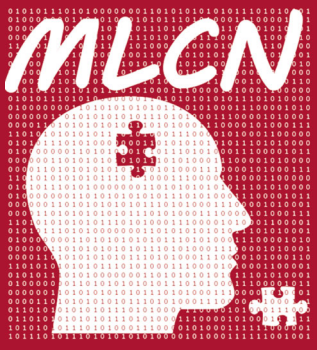
Deepti R. Bathula · Anoop Benet Nirmala · Nicha C. Dvornek · Sindhuja T. Govindarajan · Mohamad Habes · Vinod Kumar · Ahmed Nebli · Thomas Wolfers · Yiming Xiao (Eds.)

LNCS 15266

Machine Learning in Clinical Neuroimaging

7th International Workshop, MLCN 2024
Held in Conjunction with MICCAI 2024
Marrakesh, Morocco, October 10, 2024, Proceedings



Springer

MOREMEDIA

Lecture Notes in Computer Science **15266**

DISARM: Disentangled Scanner-Free Image Generation via Unsupervised Image2Image Translation

Luca Caldera[1], Lara Cavinato[1(✉)], Andrea Cappozzo[1], Isabella Cama[2], Sara Garbarino[3], Alessio Cirone[3], Raffaele Lodi[4], Fabrizio Tagliavini[5], Anna Nigri[6], Silvia De Francesco[7], Francesca Ieva[1], and RIN-Neuroimaging Network[5]

[1] MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy
lara.cavinato@polimi.it
[2] Department of Mathematics, Università di Genova, Genova, Italy
[3] IRCCS Ospedale Policlinico San Martino, Genova, Italy
[4] Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[5] Unit of Neurology (V) and Neuropathology, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[6] Neuroradiology Unit, IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
[7] Laboratory of Neuroinformatics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

Abstract. Ensuring reproducibility of Magnetic Resonance (MR) images from different scanners is crucial in multicenter studies, as scanner-induced variability is known to impact the results significantly. To address this problem, we introduce a novel unsupervised deep learning approach aimed at achieving 3 primary objectives/advantages: (1) create a *scanner-free* space that enables the uniform transfer of images across different scanners in a denoised setting, (2) impart the appearance of a specific training scanner to images, transferring its unique characteristics, (3) avoid time-consuming preprocessing of MR images. The proposed methodology is based on disentangling image information into two distinct spaces: one encoding the scanner-specific information and one capturing the anatomical/biological structure of the image. We trained our model on two open-source datasets (ADNI and PPMI) of T1-weighted brain MR images of normal control patients. We tested it

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-78761-4_10.

D. R. Bathula et al. (Eds.): MLCN 2024, LNCS 15266, pp. 102–112, 2025.
https://doi.org/10.1007/978-3-031-78761-4_10

on a real-world clinical dataset from the Italian Neuroimaging Network, comparing its performance with a state-of-the-art model. The results show the superiority of the proposed model in harmonizing images for clinical research, demonstrating its effectiveness in achieving consistent and reproducible harmonization of the MR images across (unseen) scanning environments. Code is available at https://github.com/luca2245/DISARM_Harmonization.

Keywords: Image harmonization · I2I Translation · Magnetic Resonance Imaging · Noise Disentanglement · Scanner-free Imaging

1 Introduction

With the rapid growth of brain Magnetic Resonance Imaging (MRI) datasets collected from various centers, the power of data-derived insights is becoming increasingly fundamental in understanding brain-related diseases, which can significantly enhance medical practice by providing robust statistical evidence. However, these centers often follow unharmonized acquisition protocols, leading to substantial variability in the extracted biomarkers. MRI scanners can indeed produce images with different contrast, brightness, and spatial characteristics due to variations in hardware, software, calibration, and other factors, including discrepancies that arise from machine maintenance, protocols, environmental conditions, and operator expertise. In multicenter studies involving diverse scanners and centers, this variability can confound results, highlighting the critical need for harmonized images to ensure robust analysis [19,22,27].

