Deformation-aware GAN for Medical Image Synthesis with Substantially Misaligned Pairs

Bowen Xin[1] bowen.xin@csiro.au
Tony Young[2] Tony.Young@health.nsw.gov.au
Claire Wainwright[3] Claire.Wainwright@health.qld.gov.au
Tamara Blake[4] Tamara.Blake@health.qld.gov.au
Leo Lebrat[7] leo.lebrat@csiro.au
Thomas Gaass[5] thomas.gaass@gmail.com
Thomas Benkert[6] benkert.thomas@siemens-healthineers.com
Alto Stemmer[6] alto.stemmer@siemens-healthineers.com
David Coman[4] David.Coman@health.qld.gov.au
Jason Dowling[1] jason.dowling@csiro.au

[1] *Australian E-health Research Center, CSIRO, Australia*
[2] *Liverpool and Macarthur Cancer Therapy Centres and Ingham Institute, Australia*
[3] *University of Queensland, Australia*
[4] *Queensland Children's Hospital, Australia*
[5] *Siemens Healthcare Pty Ltd, Australia*
[6] *MR Application Predevelopment, Siemens Healthcare GmbH, Erlangen, Germany*
[7] *Data61, CSIRO, Australia*

Abstract

Medical image synthesis generates additional imaging modalities that are costly, invasive or harmful to acquire, which helps to facilitate the clinical workflow. **R-tPyS-Q1:** *When training pairs are substantially misaligned (e.g., lung MRI-CT pairs with respiratory motion), accurate image synthesis remains a critical challenge. Recent works explored the directional registration module to adjust misalignment in generative adversarial networks; however, substantial misalignment will lead to 1) suboptimal data mapping caused by correspondence ambiguity, and 2) degraded image fidelity caused by morphology influence on discriminator.* To address the challenges, we propose a novel Deformation-aware GAN (DA-GAN) to dynamically correct the misalignment during the image synthesis based on multi-objective inverse consistency. Specifically, in the generative process, three levels of inverse consistency cohesively optimise symmetric registration and image generation for improved correspondence. In the adversarial process, to further improve image fidelity under misalignment, we design deformation-aware discriminators to disentangle the mismatched spatial morphology from the judgement of image fidelity. Experimental results show that DA-GAN achieved superior performance on a public dataset with simulated misalignments and a real-world lung MRI-CT dataset with respiratory motion misalignment. The results indicate the potential for a wide range of medical image synthesis tasks such as radiotherapy planning.

1. Introduction

Medical image synthesis produces additional imaging modalities to provide essential information for diagnosis or treatment planning, while bypassing the cost and extra time

© 2024 CC-BY 4.0, B. Xin *et al.*

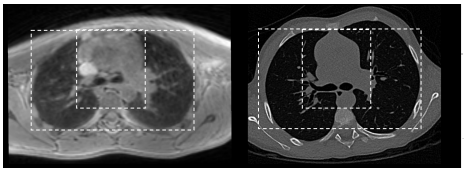

Figure 1: Substantial misalignment in lung MRI-CT pairs due to respiratory motion.

associated with additional scans. **R-3gB1-Q4:***It is particularly useful when the additional scan is invasive, harmful, costly or time-consuming (Liu et al., 2022) (a typical application is synthetic CT in MRI-only radiotherapy dose planning (Spadea et al., 2021)).* Generative adversarial networks (GANs) are widely used in medical synthesis, which usually requires either well-aligned imaging pairs (by supervised methods) or randomly unpaired data (by unsupervised methods). Specifically, supervised GANs, such as Pix2pix and its improved variants (Wang et al., 2018a,b; AlBahar and Huang, 2019), leverage pixel-wise loss on well-aligned imaging pairs to learn the unique and optimal mapping. However, well-aligned pairs are not widely available due to patient motion or organ movement, causing accumulated error and unreasonable placement in supervised methods (Pang et al., 2021). Though registration is commonly used as preprocessing to align images, it is still difficult to acquire perfectly aligned pairs, especially under substantial misalignment such as respiratory motion in lung MR-to-CT synthesis (Sotiras et al., 2013).

Unsupervised GANs are not ideal for misaligned pairs either. Specifically, unsupervised GANs enable training on randomly unpaired data by leveraging extra constraints such as cycle consistency (Zhu et al., 2017; Hoffman et al., 2018; Khorram and Fuxin, 2022), mutual information (Park et al., 2020; Zhan et al., 2022), or geometry consistency (Fu et al., 2019; Xu et al., 2022). However, they are not designed to utilise pairing information to uncover optimal mappings (minimised pixel-wise error), which is essential in tasks such as radiotherapy planning. According to (Shen et al., 2020), **R-3gB1-M3:***cycle consistency mapping used in unsupervised GANs is not strictly one-to-one mapping, which is an important condition in intra-subject medical image synthesis (Wang et al., 2021).* **R-tPyS-Q2:***Diffusion models have shown great potential in computer vision applications due to their strength in capturing distributions (Ho et al., 2020; Song et al., 2020); however, they are computationally expensive and data-hungry to train, hindering their application in the medical domain.*

R-tPyS-Q1:*One recent work RegGAN (Kong et al., 2021) explored directional registration in image synthesis on datasets with simulated misalignment; however, the real-life setting often involves large deformation between pairs (e.g., Figure 1), causing difficulty in learning unique one-to-one mapping due to a large number of local minima (Christensen and Johnson, 2001). The resulting correspondence ambiguity and asymmetric mapping error would add to pixel-wise error in supervised methods, causing a major challenge in generative modelling. The second challenge is the degraded image fidelity caused by the influence of spatial misalignment during the adversarial process. To further elaborate on the issue, the discriminator in RegGAN may recognise the real/fake images purely based on spatial morphology rather than intensity characteristics, thus leading to suboptimal image fidelity.*

In this paper, we propose a Deformation-aware GAN (DA-GAN) to jointly address the above two synthesis challenges when image pairs are substantially misaligned. Firstly,

