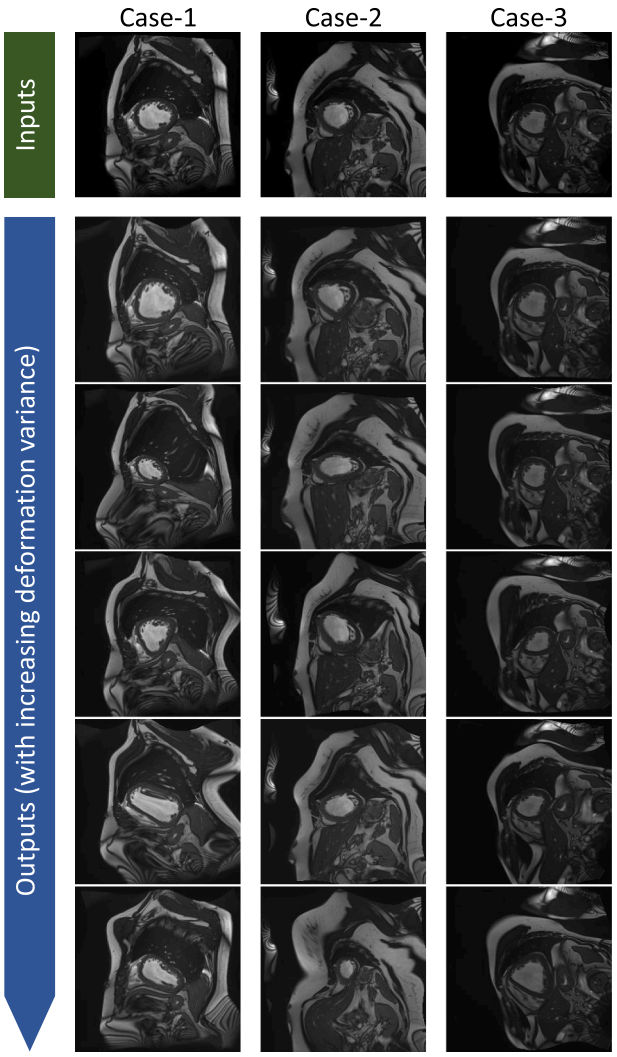

Fig. A.14. The original and deformed images of three subjects by Elastic transformation as used in BigAug for 2D cardiac MRI scans.

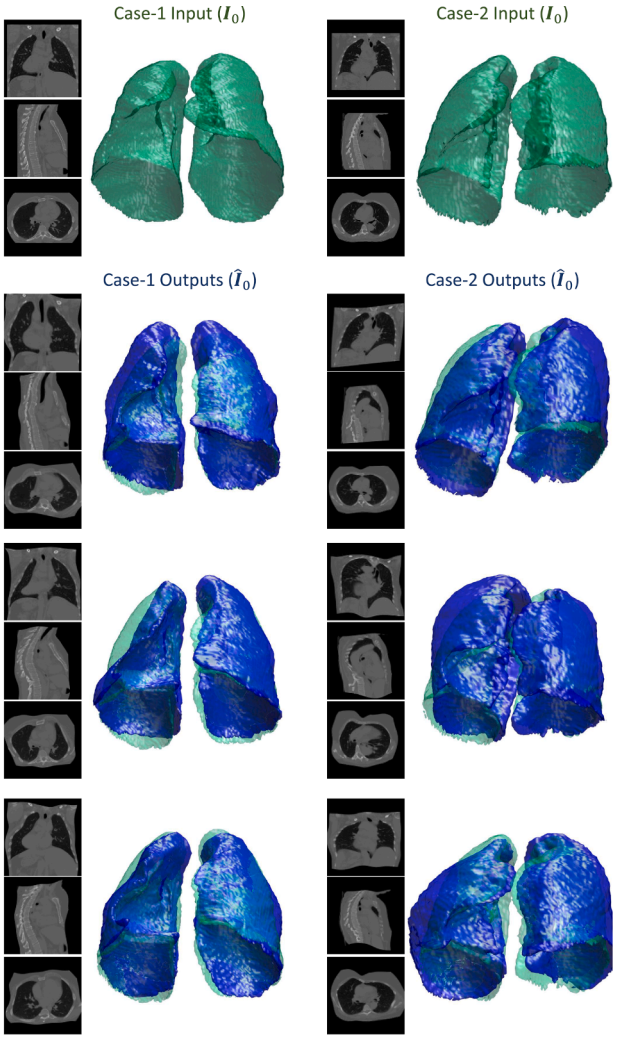

Fig. A.15. The original and deformed images of two subjects by MRBS for 3D pulmonary CT scans.