Several methodologies have been proposed for harmonizing images and specifically multicenter MRI data. Image-to-image (I2I) translation family includes methods that aim to map an input one into an output image while preserving certain semantic properties. Specifically, these generative models produce translated images, that appear to be drawn from the distribution of the target domain. They are categorized according to the type of supervision, i.e., supervised (directional or bidirectional) or unsupervised (cyclic consistency, autoencoders, disentangler), or to the type of translation, i.e., one-to-one, one-to-many, many-to-many [1]. Transformers [6] are models based on self-attention mechanism with the power of capturing long-range context dependencies. Still, they have been shown to overlook important high-frequency features that may be the site for biology- and pathology-related information [24]. Style transfer is a fully unsupervised deep learning framework [11], wherein image harmonization is treated as a style transfer problem rather than a domain transfer problem.

Only few attempts have been brought out in the medical domain, among all NeuroHarmonize [15], Calamiti [26], MURD [12] and ComBat [16,23]. At the crossroad between style transfer and I2I translation families, StarGANv2 [4] stands out as a single framework that addresses both the diversity of generated images and scalability across multiple domains, so far demonstrating significantly improved results compared to baseline methods. While showing promising results

in the harmonization task, this approach necessitates extensive preprocessing, including skull-stripping of the T1 images [18], nonuniformity correction [20], and registration to a standard space, making the end-to-end process heavy and intricate.

In this work, with the aim of overcoming inter-scanner variability of images to allow findings to be translated into clinical practice, we present DISARM, a novel model to harmonize 3D MR images by disentangling anatomical structure and scanner-specific information. The model is designed to produce *scanner-free* images while preserving the original anatomical structure and biologically informative data as to achieve robust generalizability across various scanners.

2 Proposed Methodology

We propose a harmonization model for 3D medical images extending upon DRIT++ network architecture [10], with specific application to T1-weighted brain MRI data. Particularly, our model targets the scanner-free generation of 3D MR images to clean out clinical images from batch effects.

2.1 Mathematical Formulation

Let $\mathcal{X} = \bigcup_{i=1}^{N} \mathcal{X}_i \in \mathbb{R}^{1 \text{ x } H \text{ x } W \text{ x } D}$ be a set of MR images and denote with $\mathcal{X}_i$ the set of images acquired from the $i-th$ scanner domain among N distinct domains. We assume that the images can be disentangled into two latent spaces $(\mathcal{B}, \mathcal{S})$, being $\mathcal{B}$ the space encoding the information related to the brain anatomical structure and being $\mathcal{S}$ the scanner space, specifically targeting the scanner effects. An image $\boldsymbol{x}$ drawn from $\mathcal{X}_i$ can be thus obtained as the combination of $\mathcal{B}$ and $\mathcal{S}$. The scanner-free harmonization task consists of noising out the scanner-specific information by plugging a random Gaussian noise $\mathcal{N}(0,1)$ instead of the $\mathcal{S}$ distribution. The final objective is to ensure robust generalizability across various scanners, including those not encountered during the training phase, and to deliver a model that does not necessitate time-consuming preprocessing of MR images.

2.2 Architecture

The backbone of the architecture includes two encoders and a generator. The brain structure encoder $E_b : \mathcal{X} \rightarrow \mathcal{B}$ maps an input image $\boldsymbol{x} \in \mathcal{X}$ into a lower-dimensional space $\mathcal{B}$ encoding the information related to the brain structure while the scanner encoder $E_s : (\mathcal{X}, \mathcal{C}) \rightarrow \mathcal{S}$ takes an input image $\boldsymbol{x} \in \mathcal{X}$ and its associated scanner label within the space $\mathcal{C}$ and operates as a variational autoencoder, producing a parametric distribution of the scanner effect. The scanner effect latent vector $\boldsymbol{z}_i^s$ is thus sampled as:

$$\boldsymbol{z}_i^s = \boldsymbol{\sigma}_i \cdot \boldsymbol{\varepsilon} + \boldsymbol{\mu}_i; \quad \boldsymbol{z}_i^s \in \mathcal{S}_i; \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0,1).$$

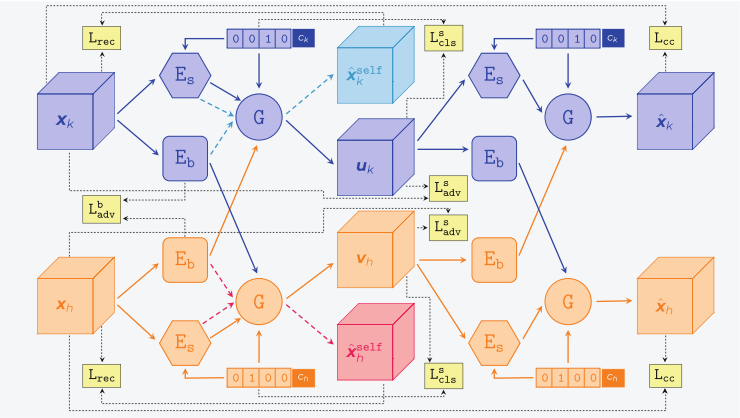


Fig. 1. High level functional diagram of DISARM network's training procedure. Discriminators, L_{lat}, and L_{KL} have not been displayed due to space reasons.