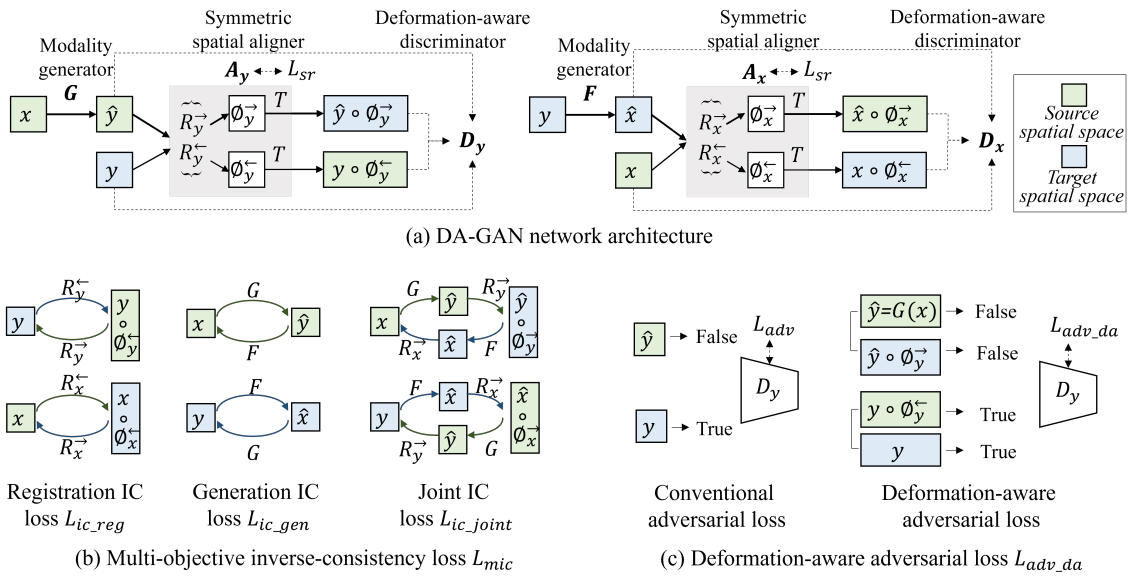

Figure 2: **R-3gB1-M2:***(a) Network architecture of DA-GAN. (b) L_{mic} loss dynamically enhances image correspondence from three objectives. (c) L_{adv_da} loss guides discriminators to learn deformation for improved image fidelity.*

inspired by the capacity of symmetric registration to jointly estimate invertible bidirectional transformation, we propose a multi-objective inverse consistency to comprehensively investigate how to cohesively incorporate symmetric registration into an image generation network. To further improve degraded image fidelity in an adversarial process, we design a deformation-aware adversarial loss, which leverages the outputs from symmetric registration to guide the discriminator to learn the intensity characteristics (that are disentangled from the mismatched spatial morphology). We comprehensively validated our proposed DA-GAN on two datasets, including a public simulation brain dataset with 6 levels of simulated misalignments, and a real-world lung MRI-CT dataset with challenging respiratory-motion misalignments.

2. Methodology

Problem formulation Suppose we have a training dataset with misaligned imaging pairs $(x_i, y_i)|_{i=1}^n$, where $x_i \in X$ and $y_i \in Y$ belong to different modalities. X and Y differ in both intensity characteristics and spatial morphology. Additionally, we denote $\hat{y}_i \in \hat{Y}$ as a transformed y_i that spatially corresponds to source imaging x_i, but $\hat{y}_i$ is unknown in the real world. In other words, both x_i and $\hat{y}_i$ are aligned in source spatial space, but only differ in intensity characteristics. With misaligned multimodal imaging pairs $(x_i, y_i)|_{i=1}^n$, our objective is to accurately synthesise the target imaging $\hat{y}_i$ that is spatially corresponding to the source image x_i for subsequent tasks such as radiotherapy treatment planning.

2.1. DA-GAN overview

DA-GAN network architecture Figure 2a presents the network architecture of our proposed DA-GAN which consists of three major components, including (1) modality generators G and F, (2) symmetric spatial aligners A_y and A_x, and (3) deformation-aware discriminators D_y and D_x. Firstly, modality generators are designed to translate the source image to the target appearance with spatial correspondence preserved, which is implemented with trainable networks G: $x \rightarrow \hat{y} \in \hat{Y}$ and F: $y \rightarrow \hat{x} \in \hat{X}$. Secondly, symmetric spatial aligners A_y and A_x are designed to exploit symmetric correspondence during image-to-image translation to optimise unique and optimal mapping. Each aligner (e.g., $A_y = \{R_y^{\rightarrow}, R_y^{\leftarrow}, T\}$) is enforced to learn bidirectional transformations $\phi_y^{\rightarrow} : \hat{y} \rightarrow y$ and $\phi_y^{\leftarrow} : y \rightarrow \hat{y}$ that are inverse to each other. The bidirectional transformations are learnt through symmetric transformation repressors $R = \{R^{\rightarrow}, R^{\leftarrow}\}$. Each transformation regressor is a CNN model trained to predict a deformation field (Kong et al., 2021), and then followed by a spatial transformer network (Jaderberg et al., 2015) to resample images to target spatial space. Thirdly, deformation-aware discriminators are denoted as $D_y : y \rightarrow \{0, 1\}$ where $y \in U(Y, \hat{Y})$ and $D_x : x \rightarrow \{0, 1\}$ where $x \in U(X, \hat{X})$.

DA-GAN objective To synthesize with misaligned pairs, DA-GAN is constrained by three loss functions, including (1) symmetric registration loss L_{sr} (in Figure 2a) for self-aligning, (2) multi-objective inverse-consistency loss L_{mic} (Figure 2b) for aligning enhancement, and (3) deformation-aware adversarial loss L_{adv_da} (Figure 2c) for synthesis enhancement. Overall, we can formulate our total loss function for DA-GAN as below:

$$\min_{G,F,R} \max_{D} L_{total}(G, F, R, D) = L_{sr} + L_{mic} + L_{adv_da} \tag{1}$$

where $R = \{R_y, R_x\}$ and $D = \{D_y, D_x\}$. These three loss functions will be introduced in the following sections, respectively.