Appendix B. Baseline results for image deformation

The baseline results for image deformations are illustrated in Fig. A.14 using Elastic transform (a part of BigAug) (Zhang et al., 2020) and Fig. A.14 using MRBS (Eppenhof and Pluim, 2019). Notably, the deformations appear unrealistic, such as the unnaturally expanded or squeezed ventricles and the distorted body shape shown in Fig. A.14 and the unnatural shearing lung in Fig. A.15. These unrealistic deformations can negatively impact the effectiveness of data augmentation or data synthesis, as validated by the experimental results in Section 4 and Section 5.

Table B.4
Network structure detail for DRDM..

func	spatial size	#chnl in/out	in	out
embed	1,1,1	1/80	t	t0
fc,act,fc	1,1,1	80/1	t0	t1
fc,act,fc	1,1,1	80/10	t0	t2
fc,act,fc	1,1,1	80/20	t0	t3
fc,act,fc	1,1,1	80/40	t0	t4
fc,act,fc	1,1,1	80/80	t0	t5
fc,act,fc	1,1,1	80/40	t0	t6
fc,act,fc	1,1,1	80/20	t0	t7
ACNN	H,W,D	c_0/10	$\hat{\boldsymbol{I}}_t$ + t1	f1
ACNN	H,W,D	10/10	f1	f1
ACNN	H,W,D	10/10	f1	f1
stride conv	H/2,W/2,D/2	10/10	f1	f2
ACNN	H/2,W/2,D/2	10/20	f2 + t2	f2
ACNN	H/2,W/2,D/2	20/20	f2	f2
ACNN	H/2,W/2,D/2	20/20	f2	f2
stride conv	H/4,W/4,D/4	20/20	f2	f3
ACNN	H/4,W/4,D/4	20/40	f3 + t3	f3
ACNN	H/4,W/4,D/4	40/40	f3	f3
ACNN	H/4,W/4,D/4	40/40	f3	f3
stride conv	H/8,W/8,D/8	40/40	f3	f4
ACNN	H/8,W/8,D/8	40/20	f4×t4	f4
ACNN	H/8,W/8,D/8	20/20	f4	f4
ACNN	H/8,W/8,D/8	20/40	f4	f4
trans conv	H/4,W/4,D/4	40/40	f4	f5
ACNN	H/4,W/4,D/4	80/40	f5|f3 + t5	f5
ACNN	H/4,W/4,D/4	40/20	f5	f5
ACNN	H/4,W/4,D/4	20/20	f5	f5
trans conv	H/2,W/2,D/2	20/20	f5	f6
ACNN	H/2,W/2,D/2	40/20	f6|f2 + t6	f6
ACNN	H/2,W/2,D/2	20/10	f6	f6
ACNN	H/2,W/2,D/2	10/10	f6	f6
trans conv	H,W,D	10/10	f6	f7
ACNN	H,W,D	20/10	f7|f1 + t7	f7
ACNN	H,W,D	10/10	f7	f7
ACNN	H,W,D	10/10	f7	f7
conv	H,W,D	10/3	f7	$\hat{\psi}_t^{(k)}$

Table B.5
Network structure detail for the i^{th} ACNN-II block.

func	kern param dila/str/pad	#chnl in/out	in	out
conv,norm	1/1/1	c_i/c_{i+1}	fi	r
act,conv	1/1/1	c_{i+1}/c_{i+1}	r	fo
act,conv	3/1/3	c_{i+1}/c_{i+1}	fo	fo
act	-	c_{i+1}/c_{i+1}	(fo + r)	fo

References

Aerts, H., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M.M., Leemans, C.R., Dekker, A., Quackenbush, J., Gillies, R.J., Lambin, P., 2015. Qdata from nsclc-radiomics. Cancer Imag. Arch. . https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI

Arsigny, V., Commowick, O., Pennec, X., Ayache, N., 2006. A log-euclidean framework for statistics on diffeomorphisms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 924–931.

