



# Harmonizing three-dimensional MRI using pseudo-warping field guided GAN

Jiaying Lin<sup>a</sup>, Zhuoshuo Li<sup>a</sup>, Youbing Zeng<sup>a</sup>, Xiaobo Liu<sup>a</sup>, Liang Li<sup>b</sup>, Neda Jahanshad<sup>c</sup>,  
Xinting Ge<sup>d</sup>, Dan Zhang<sup>e,\*</sup>, Minhua Lu<sup>f,\*</sup>, Mengting Liu<sup>a,\*</sup>, for the Alzheimer's Disease  
Neuroimaging Initiative<sup>1</sup>

<sup>a</sup> Department of Biomedical Engineering, Sun Yat-sen University, Shenzhen 518107, China

<sup>b</sup> Genuine Digital Technology Co., Ltd., Xi'an, China

<sup>c</sup> Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

<sup>d</sup> School of Information Science and Engineering, Shandong Normal University, Shandong 250358, China

<sup>e</sup> School of Cyber Science and Engineering, Ningbo University of Technology, Zhejiang 315211, China

<sup>f</sup> Department of Biomedical Engineering, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Shenzhen 518060, China

## ARTICLE INFO

Dataset link: [https://github.com/lx123-j/PWFH\\_armonization](https://github.com/lx123-j/PWFH_armonization)

### Keywords:

Slice-wise 3D MRI harmonization

GAN

Pseudo-warping field

Brain MRI

## ABSTRACT

In pursuit of cultivating automated models for magnetic resonance imaging (MRI) to aid in diagnostics, an escalating demand for extensive, multisite, and heterogeneous brain imaging datasets has emerged. This potentially introduces biased outcomes when directly applied for subsequent analysis. Researchers have endeavored to address this issue by pursuing the harmonization of MRIs. However, most existing image-based harmonization methods for MRI are tailored for 2D slices, which may introduce inter-slice variations when they are combined into a 3D volume. In this study, we aim to resolve inconsistencies between slices by introducing a pseudo-warping field. This field is created randomly and utilized to transform a slice into an artificially warped subsequent slice. The objective of this pseudo-warping field is to ensure that generators can consistently harmonize adjacent slices to another domain, without being affected by the varying content present in different slices. Furthermore, we construct unsupervised spatial and recycle loss to enhance the spatial accuracy and slice-wise consistency across the 3D images. The results demonstrate that our model effectively mitigates inter-slice variations and successfully preserves the anatomical details of the images during the harmonization process. Compared to generative harmonization models that employ 3D operators, our model exhibits greater computational efficiency and flexibility.

## 1. Introduction

Neuroimaging investigations frequently yield incongruous and occasionally conflicting outcomes owing to the constraints of limited sample sizes. The simultaneous proliferation in both the number and scale of multi-site MR imaging studies has engendered additional challenges concerning the management of extensive data volumes and the comprehension of heightened imaging variability. These challenges become particularly prominent when aggregating data acquired from diverse research facilities or through different acquisition protocols (Nieuwenhuis et al., 2017). Nevertheless, MR images obtained from different

sites inherently introduce discrepancies, which arise from variations in acquisition protocols, scanner manufacturers, hardware imperfections, and other contributing factors. Insufficient comprehension and inadequate alleviation of site effects can potentially yield deceptive or untrustworthy outcomes in a wide array of applications.

Hence, in the consolidation of multiple smaller datasets into a larger composite sample, image harmonization assumes a pivotal role in yielding robust conclusions (Keshavan et al., 2016). Leveraging its capability to mitigate or compensate for redundant inter-site disparities, harmonization ensures data homogeneity and comparability throughout statistical analyses. Several techniques have shown good

\* Corresponding authors.

E-mail addresses: [linjy277@mail2.sysu.edu.cn](mailto:linjy277@mail2.sysu.edu.cn) (J. Lin), [lizhsh23@mail2.sysu.edu.cn](mailto:lizhsh23@mail2.sysu.edu.cn) (Z. Li), [zengyb6@mail2.sysu.edu.cn](mailto:zengyb6@mail2.sysu.edu.cn) (Y. Zeng), [liuxb26@mail2.sysu.edu.cn](mailto:liuxb26@mail2.sysu.edu.cn) (X. Liu), [liliang@genuai.com](mailto:liliang@genuai.com) (L. Li), [njahansh@usc.edu](mailto:njahansh@usc.edu) (N. Jahanshad), [xintingge@163.com](mailto:xintingge@163.com) (X. Ge), [danzhang@nbut.edu.cn](mailto:danzhang@nbut.edu.cn) (D. Zhang), [luminhua@szu.edu.cn](mailto:luminhua@szu.edu.cn) (M. Lu), [liumt55@mail.sysu.edu.cn](mailto:liumt55@mail.sysu.edu.cn) (M. Liu).

<sup>1</sup> Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_Apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_Apply/ADNI_Acknowledgement_List.pdf).

<https://doi.org/10.1016/j.neuroimage.2024.120635>

Received 26 February 2024; Received in revised form 15 April 2024; Accepted 2 May 2024

Available online 8 May 2024

1053-8119/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

performance in statistical magnetic resonance imaging (MRI) harmonization. A common approach is ComBat (Fortin et al., 2018) and its variants, which statistically approximates the distribution of features derived from the image, and then makes adjustments. However, ComBat is unable to model heteroscedastic distributions and large sample differences between sites (Bayer et al., 2022), and is not suitable for multi-task scenarios. Alternatively, harmonization is directly conducted on MR images. These methods typically rely on deep learning-based generative techniques and formulate the harmonization as an image-to-image translation task, like CALAMITI (Zuo et al., 2021), MISTGAN (Raju et al., 2021), MURD (Liu and Yap, 2024), etc. The image-based methods (Wu et al., 2024; Guan et al., 2022) can effectively address the issue of cross-site variability, so the harmonization problem is often attributed to differences in style and domain.

Although some existing approaches have performed well on MRI harmonization tasks, they mainly restrict harmonization training to 2D slices rather than the entire 3D volume (Dinsdale et al., 2021; Sinha et al., 2021; Liu et al., 2021, 2023), which often results in discontinuities between adjacent slices and may not be well-suited for 3D medical images.

There are several challenges in the extension of 2D MRI harmonization methods to 3D images. (1) 3D images contain a much larger number of voxels; (2) Compared to the two-dimensional scenario, the quantity of three-dimensional training instances is relatively limited; (3) Many deep learning methods, when applied to computing 2D models, exhibit significant computational demands. Transitioning to 3D models inevitably entails even greater computational complexity. Komandur et al. (2023) presented an unsupervised 3D CycleGAN approach that uses 3D convolutional generators and discriminators in the model to harmonize entire 3D MRI scans. Hong et al. (2021) have extended the capabilities of StyleGAN2 (Karras et al., 2019) by developing 3D-StyleGAN, which enables style-mixing to transform style variations across different sites, including contrast and more. However, running these 3D models requires high computational resources, i.e. GPU configuration and memory, and thus cannot be extensively applied. In response to the demanding computational requirements of the attention-based transformer block, Zhang et al. (2022) have introduced PTNet3D, a framework that seamlessly integrates a performer block. This adept incorporation effectively alleviates the computational challenges presented by the transformer block. Nevertheless, when compared to the 2D domain, the synthesis of images in 3D space may engender a greater probability of anatomical detail distortion, thereby rendering the assurance of image quality more challenging. Cackowski et al. (2021) introduced ImUnity, a 2.5D model derived from VAE-GAN. This model integrates outputs from three models, each trained along a specific axis. While this method effectively addresses discontinuity issues in 2D generative models, it necessitates computation along all three axes.

To tackle these challenges, a feasible approach involves processing the 3D images as a series of 2D slices, and implementing extra slice-wise regularization techniques to help mitigate inconsistencies between slices. An effective strategy for ensuring consistency across slices could involve integrating information from adjacent slices. This might be achieved through deformation estimation methods that track changes from one slice to the next, accounting for both non-biologically meaningful and biologically relevant alterations. The similar technique is frequently utilized in video translation (Ilg et al., 2016; Teed and Deng, 2020; Chen et al., 2019; Bansal et al., 2018; Wang et al., 2022), where motion estimation is employed to track the movement of identical visual elements through successive frames. Nevertheless, this approach has not been used in the realm of MRI imaging. The key difference lies in the nature of 3D MRI, unlike video frames, there is no continuous “object” present across consecutive slices, despite the apparent similarity in brain tissues. Therefore, the concept of motion, as defined and used in video processing, is not directly applicable to 3D MRI analysis.

In this study, we propose to harmonize 3D MR images by utilizing 2D slices and implementing a slice-wise consistency regularization, guided by a pseudo-warping field. This technique effectively resolves inter-slice inconsistencies without resorting to potentially unreasonable and inaccurate estimation of inter-slice deformations. Our method entails generating a random warping field for each slice in a 3D MRI, instead of calculating the real deformation from one slice to the next. This artificial field is then utilized to produce a simulated subsequent slice, which is essentially a spatially warped version of the existing slice. Similarly, for the equivalent slice in the target domain, originating from the source slice, we fabricate its simulated subsequent slice using the identical warping field. Next, we apply spatial reconstruction and slice-wise consistency regularization (Wang et al., 2022) to the pair of warped slices, ensuring consistent deformation across both domains. It should be emphasized that the pseudo-warping field is not designed to accurately predict the next slice in any given domain, as this would be unfeasible in actual applications. Instead, the primary objective of this field is to uphold the consistency of the regularization process. This is achieved by ensuring uniform deformation across different domains, which in turn supports the enforcement of both spatial accuracy and slice-wise consistency. Essentially, our method focuses solely on individual slices yet enhances inter-slice consistency during the harmonization process. Moreover, by eliminating the necessity for true warping estimation, our approach achieves greater computational efficiency.

Our method was applied on three large cohorts of MR images, including Alzheimer’s Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005), UK Biobank (UKBB) (Sudlow et al., 2015) and Nathan Kline Institute-Rockland Sample (NKI-RS) (Tobe et al., 2021) for qualitative, quantitative, and downstream task evaluations. The approach presents a practical and efficient means of harmonizing 3D MRI images.