2.2. DA-GAN loss functions

Symmetric registration loss L_{sr} is designed to (1) punish dissimilarity between misaligned imaging pairs and (2) encourage local smoothness on the deformation field. The former similarity loss L_{sim} for symmetric registration is defined as:

$$\begin{aligned} \min_{G,F,R_y,R_x} L_{sim}(G, F, R_y, R_x) = \mathbb{E}_{x,y}[&||y - G(x) \circ \phi_y^{\rightarrow}||_1 + ||G(x) - y \circ \phi_y^{\leftarrow}||_1 \\ &+ ||x - F(y) \circ \phi_x^{\rightarrow}||_1 + ||F(y) - x \circ \phi_x^{\leftarrow}||_1] \end{aligned} \tag{2}$$

where $\phi_y^{\rightarrow} = R_y^{\rightarrow}(G(x), y)$, $\phi_y^{\leftarrow} = R_y^{\leftarrow}(y, G(x))$, $\phi_x^{\rightarrow} = R_x^{\rightarrow}(F(y), x)$, $\phi_y^{\leftarrow} = R_y^{\leftarrow}(x, F(y))$. Secondly, the smoothness loss L_{smt} (Balakrishnan et al., 2019) is implemented to minimize the gradient divergence of the estimated deformation field:

$$\min_{R_y,R_x} L_{smt}(R_y, R_x)) = \mathbb{E}_{x,y}[||\nabla\phi_y^{\rightarrow}||^2 + ||\nabla\phi_y^{\leftarrow}||^2 + ||\nabla\phi_x^{\rightarrow}||^2 + ||\nabla\phi_x^{\leftarrow}||^2] \tag{3}$$

Lastly, by integrating Equation 2 and 3 with loss weights λ_{reg} and λ_{smt}, we can formulate our symmetric registration loss L_{sr} as below:

$$L_{sr} = \lambda_{reg} L_{reg} + \lambda_{smt} L_{smt} \tag{4}$$

Multi-objective inverse-consistency loss L_{mic} is proposed to (1) improve image alignment during symmetric registration, and (2) improve synthesis correspondence to the source image during image generation. As illustrated in Figure 2b, this is achieved by enforcing inverse consistency from three different levels, including registration level, generation level, and joint level.

Firstly, at the registration level, we enforce inverse consistency on the forward and backward transformations $\phi^{\rightarrow}$ and $\phi^{\leftarrow}$ during symmetric registration. Specifically, the registration IC loss L_{ic_reg} is formulated as

$$\min_{R_y,R_x} L_{ic_reg}(R_y,R_x) = E_{x,y}[||y \circ \phi_y^{\leftarrow} \circ \phi_y^{\rightarrow} - y||_1 + ||x \circ \phi_x^{\leftarrow} \circ \phi_x^{\rightarrow} - x||_1] \tag{5}$$

Secondly, at the generation level, we constrain inverse consistency on the two modality generators G and F. Thus, the generation IC loss L_{ic_gen} is formulated as below:

$$\min_{G,F} L_{ic_gen}(G,F) = \mathbb{E}_x[||F(G(x)) - x||_1 + \mathbb{E}_y[||G(F(y)) - y||_1] \tag{6}$$

Lastly, we propose a third joint level inverse consistency through both image registration and generation cycle, thus jointly optimising image registration and generation. The formulation of joint inverse-consistency L_{ic_joint} is shown as below:

$$\begin{aligned}\min_{G,F,R_y,R_x} L_{ic_joint}(G,F,R_y,R_x) = \mathbb{E}_{x,y}[&||F(G(x) \circ \phi_y^{\rightarrow}) \circ \phi_x^{\leftarrow} - x||_1 \\ &+ ||G(F(y) \circ \phi_x^{\rightarrow}) \circ \phi_y^{\leftarrow} - y||_1]\end{aligned} \tag{7}$$

To summarise, the overall multi-objective inverse-consistency loss is composed of three levels of inverse consistency (with their corresponding weights denoted as λ):

$$L_{mic} = \lambda_{ic_reg} L_{ic_reg} + \lambda_{ic_gen} L_{ic_gen} + \lambda_{ic_joint} L_{ic_joint} \tag{8}$$

Deformation-aware adversarial loss L_{adv_da} is designed to disentangle the influence of spatial morphology across domains from intensity characteristic learning. We illustrate the comparison of conventional adversarial loss L_{adv} and our L_{adv_da} for an example discriminator D_y in Figure 2c. Via symmetric spatial aligner A_y, we can obtain source-shaped images and target-shaped images for both generated ($G(x)$ and $G(x) \circ \phi_y^{\rightarrow}$) and real images ($y \circ \phi_y^{\leftarrow}$ and y). Then, we feed all these images to the discriminator in the adversarial process to guide it to focus on learning intensity characteristics only. Formally, deformation-aware adversarial loss for D_y is formulated as:

$$\begin{aligned}\min_{G} \max_{D_y} L^y_{adv_da}(G,D_y) = \mathbb{E}_y[&log(D_y(y)) + log(D_y(y \circ \phi_y^{\leftarrow}))] \\ &+ \mathbb{E}_x[log(1 - D_y(G(x))) + log(1 - D_y(G(x) \circ \phi_y^{\rightarrow}))]\end{aligned} \tag{9}$$

Similarly, we can derive the other half of adversarial loss for D_x as $L^x_{adv_da}$. The total adversarial loss $L_{adv_da} = \lambda_{adv_da} L^y_{adv_da} + \lambda_{adv} L^x_{adv_da}$ where λ_{adv_da} denotes loss weight.

3. Experiments

Simulation dataset For simulation experiments, the public brain T1-T2 MRI dataset (BraTS 2018 (Menze et al., 2014)) was selected because there were well-aligned imaging pairs as ground truth. The training and testing sets contained 5760 and 768 pairs of T1 and T2 images, respectively. The data were normalised to [-1, 1], resized to 256*256, and publicly available [1] (Kong et al., 2021). As original brain images were paired and well-aligned, we simulated 6 different levels of non-affine misalignments on the dataset.