The generator $G : (\mathcal{B}, \mathcal{S}, \mathcal{C}) \rightarrow \mathcal{X}$ produces an image $\boldsymbol{x} \in \mathcal{X}$ that preserves a specified brain structure within the space $\mathcal{B}$ while incorporating a scanner attribute from the space $\mathcal{S}$ associated to its label within the space $\mathcal{C}$. To transfer a scanner attribute of scanner $\mathcal{X}_h$ to another image $\boldsymbol{x}_k \in \mathcal{X}_k$, the scanner latent space attribute $\boldsymbol{z}^s_h$ is plugged into the generator together with its label $c_h \in \mathcal{C}_h$ and the brain attribute $\boldsymbol{z}^b_{\boldsymbol{x}_k}$, obtaining the image $\boldsymbol{x}_k$ as it was acquired by the scanner h. Upon this strategy, we can achieve a scanner-free harmonization of images by sampling the scanner effect latent vector from $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)$ and plugging it into G together with input $\boldsymbol{x}$. The scanner-free space is denoted as $\mathcal{F}$ and the encoded scanner label is defined as a vector of zeros. Figures explaining the procedures are provided in the Supplementary Materials.

2.3 Losses

The training process is shown in Fig. 1 and is carried out on pairs of images $(\boldsymbol{x}_k, \boldsymbol{c}_k)$ and $(\boldsymbol{x}_h, \boldsymbol{c}_h)$ from randomly selected scanner domains (K and H) among a pool of N domains. Each image is simultaneously fed into its own E_b and E_s, yielding the corresponding latent representations $\boldsymbol{z}^s_l = E_s(\boldsymbol{x}_l, \boldsymbol{c}_l)$; $\boldsymbol{z}^b_{\boldsymbol{x}_l} = E_b(\boldsymbol{x}_l)$ for $l = \{h, k\}$. The training follows the minimization of a set of losses governing the different subnetworks in the architecture with reference to Fig. 1.

Brain Structure Adversarial Loss ($L^{\text{b}}_{\text{adv}}$)**:** It aims to ensure that the brain encoders E_b generate scanner-independent brain embeddings, that is, map images into a shared space $\mathcal{B}$ where no domain membership is distinguishable. To do this, the latent representations $\boldsymbol{z}^b_{\boldsymbol{x}_k}$ and $\boldsymbol{z}^b_{\boldsymbol{x}_h}$ are fed into a discriminator D_b which tries to identify their domain membership as follows:

$$L_{\text{adv}}^{\text{b}} = \frac{1}{2}\,\mathbb{E}_{\boldsymbol{x}_k}\Big[\log\Big[D_b(\boldsymbol{z}_{\boldsymbol{x}_k}^b)(1 - D_b(\boldsymbol{z}_{\boldsymbol{x}_k}^b))\Big]\Big] \;+ \frac{1}{2}\,\mathbb{E}_{\boldsymbol{x}_h}\Big[\log\Big[D_b(\boldsymbol{z}_{\boldsymbol{x}_h}^b)(1 - D_b(\boldsymbol{z}_{\boldsymbol{x}_h}^b))\Big]\Big]$$

Scanner Adversarial Loss ($L_{\text{adv}}^{\text{s}}$)**:** It aims to force the generators G to produce realistic images. To this purpose, the generated images $\boldsymbol{u}_k$ and $\boldsymbol{v}_h$ are fed into a discriminator D_s which tried to distinguish real images to generated images as follows:

$$L_{\text{adv}}^{\text{s}} = \frac{1}{2}\,\mathbb{E}_{\boldsymbol{x}_k}\Big[\log\Big[D_s(\boldsymbol{x}_k)(1 - D_s(\boldsymbol{u}_k))\Big]\Big] \;+\; \frac{1}{2}\,\mathbb{E}_{\boldsymbol{x}_h}\Big[\log\Big[D_s(\boldsymbol{x}_h)(1 - D_s(\boldsymbol{v}_h))\Big]\Big]$$