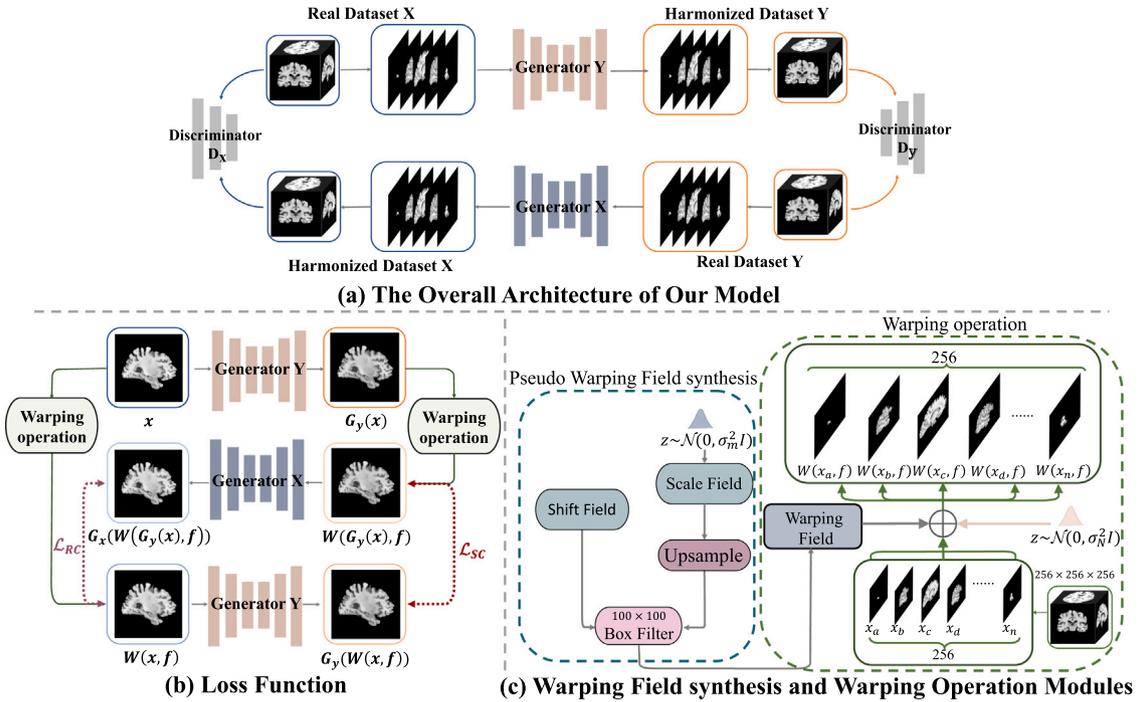
In summary, the main contributions of this work are as follows:

1. We propose an efficient DL-based harmonization model for 3D MRI images, which introduces a pseudo-warping field to enhance slice consistency and eliminates the requirement of paired images for training.
2. To ensure consistency between slices before and after harmonization, we employ unsupervised spatial and recycle loss for improved spatial accuracy.
3. Our model has relatively low GPU configuration requirements, making it easily usable in environments with limited computing resources, and provides flexibility regarding available data.

## 2. Method

### 2.1. Motivation

Traditional 2D-models focus on maintaining the reconstruction consistency of single images, there is potential for slice-wise inconsistency. This is because the generator might respond differently to adjacent slices due to their varying content. Our model, however, incorporates a warping operation to direct the generator in producing images that remain unaffected by this warping. Notably, the deformed slices are arbitrarily warped from the actual slice. Theoretically, this warping field could encompass deformations from one slice to its actual subsequent ones, and even from a slice to itself. The fundamental principle of our model is that, regardless of whether or how the slice changes, the generator is trained to consistently convert them into another domain. This ensures that all slices in an MR image are converted in a uniformly consistent manner.



**Fig. 1.** The architecture of our 3D-translation model. (a) Generator X and Generator Y are two generators which translate images between datasets X and Y. A discriminator is used to verify the authenticity of the generated images. (b) The warping operation, taking  $x$  and the pseudo-warping field  $f$  as input, generates the slice  $(W(x, f))$ .  $\mathcal{L}_{RC}$  and  $\mathcal{L}_{SC}$  are recycle consistency loss and spatial consistency loss to maintain the quality and consistency of the translated images. (c) The warping field synthesis and warping operation modules. Warping field generation utilizing the length and width of the images.

## 2.2. Network architecture

In this study, we propose a model that addresses the problem of inter-slice variance in 3D MRI harmonization using a pseudo-warping field translation with GAN. Fig. 1(a) provides an overview of the architecture of our model, which consists of three modules:

1. Two generators  $G_x$  and  $G_y$  for two domains to synthesize MR images  $x'$ ,  $y'$ .  $G_x$  translate images  $y$  from domain Y to the domain X, and vice versa with  $G_y$  ( $G_x : Y \rightarrow X$ ,  $G_y : X \rightarrow Y$ ).
2. Two discriminators ( $D_x$  and  $D_y$ ) for two domains to distinguish between real and fake images.
3. A warping operation  $W$  to generate a pseudo slice using the original slice and pseudo-warping field.  $W(x, f)$  denotes the warping operation performed on input  $x$  with pseudo-warping field  $f$ , resulting in the generation of the new slice.

First, we split the MR images into two domains depending on the datasets they come from, one for the source images  $x$ , and the other for the target images  $y$ . Our goal is to facilitate the transformation of images across domains while ensuring stability within each domain's image slices. To achieve this, we have introduced a warping operation that imports the slices into a new space, thereby enhancing the generator's capacity to translate subtle variations in slices. Our model is entirely unsupervised, negating the necessity for paired 3D MR images.

## 2.3. Pseudo-warping field

Initially, the pseudo-warping fields are randomly generated and applied to both source domain images  $x$  and their corresponding target domain harmonized image  $G_y(x)$ . This process results in two warped slices:  $W(x, f)$  and  $W(G_y(x), f)$ , as shown in Fig. 1(b), representing two arbitrarily distorted subsequent slices from the same image across both domains. Subsequently,  $W(G_y(x), f)$  is retranslated to the source domain using the generator  $G_x$ , producing  $G_x(W(G_y(x), f))$ ; similarly,  $W(x, f)$  is translated to the target domain using  $G_y$ , resulting

in  $G_y(W(x, f))$ . The entire model is trained using two specific loss functions: a Recycle Consistency Loss ( $\mathcal{L}_{RC}$ ), ensuring accurate reconstruction of  $G_x(W(G_y(x), f))$  as the fake subsequent slice of the original image  $x$  (i.e.,  $W(x, f)$ ), and a Spatial Consistency Loss ( $\mathcal{L}_{SC}$ ), ensuring that the generator produces identical subsequent images regardless of the warping's sequence relative to the generator.

For each individual slice, a random Warping Field of identical height and width was created and applied (Wang et al., 2020), as illustrated in Fig. 1(c). This random generation process ensures that the Warping Field is uniquely tailored to each slice, aligning with the specific dimensions of the input image. For both the X and Y domain, the structure of the scale field  $\tilde{f}_{sc}$  is represented by a randomly generated grid, where each element of the grid corresponds to a  $100 \times 100$  block (acts as a blur kernel for conducting mean filtering) in the input image. This grid is sampled from a normal distribution  $\mathcal{N}(0, \sigma_{sc}^2)$ . The shift field  $\tilde{f}_{sh}$  refers to a constant variance that is sampled in both domains from  $\mathcal{N}(0, \sigma_{sh}^2)$ . The final synthetic warping field  $\tilde{f} = \tilde{f}_{sc} + \tilde{f}_{sh}$ . Then, simulated slices  $W(x, f)$  are generated by warping the input image  $x$  using the warping field and extra Gaussian noise. The variance of the noise  $\Delta$  is modeled by  $\Delta \sim \mathcal{N}(0, \sigma_N^2 I)$ ,  $\sigma_N \sim \mathcal{U}(0.01, 0.02)$ . Similarly, the images of the target domain  $G_y(x)$ ,  $G_x(y)$  (translated from the source image) are used to generate the slice  $W(G_y(x), f)$ ,  $W(G_x(y), f)$  by the same warping field.

## 2.4. Adversarial loss

To translate 3D MR images between two domains, we train two generators,  $G_x$ ,  $G_y$ , to produce synthetic images in their respective domains. Then, we train two discriminators,  $D_x$ ,  $D_y$ , to discriminate between the generated images and those from their respective domains. The training process continues iteratively until the generators can synthesize images that are realistic enough to fool the discriminator and can be considered as real images in their respective domains.

The generator G requires an input image  $x$  or  $y$  to produce output images  $y' = G_y(x)$  or  $x' = G_x(y)$ . We assume a source image stack X with

$N$  slices and a target image stack  $Y$  with  $M$  slices. The discriminator  $D$  identifies whether the images are real or fake using the adversarial loss ( $\mathcal{L}_{adv}$ ):

$$\mathcal{L}_{adv} = \sum_M [\log D_y(y)] + \sum_M [\log(1 - D_y(y))] + \sum_N [\log D_x(x)] + \sum_N [\log(1 - D_x(x))] \quad (1)$$

### 2.5. Consistency losses

In Section 2.3, we highlighted that Warping Field can enhance consistency between domains. To achieve even greater stability, we introduce two additional losses: the recycle consistency loss and the spatial consistency loss. These losses aim at ensuring spatiotemporal cyclic consistency in the source domain and the target domain, respectively.

**Recycle Consistency Loss.** To ensure that the warping and generation operations do not alter the important anatomical content in the MR images, a special procedure is employed. First, a source slice is taken, denoted by  $x$  and  $y$ , and translated to the target slice, with  $y' = G_y(x)$  or  $x' = G_x(y)$ . Then, the warping operation is applied to the slices resulting in the next slices,  $W(x, f)$ ,  $W(y, f)$ ,  $W(G_x(y), f)$  and  $W(G_y(x), f)$ . Then,  $W(G_x(y), f)$  and  $W(G_y(x), f)$  are returned to the original domain  $G_y(W(G_x(y), f))$ ,  $G_x(W(G_y(x), f))$ , and make a consistency with  $W(x, f)$ ,  $W(y, f)$ , i.e.,  $x \rightarrow G_y(x) \rightarrow W(G_y(x), f) \rightarrow G_x(W(G_y(x), f)) \approx W(x, f) \leftarrow x$ . The consistency between them is enforced by an unsupervised recycle loss, represented as  $\mathcal{L}_{RC}$ :

$$\mathcal{L}_{RC} = \sum_M \|W(y, f) - G_y(W(G_x(y), f))\|_1 + \sum_N \|W(x, f) - G_x(W(G_y(x), f))\|_1 \quad (2)$$

**Spatial Consistency Loss.** In the target side, a spatial consistency loss, denoted as  $\mathcal{L}_{SC}$ , is presented. This loss is devised to measure the consistency between  $W(G_x(y), f)$ ,  $W(G_y(x), f)$  and  $G_x(W(y, f))$ ,  $G_y(W(x, f))$ , i.e.,  $x \rightarrow G_y(x) \rightarrow W(G_y(x), f) \approx G_y(W(x, f)) \leftarrow W(x, f) \leftarrow x$ , represented as  $\mathcal{L}_{SC}$ :

$$\mathcal{L}_{SC} = \sum_M \|W(G_y(x), f) - G_y(W(x, f))\|_1 + \sum_N \|W(G_x(y), f) - G_x(W(y, f))\|_1 \quad (3)$$

All losses can be summarized by a total loss function  $\mathcal{L}$ :

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{RC}\mathcal{L}_{RC} + \lambda_{SC}\mathcal{L}_{SC} \quad (4)$$

The loss weights  $\lambda_{RC}$  and  $\lambda_{SC}$  are used to control the loss function. Initially, we set both weights to 10. The training process is terminated once the generator can produce realistic images that cannot be distinguished by the discriminator.