Clinical lung MRI/CT dataset The private lung MRI-CT dataset was used to validate the proposed algorithm for a real-world radiotherapy treatment planning setting. The dataset contained 4096 pairs of ultrashort-echo time MRI (from Siemens scanners) and CT imaging for training and 1024 pairs for independent testing. Both imaging modalities were normalised to [-1, 1], cropped to lung regions, resampled to isotropic, and preliminarily registered. **R-3gB1-Q2:***However, we still observed registration errors in lung regions (Dice 0.949) as well as bones and airways.* Please refer to Appendix A for additional details on both datasets.

Experiment settings In simulation experiments on the brain dataset, 6 non-affine deformations were randomly applied on the training sets to simulate the misalignment. The non-affine deformation was implemented with elastic deformation on control points (Rand2DElastic in MONAI library (Cardoso et al., 2022)) with 6 incremental levels denoted as NA-1 to NA-6. In real-world experiments on the lung dataset, DA-GAN was compared on 9 state-of-the-art (SOTA) medical synthesis methods, including GAN (Goodfellow et al., 2020), Pix2pix (Isola et al., 2017), CycleGAN (Zhu et al., 2017), UNIT (Liu et al., 2017), MUNIT (Huang et al., 2018), NiceGAN (Chen et al., 2020), RegGAN-NC (Kong et al., 2021), RegGAN-C (Kong et al., 2021), **R-3gB1-Q3:***and additional baseline using multimodal registration CorrField (ranked 2nd best in learn2reg challenge (Hering et al., 2022)) prior to Pix2pix*. Please find more details in Appendix B.1.

All experiments were implemented in Pytorch on a 64-bit Ubuntu Linux system with one 16 GB Nvidia P100 GPU. We trained all the methods using the Adam optimiser with the learning rate 1e-4 and $(\beta_1, \beta_2) = (0.5, 0.999)$. The batch size was set to 1 with weight decay 1e04. The training included 50 epochs for both datasets. The brain imaging was evaluated with three metrics in 2D, including Normalized Mean Absolute Error (NMAE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM). The lung imaging was evaluated with MAE3D, PSNR3D, and SSIM3D. The background of the image was excluded from the computation. For reproducibility, we included more implementation details on DA-GAN modules and loss weighting in Appendix B.2.

4. Results

Results on the simulation experiments This section summarises the results of DA-GAN on brain imaging with 6 non-affine misalignments (Table 1) and visualisation on the testing set (Figure 3). Table 1 shows that DA-GAN consistently outperformed all comparison methods in three metrics from misalignment levels NA-1 to NA-6. The results of DA-GAN remained stable with increased levels of non-affine misalignment with NMAE ranging from

1. https://drive.google.com/file/d/1PiTzGQEVV7NO4nPaHeQv61WgDxoD76nL/view

Table 1: Results on the brain dataset with 6 simulated non-affine misalignments.

Non-affine (NA)	NA-1			NA-2			NA-3		
Methods	NMAE	PSNR	SSIM	NMAE	PSNR	SSIM	NMAE	PSNR	SSIM
GAN	0.111	20.12	0.784	0.102	21.73	0.824	0.162	16.65	0.700
Pix2pix	0.091	21.71	0.807	0.101	17.88	0.776	0.094	20.76	0.808
CycleGAN	0.087	23.35	0.825	0.091	23.41	0.817	0.093	22.75	0.836
RegGAN-NC	0.074	24.84	0.854	0.073	25.13	0.852	0.076	24.82	0.847
RegGAN-C	0.078	24.24	0.850	0.079	24.30	0.850	0.076	23.22	0.853
DA-GAN	**0.074**	**24.97**	**0.859**	**0.070**	**25.25**	**0.861**	**0.073**	**24.89**	**0.857**
Non-affine (NA)	NA-4			NA-5			NA-6		
Methods	NMAE	PSNR	SSIM	NMAE	PSNR	SSIM	NMAE	PSNR	SSIM
GAN	0.094	20.79	0.815	0.106	21.95	0.813	0.100	21.54	0.813
Pix2pix	0.120	17.35	0.746	0.112	18.34	0.756	0.116	15.12	0.761
CycleGAN	0.116	22.23	0.823	0.086	23.47	0.833	0.108	22.92	0.786
RegGAN-NC	0.077	24.46	0.839	0.082	22.16	0.828	0.081	22.33	0.817
RegGAN-C	0.080	23.54	0.850	0.076	23.36	0.848	0.082	22.88	0.841
DA-GAN	**0.072**	**24.58**	**0.858**	**0.075**	**24.72**	**0.858**	**0.073**	**24.61**	**0.858**

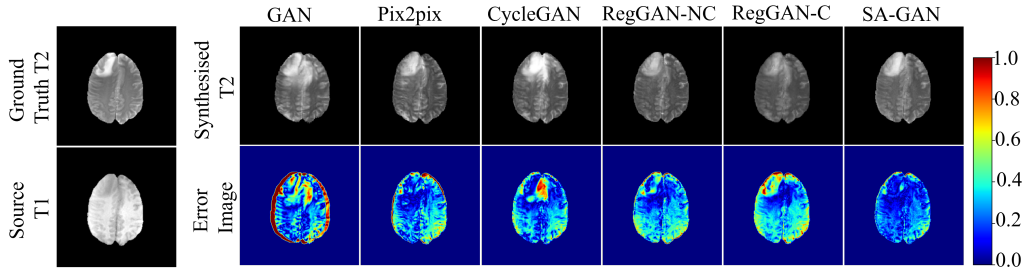

Figure 3: Visualisation of predicted and error images in the simulation experiment (NA-3).

0.070 to 0.075. Figure 3 visualises that our DA-GAN achieved less error compared with other methods on the error map.