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. Imag. 38 (8), 1788–1800. https://doi.org/10.1109/TMI.2019.2897538

Bercea, C.I., Neumayr, M., Rueckert, D., Schnabel, J.A., 2023. Mask, stitch, and re-sample: enhancing robustness and generalizability in anomaly detection through automatic diffusion models. In: ICML 3Rd Workshop on Interpretable Machine Learning in Healthcare (IMLH).

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.-M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.-M., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imag. 37 (11), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502

Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., De Bruijne, M., 2019. Semi-supervised medical image segmentation via learning consistency under transformations. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer, pp. 810–818. https://doi.org/10.1007/978-3-030-32226-7_90

Bradbury, R., Vallis, K.A., Papież, B.W., 2024. Paired diffusion: generation of related, synthetic pet-ct-segmentation scans using linked denoising diffusion probabilistic models. In: 2024Del InsThinspace IEEE International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1–5. https://doi.org/10.1109/ISBI56570.2024.10635593

Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreño, M., Albiol, A., Kong, F., Shadden, S.C., Corral Acero, J., Sundaresan, V., Saber, M., Elattar, M., Li, H., Menze, B., Khader, F., Haarburger, C., Scannell, C.M., Veta, M., Carscadden, A., Punithakumar, K., Liu, X., Tsaftaris, S.A., Huang, X., Yang, X., Li, L., Zhuang, X., Viladés, D., Descalzo, M.L., Guala, A., La Mura, L., Friedrich, M.G., Garg, R., Lebel, J., Henriques, F., Karakas, M., Çavuş, E., Petersen, S.E., Escalera, S., Seguí, S., Rodríguez-Palomares, J.F., Lekadir, K., 2021. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. IEEE Trans. Med. Imag. 40 (12), 3543–3554. https://doi.org/10.1109/TMI.2021.3090082

Christensen, G.E., Rabbitt, R.D., Miller, M.I., 1996. Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5 (10), 1435–1447. https://doi.org/10.1109/83.536892

Dang, V.N., Galati, F., Cortese, R., Di Giacomo, G., Marconetto, V., Mathur, P., Lekadir, K., Lorenzi, M., Prados, F., Zuluaga, M.A., 2022. Vessel-CAPTCHA: an efficient learning framework for vessel annotation and segmentation. Med. Image Anal. 75, 102263. https://doi.org/10.1016/j.media.2021.102263

Davis, M.H., Khotanzad, A., Flamig, D.P., Harms, S.E., 1997. A physics-based coordinate transformation for 3-d image matching. IEEE Trans. Med. Imag. 16 (3), 317–328. https://doi.org/10.1109/42.585766

Deo, Y., Jia, Y., Lassila, T., Smith, W. A.P., Lawton, T., Kang, S., Frangi, A.F., Habli, I., 2025. Metrics that matter: Evaluating image quality metrics for medical image generation. arXiv preprint arXiv:2505.07175 .

Dorjsembe, Z., Pao, H.-K., Odonchimed, S., Xiao, F., 2024. Conditional diffusion models for semantic 3d brain MRI synthesis. IEEE J. Biomed. Health Inform. 28 (7), 4084–4093. https://doi.org/10.1109/JBHI.2024.3385504

Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X., 2023. ArSDM: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, pp. 339–349. https://doi.org/10.1007/978-3-031-43895-0_32

Eppenhof, K. A.J., Pluim, J. P.W., 2019. Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans. Med. Imag. 38 (5), 1097–1105. https://doi.org/10.1109/TMI.2018.2878316

Gao, F., He, Y., Li, S., Hao, A., Cao, D., 2023. Diffusing coupling high-frequency-purifying structure feature extraction for brain multimodal registration. In: 2023Del InsThinspace IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp. 508–515. https://doi.org/10.1109/BIBM58861.2023.10385725

Gong, S., Chen, C., Gong, Y., Chan, N.Y., Ma, W., Mak, C. H.-K., Abrigo, J., Dou, Q., 2023. Diffusion model based semi-supervised learning on brain hemorrhage images for efficient midline shift quantification. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 69–81. https://doi.org/10.1007/978-3-031-34048-2_6

Gonzales, R.A., Zhang, Q., Papież, B.W., Werys, K., Lukaschuk, E., Popescu, I.A., Burrage, M.K., Shanmuganathan, M., Ferreira, V.M., Piechnik, S.K., 2021. Moconet: robust motion correction of cardiovascular magnetic resonance t1 mapping using convolutional neural networks. Front. Cardiovascul. Med. 8, 768245. https://doi.org/10.3389/fcvm.2021.768245