### 2.6. Evaluation metrics

To make a comprehensive evaluation of our results, we use several quantitative metrics to compare the visual quality of the harmonized images. We use Peak Signal-to-Noise Ratio (PSNR), Batch-wise squared Maximum Mean Discrepancy (bMMD<sup>2</sup>) (Kwon et al., 2019) and Multi-Scale Structural Similarity (MS-SSIM) (Kwon et al., 2019) to measure the similarity of source and the harmonized images based on their intensity distribution. Since bMMD<sup>2</sup> was originally designed for 2D images, we slice the 3D images into a stack of 2D images in three directions and computed the metric over the entire dataset. To measure the variances between slices within one MR image, we selected the Warping Error (WE) (Lai et al., 2018), which quantifies the inter-slice variance/stability of 3D MR images. Intuitively, a lower warping error indicates a smaller difference between adjacent images, suggesting stronger temporal stability. We computed the averages across the entire

test set to derive the final metrics.  $f_{i \Rightarrow i-1}$  represents the ground truth computed from adjacent images, while  $F$  denotes the warp function:

$$WE = \sum_{i=2}^T \sum_{f=1}^N \exp(-\alpha \|x_i - F(x_{i-1}, f_{i \Rightarrow i-1})\|_2) \|x_i - F(x_{i-1}, f_{i \Rightarrow i-1})\|_1 \quad (5)$$

## 3. Experiments

In this section, we will evaluate the performance of our method in comparison to existing methods, many of which rely on 2D MR images to reconstruct 3D MR images. We will provide further information on the specifics of our implementation, including datasets and evaluation metrics, in the upcoming sections.

### 3.1. Datasets

For qualitative and quantitative comparison, we first obtained T1-weighted brain MR images from three publicly available datasets: ADNI (Mueller et al., 2005), UKBB (Sudlow et al., 2015) and NKI-RS (Tobe et al., 2021). The NKI-RS study received ethical clearance from the Institutional Review Boards at both the Nathan Kline Institute (#226781 and #239708) and Montclair State University (#000983 A and #000983B), ensuring compliance with ethical guidelines. Written informed consent was obtained from adult participants, while minor/child participants required both written consent and assent from their legal guardians. The data used in this study came from a disease-free subset of participants, and additional information about image acquisition can be found elsewhere.

In order to compare the results quantitatively, we applied it to a subset of the ADNI-traveling subjects dataset, consisting of 44 subjects who had scans using both 1.5T and 3T scanners within 30 days, resulting in two datasets with identical anatomical structures. For robustness evaluation, we conducted three-fold cross-validation on the ADNI (traveling subjects) dataset, dividing it into three groups, each with 13, 15, and 16 MR images for both 3T and 1.5T. We cross-mapped the 3T dataset to the 1.5T dataset and vice versa. Furthermore, we demonstrated our model's capacity to adapt styles and maintain anatomical details between different datasets by utilizing unpaired training data (ADNI-CN&UKBB, UKBB&NKI-RS). The data allocation involved dedicating 65% for training/validation and the remaining 35% for testing, ensuring comprehensive evaluation. The dataset images were sliced in the sagittal direction and used for training our model as well as the comparative models. In order to validate that our model could yield satisfactory results across different planes, we augmented the training data with slices from three directions on the ADNI dataset (3T, 1.5T).

Besides, we also selected 135 Alzheimer's disease patients (ADNI-AD) and compared them to 135 healthy control subjects (ADNI-CN) from the ADNI dataset to investigate case/control effect size differences.

Note that the number of scans varied across datasets, except for the paired dataset ADNI (3T, 1.5T). For more information on the datasets, including the number of images, please refer to Table 1.

To ensure consistency in image processing, all datasets were cranially stripped using HD-BET (Isensee et al., 2019), corrected for non-uniformity with the N3 method, and linearly registered to the 1 mm<sup>3</sup> MNI152 template using FSL flirt (9 degrees of freedom). The images were then resampled to 0.8 mm<sup>3</sup> isotropic 256 × 256 × 256 voxels and converted into 256 2D images of size 256 × 256.

### 3.2. Competing methods

We performed a comprehensive comparative assessment of our model against various modern 2D and 3D image harmonization techniques, including state-of-the-art GAN-based, CNN-based and transformer-based methods. Some of them are specially designed for

**Table 1**

Number of images, Age (Mean  $\pm$  SD), Female/Male, and Field strength for the all datasets: ADNI-traveling subjects (1.5T and 3T), ADNI-CN (healthy subjects in the ADNI dataset), ADNI-AD (AD patients in the ADNI dataset), NKI-RS, UK Biobank. It should be noted that the selection of images used in this study may not be entirely representative of the entire dataset.

Datasets	Number of images	Age (Mean $\pm$ SD)	Female/Male	Field strength	
				1.5T	3T
ADNI-traveling subjects	44	68.58 $\pm$ 0.82	26/18	✓	✓
ADNI-CN	135	61.92 $\pm$ 2.92	81/54		✓
ADNI-AD	135	59.32 $\pm$ 2.91	48/87		✓
NKI-RS	181	59.77 $\pm$ 3.18	130/51		✓
UK Biobank	185	60.05 $\pm$ 3.06	123/62		✓

**Table 2**

All models' computational cost. The trainable parameters are computed for both the generator and the discriminator based on the number of filters in the first or last convolutional layer (i.e., ngf and ndf) set to 256.

Method	Input size	Trainable params (M)	Memory usage (MB)
BlindHarmony	256 $\times$ 256	29.13M	5459
ImUnity	256 $\times$ 256	9.73M	1709
RecycleGAN	256 $\times$ 256	129.98M	6061
ResViT	256 $\times$ 256	126.21M	2501
TCGAN	256 $\times$ 256	193.98M	10203
CycleGAN2D	256 $\times$ 256	114.33M	17686
CycleGAN3D	128 $\times$ 128 $\times$ 128	114.33M	36269
pix2pix3D	128 $\times$ 128 $\times$ 128	41.98M	45007
PTNet3D	256 $\times$ 256 $\times$ 256	49.32M	15203
Ours	256 $\times$ 256	21.18M	4673

MRI harmonization (ImUnity & BlinkHarmony), and some of them have been proved to be successful in general image translation tasks:

**BlindHarmony** (Jeong et al., 2023): A flow-based blind harmonization method designed for MR images. The network is trained solely on the target domain dataset and can be employed for previously unseen images from different domains.

**ImUnity** (Cackowski et al., 2021): Inspired by convolutional VAE-GANs, leverages adversarial strategies via a classical CNN discriminator to ensure realistic outputs.

**RecycleGAN** (Bansal et al., 2018): An unsupervised data-driven approach for video retargeting, capable of transferring sequential content between domains while preserving the style attributes of the target domain.

**ResViT** (Dalmaz et al., 2022): An adversarial model with a hybrid CNN-transformer architecture, aiming to attain both high localization and contextual sensitivity while preserving a high degree of realism in the synthesized images.

**TCGAN** (Li et al., 2022): A multimodal medical image synthesis technique that incorporates the transformer structure to address the limitations of CNNs and capture richer contextual information.

**CycleGAN** (Zhu et al., 2017): A framework comprising cycle-consistent generative adversarial networks designed for unpaired image-to-image translation tasks. The code for the 3D model is available at <https://github.com/davidiommi/3D-CycleGan-Pytorch-MedImaging>.

**pix2pix3D** (Isola et al., 2017): A convolutional GAN model with U-Net backbone was considered. Pix2pix has a CNN-based generator with an encoder-decoder structure tied with skip connections. We used the code extracted from Hu et al. (2022).

**PTNet3D** (Kwon et al., 2019): A 3D MRI synthesis framework that leverages attention mechanisms through transformer and performer layers. As a non-adversarial model, it dispenses with a discriminator and embraces a convolution-free architecture.

To ensure equitable comparison, both supervised and unsupervised models were evaluated simultaneously. The paired dataset (ADNI-traveling subjects) was utilized for training and testing tasks across all models. Conversely, the training and testing tasks on non-paired datasets (ADNI-CN&UKBB, NKI-RS&UKBB) were conducted exclusively

with models (BlindHarmony, ImUnity, RecycleGAN, CycleGAN) that do not require paired images.

It is noteworthy that, to account for computational limitations, it is worth noting that the previously mentioned 3D models were trained on images with dimensions of 128  $\times$  128  $\times$  128, with the exception of PTNet3D. Consequently, these models produced harmonized images with reduced resolution. Furthermore, since the ResViT, TCGAN, pix2pix3D, and PTNet3D models require paired data for training, our comparative analysis is conducted specifically on the ADNI-traveling dataset.

For all 2D methods, the 3D MR image into a stack of 2D images (sagittal plane for all datasets) and input each stack into the model to perform stack-to-stack harmonization. Then, we reconstructed the stacks into 3D volumes by assigning the obtained slices to the corresponding layers of the new 3D image data, utilizing the original Nifti image's affine, header, and other information. This operation effectively embeds the information from slices into the new 3D data, achieving a comprehensive three-dimensional reconstruction.

### 3.3. Task-specific evaluation analyses

**Brain age:** Brain age assessment evaluated cross-site prediction generalization across unharmonized and GAN harmonized scans. We used a deep-learning brain age prediction model (Gupta et al., 2021), which predicts brain age from 3D scans. For detailed network architecture, refer to Gupta et al. (2021). In harmonized experiments, the prediction model exclusively trained on UK Biobank MRI scans (train: 768 samples, validate: 128 samples), ages 47–77. Evaluation used NKI-RS and ADNI-CN datasets, treating each scan independently.