Scanner Classification Loss ($L_{\text{cls}}^{\text{s}}$)**:** The discriminator D_c is trained to predict the scanner label of the generated images, to ensure that its scanner-related information is compliant with the desired one (i.e., the one plugged in):

$$L_{\text{cls}}^{\text{s}} = \mathbb{E}_{\boldsymbol{x}_k}\Big[-\log\big[D_c(\boldsymbol{c}_k|\boldsymbol{u}_k)\big]\Big] \;+\; \mathbb{E}_{\boldsymbol{x}_h}\Big[-\log\big[D_c(\boldsymbol{c}_h|\boldsymbol{v}_h)\big]\Big]$$

Cycle Consistency Loss (L_{cc})**:** To maintain cross-cycle consistency, $\boldsymbol{u}_k$ and $\boldsymbol{v}_h$ undergo a backward translation process where they are re-generated in their own original scanner domain. The reconstructed images $\hat{\boldsymbol{x}}_k = G(E_b(\boldsymbol{v}_h), E_s(\boldsymbol{u}_k), \boldsymbol{c}_k)$ and $\hat{\boldsymbol{x}}_h = G\big(E_b(\boldsymbol{u}_k), E_s(\boldsymbol{v}_h), \boldsymbol{c}_h\big)$ need to be as maximally similar to the original $\boldsymbol{x}_k$ and $\boldsymbol{x}_h$ in terms of:

$$L_{cc} = \mathbb{E}_{\boldsymbol{x}_k,\boldsymbol{x}_h}\Big[\;\big\|\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k\big\| + \big\|\hat{\boldsymbol{x}}_h - \boldsymbol{x}_h\big\|\Big]$$

Self Reconstruction Loss (L_{rec})**:** It ensures that the reconstructed image $\hat{\boldsymbol{x}}_l^{\text{self}} = G(E_b(\boldsymbol{x}_l), E_s(\boldsymbol{x}_l), \boldsymbol{c}_l)$ closely matches the original image $\boldsymbol{x}_l$ for $l = h,\, k$. To do this, $\boldsymbol{x}_k$ and $\boldsymbol{x}_h$ are passed through the encoders (E_s, E_b) and the generator G and thus reconstructed as $\hat{\boldsymbol{x}}_k^{\text{self}}$ and $\hat{\boldsymbol{x}}_h^{\text{self}}$. The loss thus computes:

$$L_{\text{rec}} = \mathbb{E}_{\boldsymbol{x}_k,\boldsymbol{x}_h}\Big[\;\big\|\hat{\boldsymbol{x}}_k^{\text{self}} - \boldsymbol{x}_k\big\| + \big\|\hat{\boldsymbol{x}}_h^{\text{self}} - \boldsymbol{x}_h\big\|\Big]$$

Total Loss. The overall model loss is defined as follows:

$$L_{\text{tot}} = \lambda_{\text{cc}} L_{\text{cc}} + \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{lat}} L_{\text{lat}} + \lambda_{\text{KL}} L_{\text{KL}} - \lambda^{\text{b}}_{\text{adv}} L^{\text{b}}_{\text{adv}} - \lambda^{\text{s}}_{\text{cls}} L^{\text{s}}_{\text{cls}} - \lambda^{\text{s}}_{\text{adv}} L^{\text{s}}_{\text{adv}}$$

where L_{lat} serves to force the mean vector $\boldsymbol{\mu}_i$ extracted by the scanner encoder E_s to be close to a standard Gaussian, and L_{KL} is the Kullback-Leibler loss which aims to align the scanner effects representation with a prior Gaussian distribution. Values for parameters are provided in Supplementary Materials.