Results on the lung MRI-CT dataset Table 2 shows the quantitative results of DA-GAN on the lung MRI-CT with challenging respiratory-motion misalignment with visualisation shown in Figure 4. The results in Table 2 show that our DA-GAN achieved the best results of MAE3D 35.86, PSNR3D 32.49 and SSIM3D 0.731 compared with 9 SOTA methods. Figure 4 further visually highlights the superiority of our DA-GAN, especially in the challenging spine, bones and heart regions (blue arrows). These demonstrate the great potential for synthetic CT to be used for MRI-only radiotherapy.

Ablation study on DA-GAN losses **R-3gB1-C2:** *The ablation study in Table 3 shows that L_{adv_da} improved the baseline RegGAN in all metrics (e.g., PSNR-3D 0.41), while L_{mic} further improved the results in all metrics (e.g., PSNR-3D 0.36). The paired t-test shows that the contributions from both losses were statistically significant (p-value<0.001) in MAE2D, PSNR2D, SSIM2D.*

Table 2: Performance comparison on the real-world lung dataset. Abbr: SE = SimpleElastix, SG = Structure-guided, SyN = ANTs SyN.

Pre-alignment	Methods	MAE3D	PSNR3D	SSIM3D
R-3gB1-Q3:*CorrField*	Pix2pix	48.76 ± 10.87	29.55 ± 1.77	0.708 ± 0.021
SE+SG+SyN	GAN	330.66 ± 21.59	17.00 ± 0.32	0.612 ± 0.035
SE+SG+SyN	Pix2pix	39.64 ± 8.31	31.88 ± 1.48	0.710 ± 0.027
SE+SG+SyN	CycleGAN	49.32 ± 4.12	30.14 ± 0.82	0.691 ± 0.015
SE+SG+SyN	UNIT	267.9 ± 15.15	18.51 ± 0.32	0.617 ± 0.034
SE+SG+SyN	MUNIT	39.78 ± 6.56	31.62 ± 1.01	0.705 ± 0.024
SE+SG+SyN	NiceGAN	42.73 ± 8.39	31.21 ± 1.61	0.711 ± 0.022
SE+SG+SyN	RegGAN-NC	39.45 ± 5.07	31.86 ± 1.00	0.710 ± 0.021
SE+SG+SyN	RegGAN-C	40.77 ± 4.59	31.72 ± 0.97	0.710 ± 0.017
SE+SG+SyN	**DA-GAN**	**35.86 ± 6.97**	**32.49 ± 1.38**	**0.731 ± 0.023**

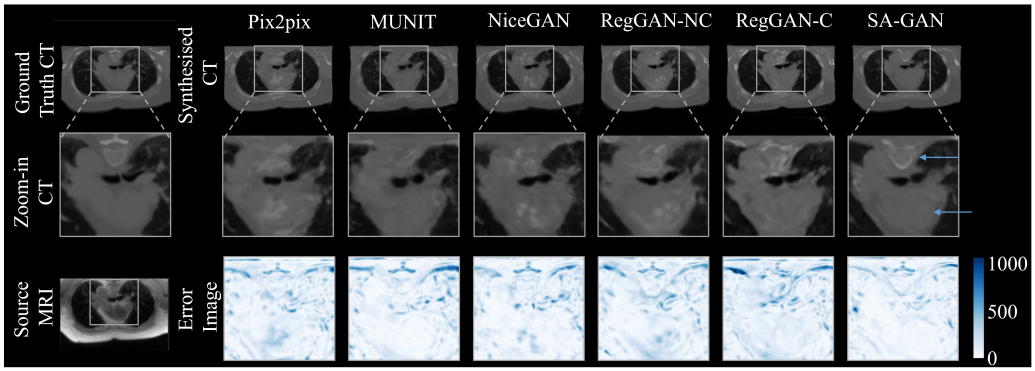

Figure 4: Visualization of synthesised images and error maps on the lung dataset. The blue arrows indicate our DA-GAN achieved better visual results at spine and heart.

Table 3: Ablation study on the proposed losses L_{mic} and L_{adv_da}.

	MAE2D	PSNR2D	SSIM2D	MAE3D	PSNR3D	SSIM3D
RegGAN-C	0.011	70.23	0.918	40.77	31.72	0.710
DA-GAN (L_{adv_da})	0.010	74.82	0.925	37.23	32.13	0.719
DA-GAN ($L_{adv_da} + L_{mic}$)	0.009	76.36	0.929	35.86	32.49	0.725

5. Conclusions

In this study, we introduce a new DA-GAN for medical image synthesis with substantially misaligned imaging pairs. We propose two novel loss functions L_{mic} and L_{adv_da} to generate high-fidelity images while adaptively learning correspondence via symmetric registration. We validated our method on a public brain dataset with both 6 simulated misalignments and a real-world lung dataset compared with 9 SOTA methods. The results demonstrated the potential towards an important step in generalisable medical image synthesis with limited data for clinical applications such as diagnosis and treatment planning.

References

Badour AlBahar and Jia-Bin Huang. Guided image-to-image translation with bi-directional feature transformation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9016–9025, 2019.

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. Monai: An open-source framework for deep learning in healthcare, 2022.

Runfa Chen, Wenbing Huang, Binghui Huang, Fuchun Sun, and Bin Fang. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8168–8177, 2020.

G.E. Christensen and H.J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20(7):568–582, 2001. doi: 10.1109/42.932742.

Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, Kun Zhang, and Dacheng Tao. Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2427–2436, 2019.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Alessa Hering, Lasse Hansen, Tony CW Mok, Albert CS Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, et al. Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, 42(3):697–712, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018.

Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.

Saeed Khorram and Li Fuxin. Cycle-consistent counterfactuals by latent transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10203–10212, 2022.

Lingke Kong, Chenyu Lian, Detian Huang, Yanle Hu, Qichao Zhou, et al. Breaking the dilemma of medical image-to-image translation. *Advances in Neural Information Processing Systems*, 34:1964–1978, 2021.

Shivani Kumar, Robba Rai, Alto Stemmer, Sonal Josan, Lois Holloway, Shalini Vinod, Daniel Moses, and Gary Liney. Feasibility of free breathing lung mri for radiotherapy using non-cartesian k-space acquisition schemes. *The British journal of radiology*, 90(1080): 20170037, 2017.

Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.

Xiao Liu, Pedro Sanchez, Spyridon Thermos, Alison Q O'Neil, and Sotirios A Tsaftaris. Learning disentangled representations in the imaging domain. *Medical Image Analysis*, page 102516, 2022.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

John P Mugler III, Samuel W Fielden, Craig H Meyer, T Altes, G Wilson Miller, Alto Stemmer, et al. Breath-hold ute lung imaging using a stack-of-spirals acquisition. In *Proc Intl Soc Mag Reson Med*, volume 23, page 1476, 2015.

Yingxue Pang, Jianxin Lin, Tao Qin, and Zhibo Chen. Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, 24:3859–3881, 2021.

Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European*

Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, pages 319–345. Springer, 2020.

Zengming Shen, S Kevin Zhou, Yifan Chen, Bogdan Georgescu, Xuqi Liu, and Thomas Huang. One-to-one mapping for unpaired image-to-image translation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1170–1179, 2020.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

Maria Francesca Spadea, Matteo Maspero, Paolo Zaffino, and Joao Seco. Deep learning based synthetic-ct generation in radiotherapy and pet: a review. *Medical physics*, 48(11): 6537–6566, 2021.

Chao Wang, Haiyong Zheng, Zhibin Yu, Ziqiang Zheng, Zhaorui Gu, and Bing Zheng. Discriminative region proposal adversarial networks for high-quality image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 770–785, 2018a.

Chengjia Wang, Guang Yang, Giorgos Papanastasiou, Sotirios A Tsaftaris, David E Newby, Calum Gray, Gillian Macnaught, and Tom J MacGillivray. Dicyc: Gan-based deformation invariant cross-domain information fusion for medical image synthesis. *Information Fusion*, 67:147–160, 2021.

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018b.

Yanwu Xu, Shaoan Xie, Wenhao Wu, Kun Zhang, Mingming Gong, and Kayhan Batmanghelich. Maximum spatial perturbation consistency for unpaired image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18311–18320, 2022.

Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. Modulated contrast for versatile image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18280–18290, 2022.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

Appendix A. Dataset details

Brain T1-T2 dataset On the brain T1-T2 MRI dataset, we simulated 6 different levels of non-affine misalignments. A visualisation of non-affine misalignment is provided in Figure 5. An overview of implementation parameters for non-affine misalignment is provided in Table 4. More specifically, the non-affine misalignment was simulated using elastic deformation on control points (Rand2DElastic in MONAI library[2]). The spacing between control points was set to [40, 40], while the magnitude was set to incremental levels from NA-1 to NA-6.

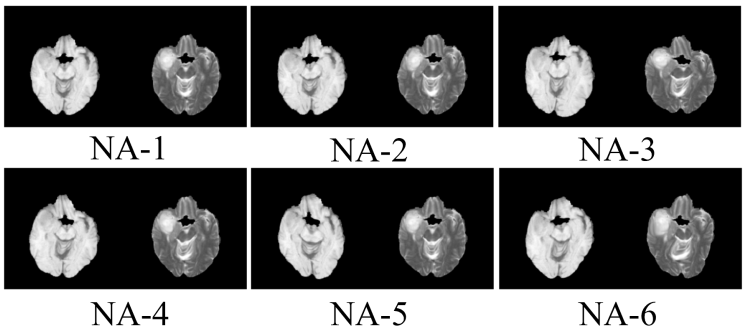

Figure 5: Example images with different levels of non-affine misalignment.

Table 4: Implementation of non-affine misalignments

Non-affine (NA)	**NA-1**	**NA-2**	**NA-3**	**NA-4**	**NA-5**	**NA-6**
Spacing	[40, 40]	[40, 40]	[40, 40]	[40, 40]	[40, 40]	[40, 40]
Magnitude	[1, 2]	[2, 3]	[3,4]	[4, 5]	[5, 6]	[6, 7]

Lung MRI-CT dataset The lung MRI-CT dataset contained paired but misaligned MRI and CT from 20 patients with lung diseases. The lung MRI was implemented with a prototype 3D free-breathing stack-of-spirals UTE VIBE sequence, provided by Siemens Healthineers (Mugler III et al., 2015; Kumar et al., 2017). UTE-MRI provided millimetre resolution and radiation-free assessment for pulmonary structural imaging (Dournes et al., 2016). 20 CTs were scanned on approximately the same day as MRI. 20 patients included 12 patients with cystic fibrosis and 8 patients with lung cancer. **R-3gB1-Q1:***For the MRI protocol, both cystic fibrosis and lung cancer cohorts used Siemens 3T scanners with flip angle 5, echo time 0.05, repetition time 2.97-3.78. For details of image size and spacing, please refer to Table 5. Both MRI and CT were normalised to [-1, 1], cropped to lung regions, resampled to isotropic spacing, and preliminarily registered.* The ground-truth paired MRI-CT images were acquired via an automatic registration pipeline including elastic registration (SimpleElastix library), structure-guided registration (demon registration in ants library) and antsRegistrationSyNQuick (ants library). **R-3gB1-Q2:***After registration, we still observed*

2. https://docs.monai.io/en/stable/transforms.html#rand2delastic

Table 5: Physical details of the lung MRI-CT dataset

Lung cancer	Image size	No. slice	Pixel spacing	Slice spacing
MRI-spiral vibe	320*320	176-210	1.5*1.5	1.5
CT	512*512	181-234	1.71*1.71	2
Cystic fibrosis	Image size	No. slice	Pixel spacing	Slice spacing
MRI-spiral vibe	512*512 or 416*416	160-241	1.1 *1.1 or 1.25*1.25	1.1 or 1.25
CT	512*512	267-902	0.46*0.46-0.78*0.78	0.4-1

registration errors on the whole lung regions with Dice 0.949 (Dice for right lung 0.947, Dice for right lung 0.950) and registration errors in bones and airways. We acquired relevant ethics approval for the study, and informed written consent from the parent or legal guardian of each child.

Appendix B. Implementation details

In this section, we first discuss the implementation of other comparison methods. Then we introduce network implementation details for different modules and loss weighting in DA-GAN.