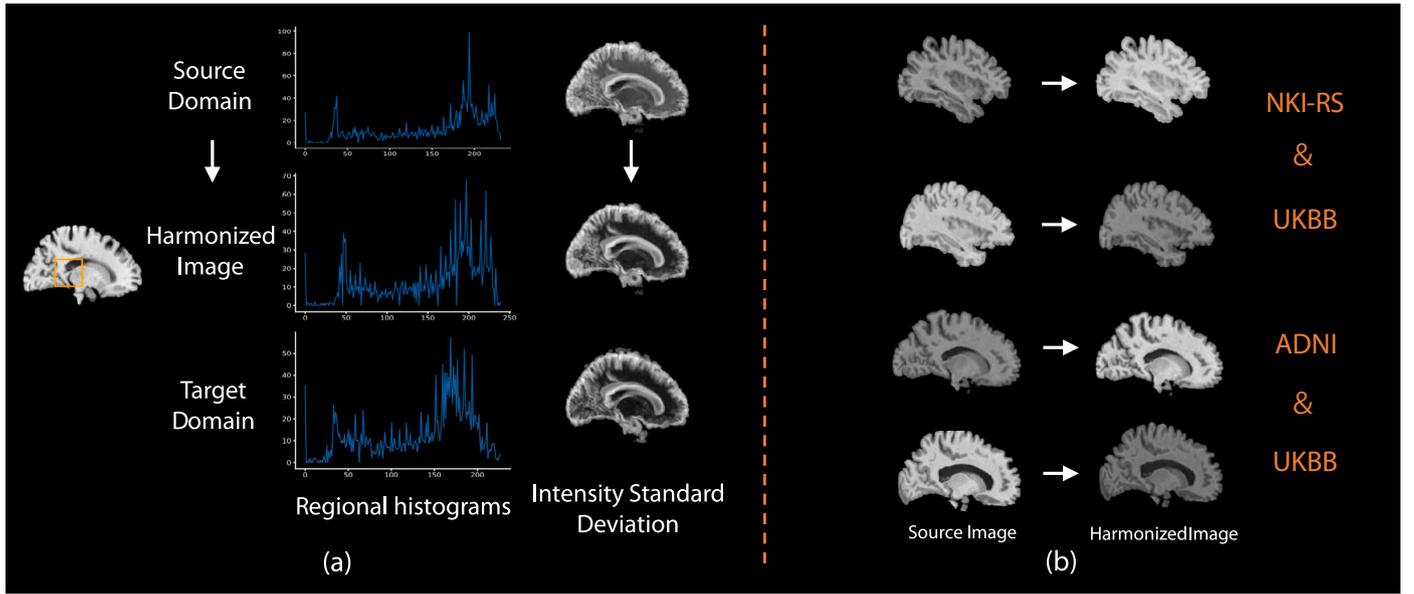
**Assessments of cortical and subcortical metrics derived from FreeSurfer:** Cortical surface reconstruction and subcortical volume segmentation were performed using FreeSurfer 7.1.0, available at <http://surfer.nmr.mgh.harvard.edu/> (Fischl, 2012; Fischl and Dale, 2000). Notably, bias field correction and skull-stripping were conducted in the image preprocessing stage before harmonization, and therefore, omitted in the recon-all script.

To evaluate harmonization's impact, we compared features, including lateral ventricle and hippocampal volumes. These comparisons helped assess changes in feature values pre and post harmonization and

**Table 3**

Datasets used for each of the validation experiments. Because the UKBB, NKI-RS, and ADNI-CN datasets are non-paired, our image similarity calculations are solely performed on the ADNI-traveling subjects.

		Validation experiment							
		Regional histograms and standard deviation	t-SNE	Warping error	Image similarity	Case/control effect size differences	Surface reconstruction	Brain age prediction	Different scanning planes
Datasets	ADNI-traveling subjects	N=44		N=44	N=44		N=44		N=44
	ADNI-CN		N=135	N=135		N=135		N=135	
	ADNI-AD					N=135			
	UKBB		N=183	N=183				N=183	
	NKI-RS		N=181	N=181				N=181	



**Fig. 2.** (a) Left: A comparison of regional histograms before and after mapping to the target domain. Right: Standard deviation (SD) maps across all scans. The images depicted in the figure represent the experimental results of the ADNI-traveling subjects. (b) Unpaired image translation. The harmonization model is trained on NKI-RS&UKBB and ADNI-CN&UKBB to showcase its effectiveness in achieving high performance in an unpaired setting.

determine cross-site similarity in extracted feature values for subjects scanned on multiple scanners.

### 3.4. Implementation details

Experiments were performed on an NVIDIA A6000 GPU with 49 GB memory utilizing batch sizes of 9, noise level of 0.001, regularization parameters  $\lambda_{RC}$  and  $\lambda_{SC}$  are set to 10, The initial learning rate was set to  $2 \times 10^{-4}$ . We used the Adam optimizer with  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$ . The learning rate was constant for the first 20 epochs for ADNI-traveling task (50 epochs for ADNI-CN&UKBB and NKI-RS&UKBB tasks) and linearly decayed to 0 in the remaining epochs. Image resolution was standardized to  $256 \times 256$  and normalized between 0 and 255. The code was implemented in Python 3.8 and PyTorch 1.12.0.

## 4. Results

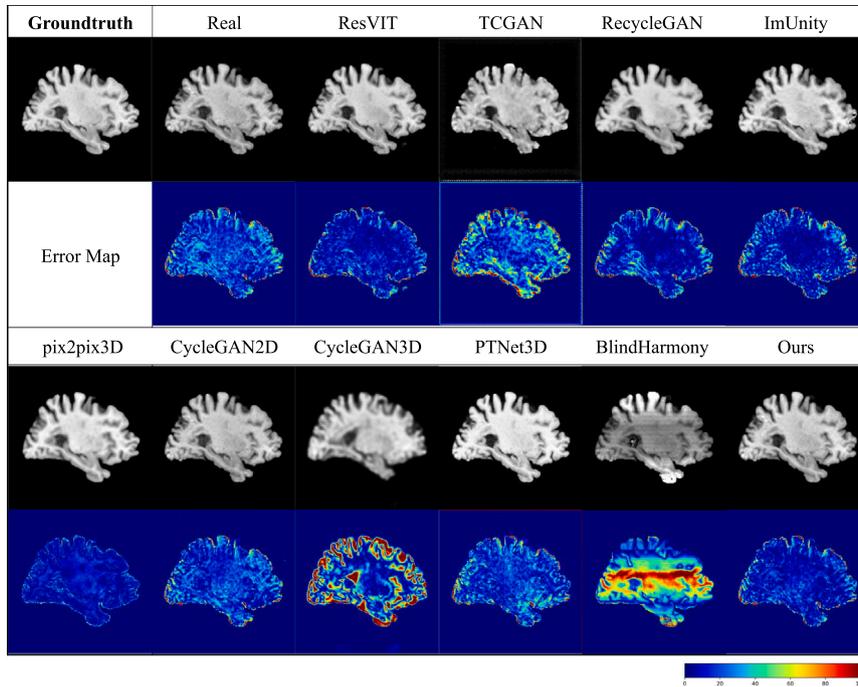
Our primary goal was to compare our method with deep learning models employing 2D and 3D operators. In Table 2, we have presented comparative insights into computational costs. Furthermore, a comprehensive evaluation and comparative analysis were performed, employing various methodologies, including inter-slice variance, image similarity, and visual comparison, as elaborated in Table 3.

### 4.1. Image-wide evaluations

**Visual Comparison.** In Fig. 2(a), the regional intensities of the harmonized image exhibit a closer alignment with the target domain

image, in contrast to the source domain image. Furthermore, it is evident that the intensity standard deviation (SD) maps of the harmonized image bear a stronger resemblance to the target domain in comparison to those of the source domain image. Fig. 2(b) showcases the unpaired results of our model, demonstrating its capacity to effectively translate site-variant styles while preserving the site-invariant content. In Fig. 3, we compute absolute error maps by comparing the target scan with the synthesized (or source) scans generated by each respective model. It is evident that our model displays fewer absolute errors.

**Inter-slice Stability Comparison.** Our study involved a comparative analysis of our model against both 3D-based models and 2D slice-based models using distinct datasets. The warping error metrics are meticulously documented in Tables 4 and 5. Among the considered methods, the outcomes derived from our method outperform those of both 2D techniques and the majority of 3D methodologies. In the case of the (ADNI-CN, UKBB) and (NKI-RS, UKBB) datasets, our model exhibits superior performance with regards to warping error. Furthermore, when assessing performance on the ADNI-traveling subjects task, our model demonstrates a performance closely aligned with pix2pix3D, which closely approximates the performance observed with real scans. However, it is essential to emphasize that the computational resources required by pix2pix3D are 9.6 times greater than those consumed by our model. On average, our results in Warping Error (WE) with a mean value of  $64.2 \pm 4.15(10e-5)$  for ADNI-traveling subjects (1.5T and 3T). The outcomes were in close proximity to those of the original images, which recorded a result of  $63.3 \pm 5.83(10e-5)$ . Similarly, the metric yielded values of  $61.7 \pm 4.86(10e-5)$  for ADNI-CN and UKBB,



**Fig. 3.** Comparisons between our model and existing harmonization models can be conducted through visualizations and absolute error maps. The absolute error is computed between the normalized ground truth and the converted scans, spanning the range of [0, 100], with lower values (indicated by the blue color) denoting minor differences. Upon careful examination, we have observed that several other models yield error maps of larger extent than the majority depicted in the figure. Pix2pix3D demonstrates the most promising outcomes. However, these achievements come at the expense of compromised image clarity and a relatively lower resolution.

**Table 4**

Inter-slice correlation comparison for ADNI-traveling subjects (3T, 1.5T). The symbol ↓ indicates that lower values are better, and the best result is highlighted in bold, while the second-best result is underlined. The term “No har” category includes the source and target images, indicates the absence of harmonization.

Dataset	Method	Warping error (WE↓) (10e-5)
ADNI (3T and 1.5T)	BlindHarmony	94.7 ± 10.6
	ImUnity	70.7 ± 3.90
	RecycleGAN	67.5 ± 4.53
	ResViT	68.6 ± 4.93
	TCGAN	103.5 ± 17.4
	CycleGAN2D	66.7 ± 5.97
	CycleGAN3D	131.7 ± 16.9
	<u>pix2pix3D</u>	<b>63.9 ± 4.26</b>
	PTNet3D	185.9 ± 4.01
	Ours	<u>64.2 ± 4.15</u>
	No har (source images)	63.3 ± 5.83

aligning closely with the original image result of  $61.5 \pm 4.39(10e-5)$ . Correspondingly, for UKBB and NKI-RS, the metric exhibited a value of  $63.4 \pm 5.35(10e-5)$ , while the outcome maintained a high degree of consistency with the original image result of  $64.9 \pm 5.38(10e-5)$ .

In summary, our model exhibits a warping error value that is similar to the original source images. It shows the least warping errors when compared to other 2D methods and is generally lower than most 3D methods, except for one task where it is slightly higher than the top-performing 3D models.