3 Experimental Setting

Datasets. We validated our approach using T1-weighted MR images from normal controls aged 60 to 80 years, with a slice thickness of 1 mm, obtained on a 3T scanner. In particular, we trained DISARM on a dataset comprising 335 images from Alzheimer's Disease Neuroimaging (ADNI3 study) [7] and Parkinson's Progression Marker Initiatives (PPMI) [13] datasets. Training images came from $N = 5$ different scanners (163 from `Prisma Fit` (**PrF**), 64 from `Prisma`, 30 from `Achieva dStream`, 38 from `Skyra`, and 40 from `TrioTim`). We tested the results on a clinical dataset from the Italian Neuroimaging Network (RIN), which we will refer to as the RIN dataset. This dataset comprises 117 images acquired with 7 scanners (17 from `Achieva dStream` (**Ac**), 20 from `DISCOVERY MR750` (**DI**), 23 from `Ingenia CX` (**In**), 40 from `Prisma` (**Pr**), 2 from `Signa HDxt` (**Si**), 12 from `Skyra` (**Sk**), 3 from `Skyra Fit` (**SkF**)). The cohort's age range is 49 to 80, with an average age of 64.5423 ± 8.6295.

Preprocessing. All the images were (1) normalized to the standard orientation to facilitate further steps, (2) applied a bias-field correction to mitigate the magnetic field variations in the MRI scanners [9,21], and (3) registered to the reference Standard MNI152-T1-1mm [8]. Moreover, we exploited a data augmentation procedure through dense random elastic deformation [14]. Implementation details are given in Supplementary Material.

Benchmarking. To benchmark our model, we evaluated the performances of the pre-trained StarGANv2 [4] on the RIN dataset. The required preprocessing included that the images were skull-stripped [18] and corrected for nonuniformity using the N3 method [20] in Freesurfer. Furthermore, the images needed to be linearly registered to a standard MNI template (ICBM 152 Nonlinear Symmetric atlas [5]) and resized to isotropic $1\,\text{mm}^3$ voxels. According to StarGANv2 implementation, harmonization was performed on a sliding window of 3 image's slices with a stride of 1. The final T1 harmonized MRI 3D volumes were then reconstructed by combining these harmonized partial volumes. As StarGANv2 model do not allow for scanner-free harmonization, the `Prisma Fit` scanner space was used as reference to be compared with the proposed model.

Evaluation Procedure. Three experiments were employed to test and compare models' performance. First, we evaluated whether the proposed model could accurately attribute the effects of one scanner to images captured by another. We thus compared grey level distributions of the output images harmonized on `Prisma Fit` scanner space by DISARM and StarGANv2. Second, we tested the differences in grey level distributions of the output images harmonized on $\mathcal{F}$ by DISARM. In both cases, we employed K-sample Anderson-Darling test [17]. Lastly, we evaluated the predictive performance in classifying patients' ages, to test the power of the harmonization model to enhance the biological information. We employed a regression model fed with the total gray volumes `TGV` [2,3] extracted from the RIN MR images before and after harmonization. We included the interactions with scanner variable, evaluating the significance of the scanner information in terms of p-values of the regression β coefficients. For this analysis, "before harmonization" also refers to before the preprocessing pipeline.

4 Results

The evaluation process involves harmonizing the 117 normal controls from the RIN dataset into both the `Prisma Fit` scanner space, using DISARM and StarGANv2, and the scanner-free space $\mathcal{F}$, using only DISARM. In Supplementary Materials, we provide example output images for both models.

4.1 Evaluation of `Prisma Fit` scanner information transfer

To test the harmonization task, we calculated the average grayscale distribution across all images for each scanner, before and after transferring the images to the `Prisma Fit` scanner space. We accounted resulting images from both the proposed model and StarGANv2. In Fig. 2, we display the boxplot of the grey-level distributions per scanner-specific image sets before and after harmonization by DISARM and StarGANv2. Of note, the harmonization of DISARM relies on the whole target scanner domain, whereas StarGANv2 relies on a single image as a reference. Therefore, the reference distribution (shown in gray) for DISARM is the average gray distribution of images acquired with `Prisma Fit`, whereas the StarGANv2 reference distribution is the one of a selected `Prisma Fit` image. The two plots (a) and (b) display results for DISARM, showing a massive alignment of the distributions after harmonization ($p = 0.39$). The two plots (c) and (d) show results for StarGANv2. Of note, (a) and (c) highlight the outcomes of the different preprocessing methods used for DISARM and StarGANv2, where the latter is much stronger. In (d), we notice how output distribution do not perfectly align with the reference image ($p = 0.03$).

4.2 Evaluation of Scanner Free Space $\mathcal{F}$

We tested the harmonization performance in the scanner-free space $\mathcal{F}$. In Fig. 3, we compare the average gray distributions for each scanner before and after the transfer to $\mathcal{F}$. We observe how the average post-harmonization distributions tend to align across all scanners, indicating successful harmonization ($p = 0.36$).