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Graf, R., Schmitt, J., Schlaeger, S., Möller, H.K., Sideri-Lampretsa, V., Sekuboyina, A., Krieg, S.M., Wiestler, B., Menze, B., Rueckert, D., et al., 2023. Denoising diffusion-based MRI to CT image translation enables automated spinal segmentation. Eur. Radiol. Exp. 7 (1), 70. https://doi.org/10.1186/s41747-023-00385-2

Gu, A., Dao, T., 2023. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 . https://doi.org/10.48550/arXiv.2312.00752

He, W., Zhang, C., Dai, J., Liu, L., Wang, T., Liu, X., Jiang, Y., Li, N., Xiong, J., Wang, L., et al., 2024. A statistical deformation model-based data augmentation method for volumetric medical image segmentation. Med. Image Anal. 91, 102984. https://doi.org/10.1016/j.media.2023.102984

Hering, A., Hansen, L., Mok, T. C.W., Chung, A. C.S., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., Vesal, S., Rusu, M., Sonn, G., Estienne, T., Vakalopoulou, M., Han, L., Huang, Y., Yap, P.-T., Balbastre, Y., Joutard, S., Modat, M., Lifshitz, G., Raviv, D., Lv, J., Li, Q., Jaouen, V., Visvikis, D., Fourcade, C., Rubeaux, M., Pan, W., Xu, Z., Jian, B., Benetti, F.D., Wodzinski, M., Gunnarsson, N., Sjölund, J., Grzech, D., Qiu, H., Li, Z., Duan, J., Großbröhmer, C., Reinertsen, I., Xiao, Y., Landman, B., Huo, Y., Murphy, K., Lessmann, N., Ginneken, B.v., Dalca, A.V., Heinrich, M.P., 2022. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Trans. Med. Imag. 42 (3), 697–712. https://doi.org/10.1109/TMI.2022.3213983

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851. https://doi.org/10.5555/3495724.3496298

Islam, M.A., Jia, S., Bruce, N. D.B., 2019. How much position information do convolutional neural networks encode? In: International Conference on Learning Representations.

Jaderberg, M., Simonyan, K., Zisserman, A., kavukcuoglu, k., 2015. Spatial transformer networks. Adv. Neural Inf. Process. Syst. 28, 2017–2025. https://doi.org/10.5555/2969442.2969465

Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., Kumar, S., 2024. Rethinking fid: towards a better evaluation metric for image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9307–9315. https://doi.org/10.1109/CVPR52733.2024.00889

Jiang, H., Imran, M., Zhang, T., Zhou, Y., Liang, M., Gong, K., Shao, W., 2025. Fast-DDPM: fast denoising diffusion probabilistic models for medical image-to-image generation. IEEE J. Biomed. Health Inform. . https://doi.org/10.1109/JBHI.2025.3565183

Ju, Z., Zhou, W., 2024. Vm-ddpm: Vision mamba diffusion for medical image synthesis. arXiv preprint arXiv:2405.05667 . https://doi.org/10.48550/arXiv.2405.05667

Kalpathy-Cramer, J., Napel, S., Goldgof, D., Zhao, B., 2015. Qin multi-site collection of lung ct data with nodule segmentations. Cancer Imag. Arch. 10, K9.

Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D., 2023. Diffusion models in medical imaging: a comprehensive survey. Med. Image Anal. 88, 102846. https://doi.org/10.1016/j.media.2023.102846

Kim, B., Han, I., Ye, J.C., 2022. Diffusemorph: unsupervised deformable image registration using diffusion model. In: European Conference on Computer Vision. Springer, pp. 347–364. https://doi.org/10.1007/978-3-031-19821-2_20

Kim, B., Ye, J.C., 2022. Diffusion deformable model for 4d temporal medical image generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 539–548. https://doi.org/10.1007/978-3-031-16431-6_51

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 . https://doi.org/10.48550/arXiv.1312.6114

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, S., Wolberg, G., Shin, S.Y., 1997. Scattered data interpolation with multilevel b-splines. IEEE Trans. Vis. Comput. Graph. 3 (3), 228–244. https://doi.org/10.1109/2945.620490