**Image Similarity Comparison.** To further compare the similarity between source and target images, we quantitatively assessed the images from both domains using three metrics, PSNR, bMMD<sup>2</sup>, and MS-SSIM (bMMD<sup>2</sup> are based on 2D images, PSNR and MS-SSIM is based on 3D images) as illustrated in Table 6. In our observations, our model consistently demonstrates superior performance across a spectrum of evaluation metrics, encompassing PSNR, bMMD<sup>2</sup>, and MS-SSIM, in both

the 3T→1.5T and 1.5T→3T tasks. The sole exception to this trend is noticeable in the 3T→1.5T direction, wherein the ResViT model exhibits a slightly elevated bMMD<sup>2</sup> score. The observed percentage of Batch-wise squared Maximum Mean Discrepancy (bMMD<sup>2</sup>) decreases were 32.6% (Site 1.5T) and 21.7% (Site 3T) respectively. While the values of Multi-Scale Structural Similarity (MS-SSIM), showcased increases from 0.972 to 0.985 (Site 1.5T) and 0.979 (Site 3T), with corresponding percentage increments of 2.34% and 0.69%. And the percentage increments for Peak Signal-to-Noise Ratio (PSNR) were 8.50% (Site 1.5T) and 7.23% (Site 3T), respectively. It is worth highlighting that the ADNI-traveling dataset manifests significant inter-site anatomical variability. Nonetheless, in comparison to the baseline models, our methodology excels in the generation of target images characterized by diminished artifact levels, heightened anatomical verisimilitude, and sharper tissue representation.

**T-distributed Stochastic Neighbor Embedding (t-SNE).** To assess the similarity of harmonized data in high-dimensional space, we utilized a t-distributed stochastic neighbor embedding (t-SNE) plot (van der Maaten and Hinton, 2008) to visualize image representations obtained from randomly selected images from the UKBB, NKI datasets. Briefly, t-SNE is a non-linear dimensionality reduction method that facilitates the visualization of high-dimensional data, where similar data points are closer together, and dissimilar points are positioned farther apart. The Fig. 4 shown in the plot was extracted from the synthesized images generated before and after the harmonization process in the model. Notably, prior to harmonization, the features exhibit a discernible separability based on the datasets of origin. However, upon the completion of the harmonization process, lower-level stylistic attributes, such as contrast, undergo a transformative shift, culminating in a cohesively embedded representation. It is noteworthy that subsequent to harmonization, the feature embedding no longer discernibly distinguishes between datasets, as visually illustrated in Fig. 4. This observation underscores the efficacy of the proposed methodology.

**Table 5**

Inter-slice correlation comparison for unpaired datasets task ADNI-CN&UKBB and NKI-RS&UKBB. As the datasets in these experiments consist of unpaired images, any model that necessitates paired images as input was excluded from this comparison.

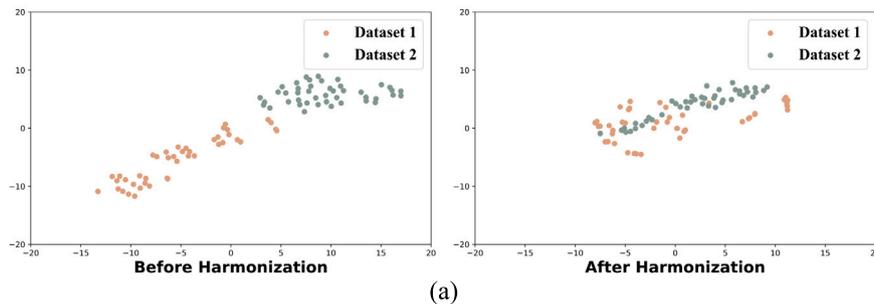
Dataset	Method	Warping error (↓) (10e-5)	Dataset	Method	Warping error (↓) (10e-5)
NKI-RS & UKBB	BlindHarmony	259.9 ± 117.2	ADNI-CN & UKBB	BlindHarmony	228.3 ± 75.6
	ImUnity	63.5 ± 3.73		ImUnity	64.3 ± 3.81
	RecycleGAN	68.5 ± 7.27		RecycleGAN	81.0 ± 15.8
	CycleGAN2D	67.0 ± 7.31		CycleGAN2D	73.6 ± 9.10
	CycleGAN3D	78.5 ± 37.2		CycleGAN3D	62.9 ± 2.44
	Ours	<b>63.4 ± 5.35</b>		Ours	<b>61.7 ± 4.86</b>
No har	64.9 ± 5.38	No har	61.5 ± 4.39		

**Table 6**

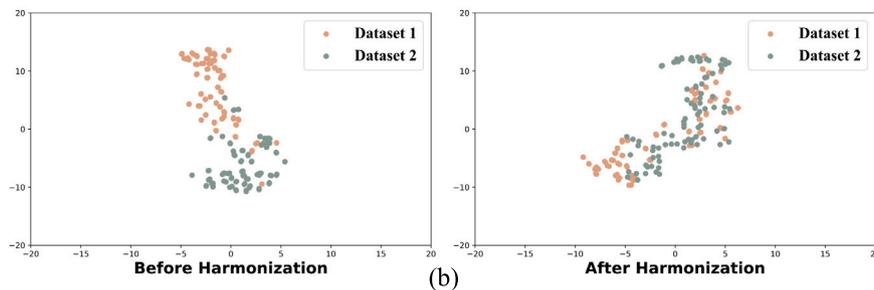
Image similarity results for ADNI-Traveling subjects. The values were computed using the target domain image as a reference. The term “No har” category includes the source and target images, indicates the absence of harmonization. Although the dataset is paired, there may be differences in anatomical details (as shown in Fig. 3), which imposes an upper limit on the similarity values. Bold indicated the best performance.

Models	PSNR (↑)	bMMD <sup>2</sup> (↓)	MS-SSIM (↑)	Models	PSNR (↑)	bMMD <sup>2</sup> (↓)	MS-SSIM (↑)
BlindHarmony	19.75 ± 1.21	374.5 ± 254.2	0.949 ± 0.016	BlindHarmony	21.72 ± 1.77	404.2 ± 132.4	0.949 ± 0.025
ImUnity	25.44 ± 1.82	205.0 ± 108.6	0.977 ± 0.014	ImUnity	23.76 ± 2.58	323.3 ± 195.8	0.971 ± 0.017
RecycleGAN	25.54 ± 1.97	205.9 ± 116.4	0.977 ± 0.015	RecycleGAN	24.74 ± 2.72	266.7 ± 173.5	0.972 ± 0.018
ResViT	22.68 ± 2.39	<b>175.2 ± 112.5</b>	0.963 ± 0.018	ResViT	23.19 ± 1.04	247.8 ± 161.4	0.963 ± 0.015
TCGAN	23.04 ± 1.21	337.3 ± 110.3	0.937 ± 0.018	TCGAN	22.56 ± 2.01	400.2 ± 188.5	0.959 ± 0.017
CycleGAN2D	25.13 ± 2.96	248.9 ± 173.0	0.975 ± 0.018	CycleGAN2D	25.08 ± 2.91	249.9 ± 169.0	0.973 ± 0.017
CycleGAN3D	18.03 ± 1.55	344.0 ± 105.6	0.866 ± 0.046	CycleGAN3D	17.16 ± 2.28	416.1 ± 118.9	0.809 ± 0.079
pix2pix3D	26.68 ± 1.97	188.9 ± 109.6	0.981 ± 0.013	pix2pix3D	24.32 ± 3.93	218.8 ± 105.1	0.967 ± 0.026
PTNet3D	18.42 ± 2.80	223.9 ± 105.9	0.880 ± 0.010	PTNet3D	21.47 ± 0.87	399.5 ± 239.1	0.874 ± 0.012
Ours	<b>26.79 ± 2.60</b>	186.2 ± 140.2	<b>0.985 ± 0.017</b>	Ours	<b>26.47 ± 2.61</b>	<b>216.3 ± 138.3</b>	<b>0.979 ± 0.017</b>
No har (source images)	24.69 ± 3.027	276.5 ± 183.6	0.972 ± 0.017	No har (source images)	24.69 ± 3.027	276.5 ± 183.6	0.972 ± 0.017

**t-SNE Visualization of UKBB and ADNI-CN Datasets**



**t-SNE Visualization of UKBB and NKI-RS Datasets**



**Fig. 4.** T-SNE representation of the images from two datasets (UKBB, NKI-RS)(UKBB, ADNI-CN) before and after harmonization.

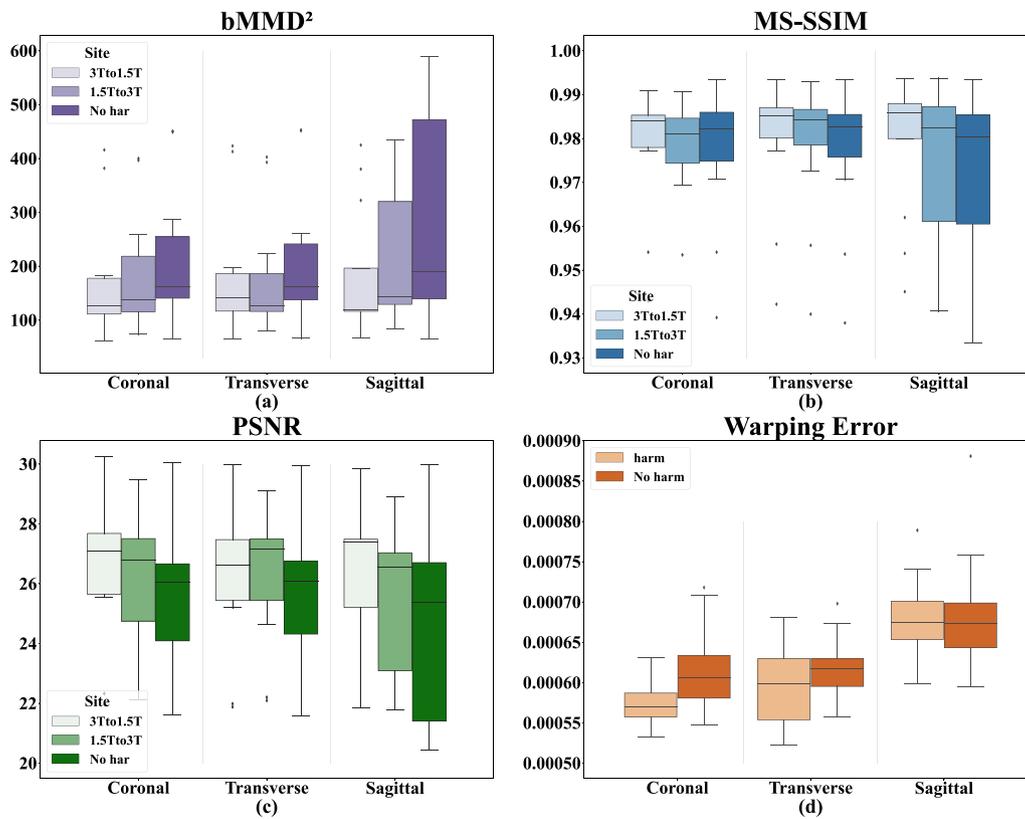


Fig. 5. Comparison of Different Scanning Planes. The results of our study demonstrate that our method is capable of achieving favorable outcomes in any scanning planes, as illustrates in the figure above. The term “No har” category includes the source and target images, indicates the absence of harmonization.