B.1. Implementation of comparison methods

CorrField is an unsupervised multimodal registration baseline, which estimates estimates sparse correspondences on image-based Förstner keypoints using exact message passing on a minimum spanning tree. The open-source code is on https://grand-challenge.org/algorithms/corrfield/.

GAN uses a generator to translate the source image to the target space, while using a discriminator to determine whether images were generated for from the real data. In this way, GANs iteratively improve the image fidelity via min-max game. In the implementation, the generator and discriminator shared the same network architecture as ours for a fair comparison. The open-source code is on https://github.com/Kid-Liet/Reg-GAN.

Pix2pix is a typical supervised GAN using L1 loss functions to enforce pixel-wise similarity between the predicted images and ground-truth images. In the implementation, the generator and discriminator shared the same network architecture as ours for a fair comparison. The open-source code is on https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

CycleGAN proposes a cycle consistency loss to constrain two generators that are reverse to each other. A cycleGAN consists of two sets of generators and discriminators. In the implementation, the generator and discriminator shared the same network architecture as ours for a fair comparison. The open-source code is on https://github.com/junyanz/pytorchCycleGAN-and-pix2pix.

UNIT leverages the assumption that the latent coding space is shared by different modalities for image-to-image translation. The open-source code is on https://github.com/mingyuliutw/UNIT.

MUNIT disentangles the representation to a content representation and a style representation for image-to-image translation. The open-source code is on https://github.com/NVlabs/MUNIT.

NICE-GAN is a compact and effective network architecture for image-to-image translation, which is achieved by reusing the discriminator for encoding. The open-source code is on https://github.com/alpc91/NICE-GAN-pytorch.

RegGAN-NC is a variant of RegGAN (without cycle consistency). It incorporates a registration module to adaptively fit misaligned distribution. Specifically, it consists of a generator, a registration module and a discriminator. The open-source code is on https://github.com/Kid-Liet/Reg-GAN.

RegGAN-C is a variant of RegGAN with Cycle consistency. It consists of two generators, two discriminators and one registration module. The open-source code is on https://github.com/Kid-Liet/Reg-GAN.

B.2. Implementation of DA-GAN

DA-GAN was implemented in PyTorch with Python 3.7. The network architecture of DA-GAN consists of two modality generators, two symmetric spatial aligners, and two deformation-aware discriminators.

Modality generator Each modality generator uses 2 downsampling layers, 9 residual blocks, and 2 upsampling layers, following the implementation of Johnson et al. (Johnson et al., 2016). Specifically, we use Ck to denote 7*7 convolution-InstanceNorm-ReLU layer with k filters and stride 1, Dk to denote the downsampling 3*3 convolution-instanceNorm-ReLU layer with k filters and stride 2, Rk to denote a residual block containing two 3*3 convolutional layers with same number of filters, Uk to denote the upsampling 3*3 fractional-strided-convolution-instanceNorm-ReLU layer with k filters and stride 0.5. The architecture of a modality generator is C64-D128-D256-R256-R256-R256-R256-R256-R256-R256-R256-R256-U128-U64-C1.

Symmetric spatial aligner Each symmetric spatial aligner consists of 4 transformation regressors and 4 spatial transformer networks. Each transformation regressor uses ResUnet architecture, containing 7 encoder layers, 3 residual blocks, 7 decoder layers, and skip connections, following the implementation of RegGAN (Kong et al., 2021). Specifically, we use Dk to denote downsampling 3*3 convolution-LeakyReLU (LeakyReLU with a slope of 0.2) with k filters and stride 1, Rk to denote a residual block containing two 3*3 convolutional layers, Uk to denote upsampling 3*3 convolution-LeakyReLU (LeakyReLU with a slope of 0.2) with k filters and stride 1. The backbone architecture of the transformation regressor is D32-D64-D64-D64-D64-D64-D64-R64-R64-R64-U64-U64-U64-U64-U64-U32, followed by a refinement layer (a residual block and a 1*1 convolutional layer) and an output layer (3*3 convolution layer).

Deformation-aware discriminator For deformation-aware discriminator, it was implemented with a 70*70 PatchGAN (Isola et al., 2017). Specifically, If we denote a 4*4 convolution-instanceNorm-LeakyReLU layer with k filters and stride 2 as Ck (LeakyReLU with a slope of 0.2). The architecture of the discriminator is C64-C128-C256-C512. After the last layer, another convolution is applied to generate 1-dimensional classification results. No InstanceNorm is used for the first $C64$ layer.

Loss weights Lastly, the weights for different losses in DA-GAN were shown in Table 6, which followed RegGAN (Kong et al., 2021) for a fair comparison.

Table 6: Weights for each loss function in DA-GAN.

Loss	L_{sc}		L_{mic}			L_{adv_da}
Weights	λ_{reg}	λ_{smt}	λ_{ic_reg}	λ_{ic_gen}	λ_{ic_joint}	λ_{adv_da}
Values	20	10	10	10	10	1

Appendix C. Convergence analysis during training

Figure 6 shows the validation NMAE during the training process for different levels of non-affine misalignments in the brain dataset. The results demonstrate that DA-GAN successfully converged under different levels of non-affine misalignments.

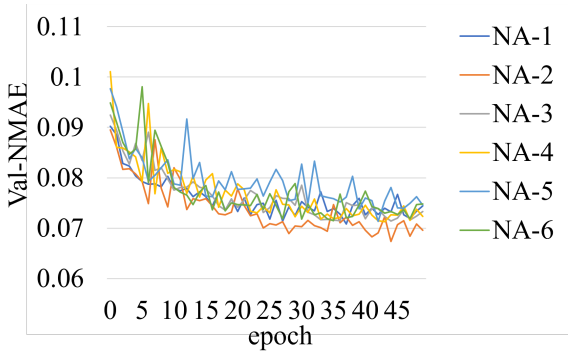

Figure 6: Convergence analysis of DA-GAN on brain dataset with non-affine misalignments.