Li, J., Cao, H., Wang, J., Liu, F., Dou, Q., Chen, G., Heng, P.-A., 2023. Fast non-markovian diffusion model for weakly supervised anomaly detection in brain MR images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 579–589. https://doi.org/10.1007/978-3-031-43904-9_56

Liang, Z., Anthony, H., Wagner, F., Kamnitsas, K., 2023. Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 168–181. https://doi.org/10.1007/978-3-031-47425-5_16

Martín-Isla, C., Campello, V.M., Izquierdo, C., Kushibar, K., Sendra-Balcells, C., Gkontra, P., Sojoudi, A., Fulton, M.J., Arega, T.W., Punithakumar, K., Li, L., Sun, X., Khalil, Y.A., Liu, D., Jabbar, S., Queirós, S., Galati, F., Mazher, M., Gao, Z., Beetz, M., Tautz, L., Galazis, C., Varela, M., Hüllebrand, M., Grau, V., Zhuang, X., Puig, D., Zuluaga, M.A., Mohy-ud Din, H., Metaxas, D., Breeuwer, M., van der, G. R.J., Noga, M., Bricq, S., Rentschler, M.E., Guala, A., Petersen, S.E., Escalera, S., Palomares, J. F.R., Lekadir, K., 2023. Deep learning segmentation of the right ventricle in cardiac mri: the m&ms challenge. IEEE J. Biomed. Health Inform. 27 (7), 3302–3313. https://doi.org/10.1109/JBHI.2023.3267857

Mo, Y., Liu, F., Yang, G., Wang, S., Zheng, J., Wu, F., Papież, B.W., McIlwraith, D., He, T., Guo, Y., 2024. Labelling with dynamics: a data-efficient learning paradigm for medical image segmentation. Med. Image Anal. , 103196. https://doi.org/10.1016/j.media.2024.103196

Papież, B.W, Heinrich, M.P., Fehrenbach, J., Risser, L., Schnabel, J.A., 2014. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. Med. Image Anal. 18 (8), 1299–1311.

Pinaya, W. H.L., Tudosiu, P.-D., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J., 2022. Brain imaging generation with latent diffusion models. In: MICCAI Workshop on Deep Generative Models. Springer, pp. 117–126. https://doi.org/10.1007/978-3-031-18576-2_12

Qiao, T., Zhang, J., Xu, D., Tao, D., 2019. Mirrorgan: learning text-to-image generation by redescription. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1505–1514. https://doi.org/10.1109/CVPR.2019.00160

Qin, Y., Li, X., 2023. Fsdiffreg: feature-wise and score-wise diffusion-guided unsupervised deformable image registration for cardiac images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 655–665. https://doi.org/10.1007/978-3-031-43999-5_62

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A.J., Wright, G.A., 2009. Evaluation framework for algorithms segmenting short axis cardiac MRI. MIDAS J. . https://doi.org/10.54294/g80ruo

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 11Del-Ins–13, 2017, Proceedings, Part I 20. Springer, pp. 266–274. https://doi.org/10.1007/978-3-319-66182-7_31

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01042

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. PMLR, pp. 2256–2265. https://doi.org/10.5555/3045118.3045358

Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B. P.F., Išgum, I., Staring, M., 2017. Nonrigid image registration using multi-scale 3d convolutional neural networks. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11Del-Ins–13, 2017, Proceedings, Part I 20. Springer, pp. 232–239. https://doi.org/10.1007/978-3-319-66182-7_27

Song, J., Meng, C., Ermon, S., 2020. Denoising diffusion implicit models. In: International Conference on Learning Representations.

Starck, S., Sideri-Lampretsa, V., Kainz, B., Menten, M., Mueller, T., Rueckert, D., 2024. Diff-def: Diffusion-generated deformation fields for conditional atlases. arXiv preprint arXiv:2403.16776 . https://doi.org/10.48550/arXiv.2403.16776

Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., Pinto, K., Karim, R., Ammar, M., Daoudi, A., Margeta, J., Sandoval, Z., Stender, B., Zheng, Y., Zuluaga, M.A., Betancur, J., Ayache, N., Chikh, M.A., Dillenseger, J.-L., Kelm, B.M., Mahmoudi, S., Ourselin, S., Schlaefer, A., Schaeffter, T., Razavi, R., Rhode, K.S., 2015. Benchmark for algorithms segmenting the left atrium from 3d CT and MRI datasets. IEEE Trans. Med. Imag. 34 (7), 1460–1473. https://doi.org/10.1109/TMI.2015.2398818