#### 4.2. Comparison of different scanning planes

To guarantee the robustness of our model across sagittal, transverse, and coronal positions, we conducted training on all three slices using the ADNI-traveling dataset, comprising 31 Nifti images for the training set and 13 images for the test set. The corresponding test results are illustrated in Fig. 5. As evident from the graphical representation, in the majority of instances, our approach exhibited an enhanced performance concerning both image similarity and warping error across a range of scenarios involving three scanning planes, when contrasted with the baseline methods.

#### 4.3. Task-specific evaluation of downstream analyses on 3D reconstructions

**Surface Reconstruction Comparison.** We next run Freesurfer v7.1 on all the images and compare the cortical volume differences between 1.5T images and 3T images before and after the harmonization. The volume differences (delta volume) are measured by dividing the volume differences values by cortical structural volumes (average of 1.5T and 3T). Results suggest that after harmonization, volume differences between 1.5T image and 3T images are smaller than before harmonization for all the brain structures we evaluate. We extended our comparison of volume differences to include three harmonization models: the unsupervised ImUnity model, and the supervised models of ResVit and pix2pix3D, all of which showed relatively well performance in our prior evaluations. We further harmonized the structural volume using combat and then compare the volume differences between combat and our method. Results indicate that our method outperforms other methods in every application (Fig. 6).

**Maintaining Case/Control Effect Size Differences:** In order to validate the efficacy of our approach in preserving pathological features following harmonization, we conducted an evaluation of hippocampal volumes in Alzheimer’s disease (AD) patients (ADNI-AD) compared

to healthy individuals (ADNI) within the ADNI dataset. This assessment was performed both prior to and after harmonization, utilizing a reference from the UK Biobank (UKBB). Additionally, we accounted for covariates including age, gender, and intracranial volume in our analysis. The measurement of hippocampal volume was conducted separately for both the left and right hemispheres. As illustrated in Fig. 6, before harmonization, the hippocampal volumes in AD patients (left:  $2046.7 \pm 962.9 \text{ mm}^3$ ; right:  $2763.9 \pm 979.3 \text{ mm}^3$ ) were noticeably lesser than those in healthy individuals (left:  $3111.2 \pm 1009.6 \text{ mm}^3$ ; right:  $3906.4 \pm 996.0 \text{ mm}^3$ ; left:  $p < 0.0001$ , Cohen’s  $d = -1.08$ ; right:  $p < 0.0001$ , Cohen’s  $d = -1.16$ ).

After harmonization, these differences were pronounced. The hippocampal volumes in AD patients (left:  $2123.3 \pm 959.9 \text{ mm}^3$ ; right:  $2837.7 \pm 979.7 \text{ mm}^3$ ) when contrasted with the controls (left:  $3187.7 \pm 1010.6 \text{ mm}^3$ ; right:  $3978.9 \pm 998.7 \text{ mm}^3$ ) continued to show significant disparities. The effect sizes were almost identical to those observed pre-harmonization (left:  $p < 0.0001$ , Cohen’s  $d = -1.08$ ; right:  $p < 0.0001$ , Cohen’s  $d = -1.16$ ), as shown in Fig. 7.

**Age Prediction.** After completing the model training, we applied it to individuals from three datasets: the UK Biobank ( $n=85$ , age range 55–65 years old) and NKI-RS ( $n=81$ , age range 55–65 years old) and ADNI-CN ( $n=35$ , age range 56–65 years old). In the UK Biobank healthy brain age test set, we achieved a mean absolute error (MAE) of 2.6054 years between the true chronological age and the predicted brain age, with a corresponding mean squared error (MSE) of 10.8310. Next, we proceeded to harmonize the NKI-RS and ADNI-CN images to align them with the UK Biobank dataset. Utilizing the brain age prediction model, we then compared predictions before and after the harmonization process. The specifics are outlined in the Table 7. It is clear that images, after undergoing our model’s harmonization process, show a noteworthy enhancement in predicted brain age compared to images without harmonization. Additionally, we benchmarked our outcomes against ImUnity, the top-performing unsupervised model

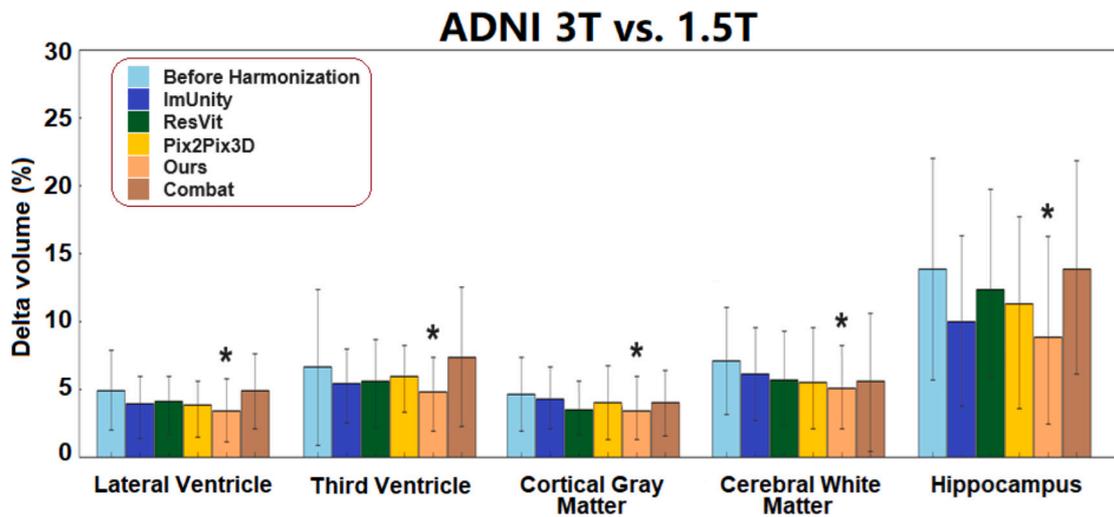


Fig. 6. Cortical structural volume differences for traveling subjects scanned by 1.5T scanner and 3T scanner within 30 days. The evaluation compared the volume differences prior to harmonization with those following various methods. An asterisk indicates the method that demonstrated the best performance across all conditions.

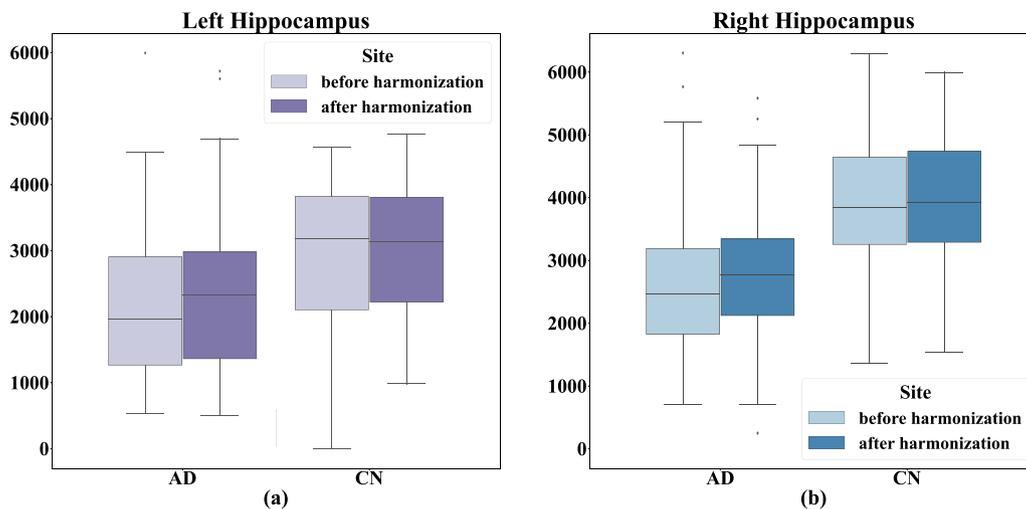


Fig. 7. Comparisons of hippocampal volume were conducted between participants diagnosed with Alzheimer’s Disease (AD) and cognitively healthy controls within the ADNI dataset, both before and after the harmonization process. The graphical representations indicate that harmonization does not significantly alter the statistical case/control differences observed within the respective cohorts.

Table 7  
Age prediction results in ADNI-CN2UKBB and NKI-RS2UKBB tasks.

	ADNI-CN		NKI-RS	
	MAE	MSE	MAE	MSE
Real scans	6.3538	48.3576	6.5608	50.9944
ImUnity	2.8924	13.0498	3.2416	18.1793
Our	<b>2.7284</b>	<b>11.2199</b>	<b>3.2326</b>	<b>17.3404</b>

that does not require paired images from previous evaluations. The comparison indicates that our model demonstrates smaller gaps in age prediction.

#### 4.4. Ablation studies

To assess the impact of the pseudo-warping field and loss function components, we evaluate our model on ADNI-traveling, UKBB to ADNI, and UKBB to NKI tasks, comparing harmonized results across various configurations. Our model, incorporating three modules, significantly improves overall performance compared to the baseline.

From Table 8, we observe a significant decline in performance across various metrics in all three tasks when the warping field and both loss components are omitted. The introduction of the three modules has noticeably enhanced the model’s ability to preserve the original image’s structure, texture, intensity, and contrast. Furthermore, it ensures a consistent alignment between the generated images and the original images at both pixel-level and global distribution. Additionally, the improvement in warping error indicates a substantial enhancement in inter-slice continuity upon incorporating this module.