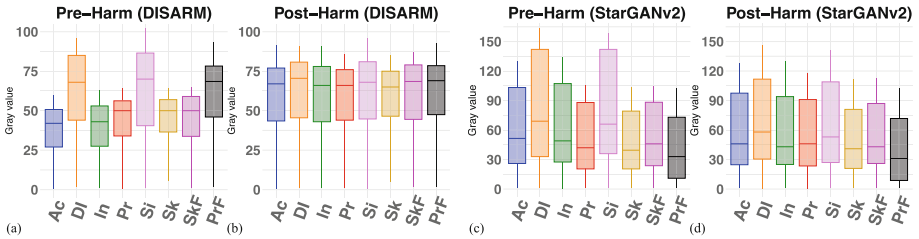


Fig. 2. Average gray distributions with Average Pool in each scanner pre- and post-transfer to `Prisma Fit` scanner space. DISARM result: (a) distributions after Preprocessing and (b) distributions post-harmonization; StarGAN result: (c) distributions after Preprocessing and (d) distributions post-harmonization. (Color figure online)

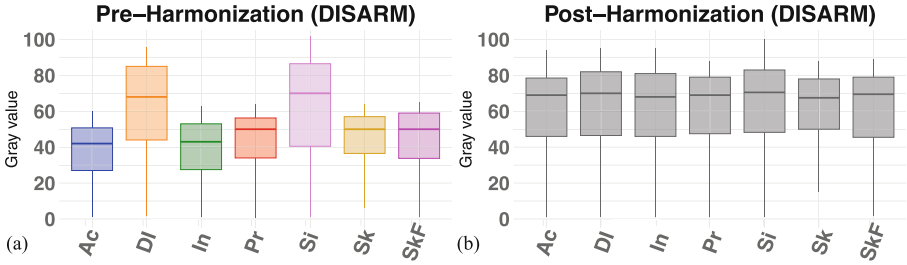


Fig. 3. The average gray distributions in each scanner before (a) and after (b) the transfer to Scanner-Free space $\mathcal{F}$. (Color figure online)

4.3 Age Prediction Task

In Table 1, we present the results of the regression models before and after harmonization for both DISARM and StarGANv2 models. Resulting regression lines are displayed in Supplementary Materials. Before harmonization, interactions with different scanners are significant ($p < 0.05$). After DISARM harmonization, the interaction with various scanners is no longer significant ($p > 0.05$), whereas after StarGANv2 harmonization, it remains weakly significant for two scanners ($p < 0.10$). This reflects the findings of the first experiment, highlighting how the scanner-related information still impact on the age regression task even after the harmonization procedure and sensibly contribute to the prediction. The Root Mean Squared Error (RMSE) before harmonization is 7.19. After harmonization with DISARM, it is 7.05, and after harmonization with StarGANv2, it is 7.45.

4.4 Preservation of Anatomical Structure

To evaluate the preservation of anatomical structures after harmonization, we use the Structural Similarity Index Measure (SSIM). We focus on the structural comparison between the original image and its harmonized version since we lack a ground truth for comparing also contrast and luminance [25]. For a set of 100

random image pairs from the RIN dataset, which exhibit different anatomical structures, the average SSIM value is 0.70 ± 0.07. The average SSIM between all 117 test images of the RIN dataset and their harmonized versions in the scanner-free space, obtained using DISARM, is 0.96 ± 0.007. When harmonization is performed using the `Prisma Fit` scanner as reference, the average SSIM value is 0.955 ± 0.008.

Table 1. Significance of Total Gray Volume (TGV) and TGV-scanner interactions in age regression models before and after DISARM and StarGANv2.