Uzunova, H., Wilms, M., Handels, H., Ehrhardt, J., 2017. Training CNNs for image registration from few samples with model-based data augmentation. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11Del-Ins–13, 2017, Proceedings, Part I 20. Springer, pp. 223–231. https://doi.org/10.1007/978-3-319-66182-7_26

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45 (1), S61–S72. https://doi.org/10.1016/j.neuroimage.2008.10.040

Wang, A.Q., Huang, F., Trang, B., Peng, W., Abbasi, M., Pohl, K., Sabuncu, M.R., Adeli, E., 2025. Generating novel brain morphology by deforming learned templates. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 207–217. https://doi.org/10.1007/978-3-032-04937-7_20

Wang, Z., Li, T., Zheng, J.-Q., Huang, B., 2022a. When cnn meet with vit: towards semi-supervised learning for multi-class medical image semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 424–441. https://doi.org/10.1007/978-3-031-25082-8_28

Wang, Z., Zheng, J.-Q., Voiculescu, I., 2022b. An uncertainty-aware transformer for MRI cardiac semantic segmentation via mean teachers. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 494–507. https://doi.org/10.1007/978-3-031-12053-4_37

Wilms, M., Handels, H., Ehrhardt, J., 2017. Multi-resolution multi-object statistical shape models based on the locality assumption. Med. Image Anal. 38, 17–29. https://doi.org/10.1016/j.media.2017.02.003

Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C., 2022. Diffusion models for medical anomaly detection. In: International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, pp. 35–45. https://doi.org/10.1007/978-3-031-16452-1_4

Wu, N., Jayakumar, N., Xing, J., Zhang, M., 2025. Igg: image generation informed by geodesic dynamics in deformation spaces. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 232–246. https://doi.org/10.1007/978-3-031-96628-6_16

Zhan, F., Yu, Y., Wu, R., Zhang, J., Lu, S., Liu, L., Kortylewski, A., Theobalt, C., Xing, E., 2023. Multimodal image synthesis and editing: a survey and taxonomy. IEEE Trans. Pattern Anal. Mach. Intell. . https://doi.org/10.1109/TPAMI.2023.3305243

Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., Xu, Z., 2020. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Trans. Med. Imag. 39 (7), 2531-2540. https://doi.org/10.1109/TMI.2020.2973595

Zhang, L., Wu, F., Bronik, K., Papiez, B.W., 2025. Diffuseg: domain-driven diffusion for medical image segmentation. IEEE J. Biomed. Health Inform. . https://doi.org/10.1109/JBHI.2025.3526806

Zhang, Z., Yao, L., Wang, B., Jha, D., Durak, G., Keles, E., Medetalibeyoglu, A., Bagci, U., 2024. Diffboost: enhancing medical image segmentation via text-guided diffusion model. IEEE Trans. Med. Imag. . https://doi.org/10.1109/TMI.2024.3519307

Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V., 2019. Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8543–8553. https://doi.org/10.1109/CVPR.2019.00874

Zheng, J.-Q., Lim, N.H., Papież, B.W., 2023. Accurate volume alignment of arbitrarily oriented tibiae based on a mutual attention network for osteoarthritis analysis. Comput. Med. Imag. Graphic. 106, 102204. https://doi.org/10.1016/j.compmedimag.2023.102204

Zheng, J.-Q., Wang, Z., Huang, B., Lim, N.H., Papiez, B.W., 2024. Residual aligner-based network (RAN): motion-separable structure for coarse-to-fine deformable image registration. Med. Image Anal. , 103038. https://doi.org/10.1016/j.media.2023.103038

Zheng, J.-Q., Wang, Z., Huang, B., Vincent, T., Lim, N.H., Papież, B.W., 2022. Recursive deformable image registration network with mutual attention. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 75–86. https://doi.org/10.1007/978-3-031-12053-4_6

Zhou, X.-Y., Zheng, J.-Q., Li, P., Yang, G.-Z., 2020. ACNN: A full resolution dcnn for medical image segmentation. In: 2020Del InsThinspace IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 8455–8461. https://doi.org/10.1109/ICRA40945.2020.9197328