#### 5. Discussion

Conventional 2D-based harmonization approaches, which directly convert 3D images into a stack of 2D MRI slices and harmonize them independently, fail to preserve the essential inter-slice continuity within the MRI slices. On the other hand, 3D-based methods often require substantial computational resources. In this research, we present a novel GAN-based harmonization framework that leverages a pseudo-warping field to address these limitations. Under the guidance of the pseudo-warping field, our method achieves high-quality harmonization while maintaining inter-slice consistency effectively. The efficiency of

**Table 8**  
Evaluating our model’s performance and ablating three modules.

		ADNI-traveling				UKBB&NKI-RS	UKBB&ADNI
		PSNR	MS-SSIM	bMMD <sup>2</sup>	WE	WE	WE
w/o $\mathcal{L}_{RC} + \mathcal{L}_{SC}$	1.5T→3T	16.43 ± 1.079	0.837 ± 0.014	623.8 ± 146.4	183.0 ± 12.6	164.7 ± 33.3	123.0 ± 33.4
	3T→1.5T	18.42 ± 0.456	0.855 ± 0.012	602.8 ± 143.6			
w/o $\mathcal{L}_{SC}$	1.5T→3T	20.49 ± 0.949	0.918 ± 0.065	596.4 ± 138.5	73.7 ± 5.67	177.2 ± 50.9	110.0 ± 38.2
	3T→1.5T	20.71 ± 1.13	0.936 ± 0.015	572.2 ± 153.8			
w/o $\mathcal{L}_{RC}$	1.5T→3T	18.85 ± 1.003	0.857 ± 0.016	607.1 ± 148.2	150.6 ± 17.1	159.9 ± 55.3	101.1 ± 44.5
	3T→1.5T	18.94 ± 0.56	0.870 ± 0.012	586.8 ± 153.1			
w/o warping field	1.5T→3T	22.48 ± 1.15	0.944 ± 0.016	382.6 ± 122.1	101.8 ± 9.01	104 ± 10.0	70.8 ± 16.1
	3T→1.5T	22.40 ± 1.22	0.940 ± 0.017	390.4 ± 122.3			
Baseline	1.5T→3T	26.79 ± 2.60	0.985 ± 0.017	186.2 ± 140.2	64.2 ± 4.15	63.4 ± 5.35	61.7 ± 4.86
	3T→1.5T	26.47 ± 2.61	0.979 ± 0.017	216.3 ± 138.3			

our model is primarily due to its focus on processing 2D slices, avoiding the extensive resource demands of 3D operations. Additionally, it sidesteps the traditional necessity of precisely estimating deformation between consecutive 2D slices to preserve inter-slice continuity. Utilizing a pseudo-warping field, our approach distinctively facilitates the harmonization of MR images across varied domains. In our extensive tests with both paired and unpaired datasets, our model demonstrated exceptional performance, outperformed various other learning-based and transformer-based techniques.

The concept of a pseudo-warping field stems from motion estimation techniques used in video-to-video translation tasks. In these tasks, motion is estimated to link adjacent frames, guiding generators to produce temporally consistent frames. However, this approach is not directly suitable in 3D MR imaging scenarios. In video, motion is defined as the movement of the same object across frames, but in brain MRI, each slice essentially contains different, yet similar, objects (brain tissues). Therefore, deformation in adjacent MR slices is not just about object movement; it also involves changes in the objects themselves. To adapt video translation techniques to MRI, it is necessary to limit motion estimation or warping operation to within single slices. Meanwhile, the ultimate goal of the motion estimation is not to faithfully predict the next frame, but to ensure the generator to produce smooth and consistent frames throughout the video. In other words, if the frames can be translated consistently, the accuracy of the estimated motion between them becomes less important.

Given these factors, we propose to employ a pseudo-warping field. For each individual slice, the pseudo-field warp it into an artificial subsequent slice without interfering the actual next slice. Therefore, it is ideal for our scenarios focused on individual slices. On the other hand, due to the artificial and random nature of the warping operation, the pseudo-warping field offers a method where slices that have been deformed using any arbitrary warping operation can be converted by the generator into images of a different domain. Importantly, this conversion is executed smoothly and consistently, aligning well with the characteristics of the current, actual slice. Theoretically, this warping field could include deformations from one slice to the next, as well as deformations within the same slice. Hence, the fundamental principle of our model is to train the generators to transform these slices into another domain with strong consistency, regardless of whether and how the slice changes. The slice-wise inconsistency can arise because the generator may react differently to neighboring slices due to their varying content. By incorporating the proposed warping operation, our model guides the generator to produce images unaffected by this operation, ensuring that all slices in an MRI image are converted with consistent quality.

Another merit of employing the pseudo-warping field is that it eliminates the need to learn the actual warping between slices. Instead, all warping fields are randomly synthesized and directly applied to images in both the source and target domains. Moreover, the parameters of these fields are not updated during the learning process.

This approach can significantly reduce the size of the model and save computational resources, offering a more efficient solution for handling image harmonizations.

To establish a dependable and efficient framework for spatial accuracy and slice-wise consistency regularization, we have developed two unsupervised loss functions: the recycle consistency loss and the spatial consistency loss based on a pair of generated fake subsequent slices in both the source and target domains. The term of “recycle” arises from the process involving paired fake subsequent slices in the source domain. One of these slices is directly warped from the current slice, while the other undergoes a multi-step transformation: it is first converted to the target domain using one generator, then warped, and finally reconverted back to the source domain using a different generator. Concurrently, a spatial loss is utilized to maintain spatial and slice-wise consistency within the target domain. These two types of losses work together to ensure that our model not only harmonizes unpaired images but also preserves the continuity between slices through these domain transformations.

Our model, compared to other CNN and transformer models, offers advantages: fewer parameters, improved inter-slice consistency, and reduced computational demands. For 3D models, due to limitations in computational resources, an inevitable trade-off exists that could result in lower resolution in harmonized images. Thus, current 3D methods may not suit high-resolution image processing.

Furthermore, our model effectively preserves anatomical details in images following the harmonization process, while also accommodating stylistic variations across different sites, thereby facilitating further research endeavors. To validate this assertion, we conducted quantitative analyses of image similarity, assessing the resemblance between original and generated images using metrics such as MS-SSIM, PSNR, and bMMD<sup>2</sup>, and we further conducted a comprehensive analysis by comparing our model with others through error maps, regional histograms, and intensity standard deviations. The results demonstrate that the synthesized scans generated by our model exhibit higher reliability and accuracy compared to other models.

Preserving relevant biological and clinical patterns in the images without excessive correction is a crucial aspect of harmonization. In addition to the direct comparisons using quantitative metrics and qualitative evaluation, we performed an indirect evaluation of our model by assessing surface reconstruction, maintaining case/control effect size differences in hippocampal volumes, and brain age prediction on harmonized scans, comparing them to other methods. The outcomes highlight that harmonized scans produced by our model preserve anatomical details to a significant extent, ensuring their reliability for clinical analysis when used by healthcare professionals.

While our study has demonstrated superior performance compared to other methods, it is essential to acknowledge certain limitations. Identifying avenues for further improvement is crucial. (1) Enabling MRI harmonization when the target domain is inaccessible. (2) Achieving multi-domain harmonization simultaneously. (3) Utilizing

approaches with more scientifically grounded techniques for 2D slice reconstruction. (4) Adapting our model for 4D imaging techniques like fMRI and diffusion MRI. These enhancements will increase our model's usefulness across various scenarios.

## 6. Conclusion

Efficient 3D MRI harmonization is scarce for large-scale neuroimaging studies. In this study, we propose an unsupervised 3D MR image harmonization method that can minimize the inter-slice variances using warping field and two consistency losses. Instead of using 3D operators to replace the 2D operators in classical generative harmonization models, our model is more efficient and still achieves good harmonization results. First, we decompose the 3D MR volumes into a stack of slices, then treat the brain tissue between adjacent slices as an object with small warpings. In this way, the deviations between slices can be represented and constrained by the warping field. The biologically unrelated information (contrast, brightness, etc.) of the harmonized image is determined by the target domain, while the anatomical contents are extracted from the source domain. Our results show that the proposed method efficiently and successfully harmonize images from two domains, with consistent anatomical details in source domain and faithful representations in the target domain.

## CRedit authorship contribution statement

**Jiaying Lin:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zhuoshuo Li:** Investigation, Data curation. **Youbing Zeng:** Investigation, Data curation. **Xiaobo Liu:** Writing – original draft, Visualization. **Liang Li:** Visualization, Data curation. **Neda Jahanshad:** Investigation, Visualization. **Xinting Ge:** Validation, Data curation. **Dan Zhang:** Supervision, Writing – review & editing. **Minhua Lu:** Writing – review & editing, Supervision, Funding acquisition. **Mengting Liu:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

## Data availability

The data used in these experiments are available on application to the relevant studies. Source code used is released in <https://github.com/lx123-j/PWFHarmonization>.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to enhance language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Acknowledgments

This work was supported in part by the Shenzhen Science and Technology Innovation Program under (Grant JCYJ20220818102414031), “Young Innovation Team Program” of Shandong Provincial (Grant No. 2022KJ250), and Major Projects of Shanxi Province (Grant No. 202101030201001). This research has been conducted using the UK Biobank Resource under Application Number ‘11559’. Data used in preparation of this article were also obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>) and Nathan Kline Institute Rockland sample (NKI-RS), which are available online as part of the International Neuroimaging Data-sharing Initiative ([https://fcon\\_1000.projects.nitrc.org/indi/enhanced/](https://fcon_1000.projects.nitrc.org/indi/enhanced/)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. The dataset from the Nathan Kline Institute-Rockland Sample was rendered accessible through the auspices of the International Neuroimaging Data-sharing Initiative. The scholarly pursuits of Franziskus Liem received backing from the Swiss National Science Foundation (P2ZHP1 155200).

## References

- Bansal, A., Ma, S., Ramanan, D., Sheikh, Y., 2018. Recycle-GAN: Unsupervised video retargeting. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*. Springer International Publishing, Cham, pp. 122–138.
- Bayer, J.M.M., Thompson, P.M., Ching, C.R.K., Liu, M., Chen, A.A., Panzenhagen, A.C., Jahanshad, N., Marquand, A.F., Schmaal, L., Sämann, P.G., 2022. Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Front. Neurol.* 13, URL <https://api.semanticscholar.org/CorpusID:253239971>.
- Cackowski, S., Barbier, E.L., Dojat, M., Christen, T., 2021. Imunity: a generalizable VAE-GAN solution for multicenter MR image harmonization. *Med. Image Anal.* 88, 102799, URL <https://api.semanticscholar.org/CorpusID:237503710>.
- Chen, Y., Pan, Y., Yao, T., Tian, X., Mei, T., 2019. Mocycle-GAN: Unpaired video-to-video translation. In: *Proceedings of the 27th ACM International Conference on Multimedia*. URL <https://api.semanticscholar.org/CorpusID:201666368>.
- Dalmaz, O., Yurt, M., Çukur, T., 2022. ResViT: Residual vision transformers for multimodal medical image synthesis. *IEEE Trans. Med. Imaging* 41 (10), 2598–2614. <http://dx.doi.org/10.1109/TMI.2022.3167808>.
- Dinsdale, N.K., Jenkinson, M., Namburete, A.I., 2021. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage* 228, 117689. <http://dx.doi.org/10.1016/j.neuroimage.2020.117689>, URL <https://www.sciencedirect.com/science/article/pii/S1053811920311745>.
- Fischl, B., 2012. FreeSurfer. *NeuroImage* 62 (2), 774–781. <http://dx.doi.org/10.1016/j.neuroimage.2012.01.021>, 20 YEARS OF fMRI. URL <https://www.sciencedirect.com/science/article/pii/S1053811912000389>.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. USA* 97 (20), 11050–11055, URL <http://www.jstor.org/stable/123250>.
- Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* 167, 104–120. <http://dx.doi.org/10.1016/j.neuroimage.2017.11.024>, URL <https://www.sciencedirect.com/science/article/pii/S105381191730931X>.
- Guan, H., Liu, S., Lin, W., Yap, P.-T., Liu, M., 2022. Fast image-level MRI harmonization via spectrum analysis. In: Lian, C., Cao, X., Reikik, I., Xu, X., Cui, Z. (Eds.), *Machine Learning in Medical Imaging*. Springer Nature Switzerland, Cham, pp. 201–209.
- Gupta, U., Lam, P., Steeg, G., Thompson, P., 2021. Improved brain age estimation with slice-based set networks. pp. 840–844. <http://dx.doi.org/10.1109/ISBI48211.2021.9434081>.
- Hong, S., Marinescu, R., Dalca, A.V., Bonkhoff, A.K., Bretzner, M., Rost, N.S., Golland, P., 2021. 3D-StyleGAN: A style-based generative adversarial network for generative modeling of three-dimensional medical images. In: Engelhardt, S., Oksuz, I., Zhu, D., Yuan, Y., Mukhopadhyay, A., Heller, N., Huang, S.X., Nguyen, H., Sznitman, R., Xue, Y. (Eds.), *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*. Springer International Publishing, Cham, pp. 24–34.
- Hu, Q., Li, H., Zhang, J., 2022. Domain-adaptive 3D medical image synthesis: An efficient unsupervised approach. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (Eds.), *Medical Image Computing and Computer Assisted Intervention. MICCAI 2022*, Springer Nature Switzerland, Cham, pp. 495–504.
- Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2016. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1647–1655, URL <https://api.semanticscholar.org/CorpusID:3759573>.
- Isensee, F., Schell, M., Tursunova, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K., Kickingereder, P., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* 40, 4952–4964, URL <https://api.semanticscholar.org/CorpusID:59523686>.
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 5967–5976. <http://dx.doi.org/10.1109/CVPR.2017.632>.