Appendix D. Comparison of complexity and accuracy

R-tPyS-Q1: *As shown in Table 7, our model achieves better synthesis accuracy (9.1% increase of MAE3D) at the cost of additional complexity (doubled no. parameters and size of the model) compared with RegGAN-NC. However, we must emphasize that synthesis accuracy is clinically critical in radiotherapy planning because it reduces unnecessary radiation-related toxicity in patients and is vital to patients' safety (Burnet et al., 2004). Further, our model complexity is still smaller than state-of-the-art GANs (e.g., MUNIT and NiceGAN), and our inference time is fast (0.013s fps).*

Appendix E. Sensitivity analysis on hyperparameters

R-tPyS-W2: *We conducted a sensitivity analysis on hyperparameters of loss weights on the lung MRI-CT dataset. As shown in Table 8, the results (fluctuating slightly from 0.009 to*

Table 7: Comparison of complexity and accuracy of the generative models on the lung MRI-CT dataset

Lung MRI-CT data	MUNIT	NiceGAN	RegGAN-NC	RegGAN-C	Ours
MAE3D	39.8	42.7	39.5	40.8	35.9
No. Parameters (M)	44.3	112.3	16.2	30.3	36.5
Size of the model (mb)	169.0	428.3	61.8	115.7	139.2
fps (s)	0.02	0.10	0.01	0.01	0.01

0.010) indicate that the model performance is robust across different choices of hyperparameter values.

Table 8: Sensitivity analysis on hyperparameters of loss weights on the lung MRI-CT dataset

lambda_reg	MAE2D	lambda_smt	MAE2D	lambda_ic	MAE2D
5	0.009	1	0.011	1	0.010
10	0.010	5	0.010	5	0.010
20	0.009	10	0.009	10	0.009
30	0.010	20	0.010	20	0.009

Appendix F. Complete analysis for ablation study

R-3gB1-C2: *Table 3 shows the results of an ablation study on the multi-objective inverse consistency loss L_{mic} and deformation-aware adversarial loss L_{adv_da}. Firstly, for L_{mic}, we experimented on the 7 settings on different combinations of $\{L_{ic_reg}, L_{ic_gen}, L_{ic_joint}\}$. As shown in Table 3, by comparing the settings A-F and G2, we concluded that our proposed L_{mic} that was composed of all three IC losses achieved the best results. Secondly, for L_{adv_da}, we compared DA-GAN using conventional adversarial loss L_{adv} (G1) and deformation-aware adversarial loss L_{adv_da} (G2). The results show that L_{adv_da} outperformed its counterpart. Overall, the ablation study demonstrated the effectiveness of both proposed components L_{mic} and L_{adv_da}.*

Appendix G. Inverse consistency in registration

R-3gB1-C4: *Inverse consistency is often implemented by symmetrising cost functions (Christensen and Johnson, 2001) or computing the cost function in "mid-space" (Reuter et al., 2010). A latest work proposed a multi-step IC registration to allow for coarse-to-fine registration ((Greer et al., 2023)). As our work primarily focuses on image synthesis with better-handled misalignment, we design a single-step multi-objective IC loss which jointly optimises the image generation and registration. For efficiency reasons, our solution does not rely on scaling and squaring technique and its costly numerical integration. The results show that the proposed IC loss significantly contributes to the improved generative performance (p-value < 0.001)*

Table 9: Ablation study on the proposed losses L_{mic} and L_{adv_da}.

	L_{ic_reg}	L_{ic_gen}	L_{ic_joint}	L_{adv}	L_{adv_da}	MAE3D	PSNR3D	SSIM3D
A	Y				Y	38.88±5.86	31.78±1.30	0.713±0.018
B		Y			Y	37.23±6.00	32.13±1.21	0.719±0.021
C			Y		Y	40.99±7.53	31.52±1.35	0.709±0.023
D	Y	Y			Y	41.15±7.31	31.73±1.31	0.712±0.021
E	Y		Y		Y	40.08±7.72	31.66±1.22	0.712±0.024
F		Y	Y		Y	36.86±7.58	32.32±1.43	0.720±0.026
G1	Y	Y	Y	Y		40.57±12.41	31.72±2.34	0.721±0.027
G2	**Y**	**Y**	**Y**		**Y**	**35.86±6.91**	**32.49±1.38**	**0.725±0.024**

on misaligned data. **R-tPyS-D2:***In future work, we will investigate the influence of better regularisation methods, such as Total Variation Regularisation (Vishnevskiy et al., 2016).*

Reference

Burnet, N. G., S. J. Thomas, K. E. Burton, and S. J. Jefferies (2004). Defining the tumour and target volumes for radiotherapy. *Cancer Imaging 4*(2), 153.

Christensen, G. E. and H. J. Johnson (2001). Consistent image registration. *IEEE transactions on medical imaging 20*(7), 568–582.

Dournes, G., F. Menut, J. Macey, M. Fayon, J.-F. Chateil, M. Salel, O. Corneloup, M. Montaudon, P. Berger, and F. Laurent (2016). Lung morphology assessment of cystic fibrosis using mri with ultra-short echo time at submillimeter spatial resolution. *European Radiology 26*, 3811–3820.

Greer, H., L. Tian, F.-X. Vialard, R. Kwitt, S. Bouix, R. San Jose Estepar, R. Rushmore, and M. Niethammer (2023). Inverse consistency by construction for multistep deep registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 688–698. Springer.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134.

Johnson, J., A. Alahi, and L. Fei-Fei (2016). Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pp. 694–711. Springer.

Kong, L., C. Lian, D. Huang, Y. Hu, Q. Zhou, et al. (2021). Breaking the dilemma of medical image-to-image translation. *Advances in Neural Information Processing Systems 34*, 1964–1978.

Reuter, M., H. D. Rosas, and B. Fischl (2010). Highly accurate inverse consistent registration: a robust approach. *Neuroimage 53*(4), 1181–1196.

Vishnevskiy, V., T. Gass, G. Szekely, C. Tanner, and O. Goksel (2016). Isotropic total variation regularization of displacements in parametric image registration. *IEEE transactions on medical imaging 36*(2), 385–395.