- Jeong, H., Byun, H., Kang, D.U., Lee, J., 2023. BlindHarmony: “blind” harmonization for MR images via flow model. In: 2023 IEEE/CVF International Conference on Computer Vision. ICCV, pp. 21072–21082, URL <https://api.semanticscholar.org/CorpusID:258762749>.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2019. Analyzing and improving the image quality of StyleGAN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 8107–8116, URL <https://api.semanticscholar.org/CorpusID:209202273>.
- Keshavan, A., Paul, F., Beyer, M.K., Zhu, A.H., Papinutto, N., Shinohara, R.T., Stern, W.A., Amann, M., Bakshi, R., Bischof, A., Carriero, A., Comabella, M., Crane, J.C., D’alfonso, S., Demaerel, P., Dubois, B., Filippi, M., Fleischer, V., Fontaine, B., Gaetano, L., Goris, A., Graetz, C., Gröger, A., Groppa, S., Hafner, D.A., Harbo, H.F., Hemmer, B., Jordan, K.M., Kappos, L., Kirkish, G., Lufriu, S., Magon, S., Martinelli-Boneschi, F., McCauley, J.L., Montalban, X., Mühlau, M., Pelletier, D., Pattany, P.M., Pericak-Vance, M.A., Cournu-Rebeix, I., Rocca, M.A., Rovira, À., Schlaeger, R., Saiz, A., Sprenger, T., Stecco, A., Uitdehaag, B.M.J., Villoslada, P., Wattjes, M.P., Weiner, H.L., Wuerfel, J., Zimmer, C., Zipp, F., Hauser, S.L., Oksenberg, J.R., Henry, R.G., 2016. Power estimation for non-standardized multisite studies. *NeuroImage* 134, 281–294, URL <https://api.semanticscholar.org/CorpusID:3096020>.
- Komandur, D., Gupta, U., Chattopadhyay, T., Dhinagar, N.J., Thomopoulos, S.I., Chen, J.C., Beavers, D., Steeg, G.V., Thompson, P.M., 2023. Unsupervised harmonization of brain MRI using 3D CycleGANs and its effect on brain age prediction. In: 2023 19th International Symposium on Medical Information Processing and Analysis. SIPAIM, pp. 1–5. <http://dx.doi.org/10.1109/SIPAIM56729.2023.10373501>.
- Kwon, G., Han, C., Kim, D.S., 2019. Generation of 3D brain MRI using auto-encoding generative adversarial networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. URL <https://api.semanticscholar.org/CorpusID:199472854>.
- Lai, W.S., Huang, J.B., Wang, O., Shechtman, E., Yumer, E., Yang, M.H., 2018. Learning blind video temporal consistency. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*. Springer International Publishing, Cham, pp. 179–195.
- Li, J., Yang, Y., Zhang, F., Li, M., Qu, Z., Hu, S., 2022. TCGAN: A transformer-enhanced GAN for PET synthetic CT. *Biomed. Opt. Express* 13, <http://dx.doi.org/10.1364/BOE.467683>.
- Liu, M., Maiti, P., Thomopoulos, S., Zhu, A., Chai, Y., Kim, H., Jahanshad, N., 2021. Style transfer using generative adversarial networks for multi-site MRI harmonization. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), *Medical Image Computing and Computer Assisted Intervention. MICCAI 2021*, Springer International Publishing, Cham, pp. 313–322.
- Liu, S., Yap, P.T., 2024. Learning multi-site harmonization of magnetic resonance images without traveling human phantoms. *Commun. Eng.* 3 (1), 6. <http://dx.doi.org/10.1038/s44172-023-00140-w>.
- Liu, M., Zhu, A.H., Maiti, P., Thomopoulos, S.I., Gadewar, S., Chai, Y., Kim, H., Jahanshad, N., for the Alzheimer’s Disease Neuroimaging Initiative, 2023. Style transfer generative adversarial networks to harmonize multisite mri to a single reference image to avoid overcorrection. *Hum. Brain Mapp.* 44 (14), 4875–4892. <http://dx.doi.org/10.1002/hbm.26422>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.26422>.
- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. Ways toward an early diagnosis in Alzheimer’s disease: The Alzheimer’s disease neuroimaging initiative (ADNI). *Alzheimer’s Dementia* 1 (1), 55–66. <http://dx.doi.org/10.1016/j.jalz.2005.06.003>, URL <https://www.sciencedirect.com/science/article/pii/S155252600500004X>.
- Nieuwenhuis, M., Schnack, H.G., van Haren, N.E.M., Lappin, J.M., Morgan, C., Reinders, A.A., Tordesillas-Gutiérrez, D., Roiz-Santiañez, R., Schaufelberger, M.S., Rosa, P.G., Zanetti, M.V., Busatto, G.F., Crespo-Facorro, B., McGorry, P.D., Velakoulis, D., Pantelis, C., Wood, S.J., Kahn, R.S., Miranda, J.M., Dazzan, P., 2017. Multi-center MRI prediction models: Predicting sex and illness course in first episode psychosis patients. *Neuroimage* 145, 246–253, URL <https://api.semanticscholar.org/CorpusID:15314851>.
- Raju, J.C., Gayatri, K.S., Ram, K., Rangasami, R., Ramachandran, R., Sivaprakasam, M., 2021. MIST GAN: Modality imputation using style transfer for MRI. In: Lian, C., Cao, X., Rekić, I., Xu, X., Yan, P. (Eds.), *Machine Learning in Medical Imaging*. Springer International Publishing, Cham, pp. 327–336.
- Sinha, S., Thomopoulos, S.I., Lam, P., Muir, A.M., Thompson, P.M., 2021. Alzheimer’s disease classification accuracy is improved by MRI harmonization based on attention-guided generative adversarial networks. In: Proceedings of SPIE—the International Society for Optical Engineering, vol. 12088, URL <https://api.semanticscholar.org/CorpusID:236502514>.
- Sudlow, C.L.M., Gallacher, J.E., Allen, N.E., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M.J., Liu, B., Matthews, P.M., Ong, G.J., Pell, J.P., Sillman, A.J., Young, A., Sprosen, T., Peakman, T.C., Collins, R., 2015. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12, URL <https://api.semanticscholar.org/CorpusID:16996204>.
- Teed, Z., Deng, J., 2020. RAFT: Recurrent all-pairs field transforms for optical flow. In: European Conference on Computer Vision. URL <https://api.semanticscholar.org/CorpusID:214667893>.
- Tobe, R.H., MacKay-Brandt, A., Lim, R., Kramer, M., Breland, M.M., Tu, L., Tian, Y., Trautman, K.D., Hu, C., Sangoi, R., Alexander, L.M., Gabbay, V., Castellanos, F.X., Leventhal, B.L., Craddock, R.C., Colcombe, S.J., Franco, A.R., Milham, M.P., 2021. A longitudinal resource for studying connectome development and its psychiatric associations during childhood. *Sci. Data* 9, URL <https://api.semanticscholar.org/CorpusID:232204171>.
- van der Maaten, L., Hinton, G.E., 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605, URL <https://api.semanticscholar.org/CorpusID:5855042>.
- Wang, K., Akash, K., Misu, T., 2022. Learning temporally and semantically consistent unpaired video-to-video translation through pseudo-supervision from synthetic optical flow. In: AAAI Conference on Artificial Intelligence, vol. 36, (no. 3), pp. 2477–2486. <http://dx.doi.org/10.1609/aaai.v36i3.20148>.
- Wang, W., Yang, S., Xu, J., Liu, J., 2020. Consistent video style transfer via relaxation and regularization. *IEEE Trans. Image Process.* 29, 9125–9139. <http://dx.doi.org/10.1109/TIP.2020.3024018>.
- Wu, M., Zhang, L., Yap, P.-T., Lin, W., Zhu, H., Liu, M., 2024. Structural MRI harmonization via disentangled latent energy-based style translation. In: Cao, X., Xu, X., Rekić, I., Cui, Z., Ouyang, X. (Eds.), *Machine Learning in Medical Imaging*. Springer Nature Switzerland, Cham, pp. 1–11.
- Zhang, X., He, X., Guo, J., Etehadji, N., Aw, N., Semanek, D., Posner, J., Laine, A., Wang, Y., 2022. PTNet3D: A 3D high-resolution longitudinal infant brain MRI synthesizer based on transformers. *IEEE Trans. Med. Imaging* 41 (10), 2925–2940. <http://dx.doi.org/10.1109/TMI.2022.3174827>.
- Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2242–2251. <http://dx.doi.org/10.1109/ICCV.2017.244>.
- Zuo, L., Dewey, B.E., Liu, Y., He, Y., Newsome, S.D., Mowry, E.M., Resnick, S.M., Prince, J.L., Carass, A., 2021. Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. *NeuroImage* 243, 118569. <http://dx.doi.org/10.1016/j.neuroimage.2021.118569>, URL <https://www.sciencedirect.com/science/article/pii/S1053811921008429>.