	Pre-Harmonization	DISARM	StarGANv2
Variables	P-value	P-value	P-value
(Intercept)	**<0.0001**	**<0.0001**	**<0.0001**
TGV	**<0.0001**	**<0.0001**	**<0.0001**
TGV-Achieva	**0.0321**	0.381	**0.078**
TGV-DISCOVERY	**0.0827**	0.101	0.332
TGV-Ingenia	**0.0170**	0.522	**0.081**
TGV-Skyra	0.8977	0.686	0.901

5 Conclusions

We introduced a novel method to harmonize T1-weighted MR images from various scanners, operating across the entire 3D volume to account for spatial correlations among different brain structures during reconstruction. Our approach allows for image transfer from different scanners in two distinct ways: (1) transferring images to a scanner-free space, ensuring consistent appearances regardless of the original scanner source; (2) mapping images to the space of one of the scanners used in the model's training, embedding the unique characteristics of the selected scanner into the transferred image. The model demonstrated robust generalization also to scanners not included in the training data, though the relatively small size of our training dataset indicates room for improvement with more data. Additionally, our method eliminates the need for time-consuming preprocessing steps like skull-stripping, which can be flawed and might remove parts of the brain or leave non-brain tissue remnants. This is particularly beneficial for applications requiring analysis of the entire head image, such as studies involving head trauma or skull deformities. Future expansions of the proposed approach would be carried out using travelling heads and include: (1) eliminating the bias field correction step to further streamline the preprocessing pipeline; (2) improve the model by integrating custom losses aimed at enhancing the fit for the scanner-free space; (3) improving the preservation of brain structure in harmonized images; and (4) extending the model to other imaging modalities and anatomical sites.

Acknowledgments. This study was funded by the Italian Minister of Health (RCR; RRC-2016-2361095; RRC-2017-2364915; RRC-2018-2365796; RCR-2019-23669119_001; RCR 2020-23670067; RCR 2022-23682285) and the Ministry of Economy and Finance (CCR-2017-23669078).

L. Cavinato is funded by the National Plan for NRRP Complementary Investments - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM).

The authors acknowledge the support by MUR grant Dipartimento di Eccellenza 2023-2027.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alotaibi, A.: Deep generative adversarial networks for image-to-image translation: a review. Symmetry **12**(10), 1705 (2020)
2. Bauer, T., et al.: Subcortical grey matter volume and asymmetry in the long-term course of Rasmussen's encephalitis. Brain Commun. **5**(6), fcad324 (2023)
3. Charroud, C., Turella, L.: Subcortical grey matter changes associated with motor symptoms evaluated by the unified Parkinson's disease rating scale (Part III): a longitudinal study in Parkinson's disease. NeuroImage: Clinical **31**, 102745 (2021)
4. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: Stargan v2: diverse image synthesis for multiple domains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8188–8197 (2020)
5. Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage **47**, S102 (2009)
6. Guo, Z., Gu, Z., Zheng, B., Dong, J., Zheng, H.: Transformer for image harmonization and beyond. IEEE Trans. Pattern Anal. Mach. Intell. (2022)
7. Jack Jr, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magnet. Resonance Imaging Off. J. Int. Soc. Magnet. Reson. Med. **27**(4), 685–691 (2008)
8. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage **17**(2), 825–841 (2002)
9. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
10. Lee, H.Y.: Drit++: diverse image-to-image translation via disentangled representations. Int. J. Comput. Vision **128**, 2402–2417 (2020)
11. Liu, M., et al.: Style transfer generative adversarial networks to harmonize multisite MRI to a single reference image to avoid overcorrection. Hum. Brain Mapp. **44**(14), 4875–4892 (2023)
12. Liu, S., Yap, P.T.: Learning multi-site harmonization of magnetic resonance images without traveling human phantoms. Commun. Eng. **3**(1), 6 (2024)
13. Marek, K., et al.: The Parkinson progression marker initiative (PPMI). Prog. Neurobiol. **95**(4), 629–635 (2011)
14. Pérez-García, F., Sparks, R., Ourselin, S.: Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput. Methods Programs Biomed. **208**, 106236 (2021)

15. Pomponio, R., et al.: Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage **208**, 116450 (2020)
16. Radua, J., et al.: Increased power by harmonizing structural MRI site differences with the combat batch adjustment method in enigma. Neuroimage **218**, 116956 (2020)
17. Scholz, F.W., Stephens, M.A.: K-sample anderson-darling tests. J. Am. Stat. Assoc. **82**(399), 918–924 (1987)
18. Ségonne, F., et al.: A hybrid approach to the skull stripping problem in MRI. Neuroimage **22**(3), 1060–1075 (2004)
19. Shinohara, R.T., et al.: Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. Am. J. Neuroradiol. **38**(8), 1501–1509 (2017)
20. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging **17**(1), 87–97 (1998)
21. Smith, S.M., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage **23**, S208–S219 (2004)
22. Takao, H., Hayashi, N., Ohtomo, K.: Effect of scanner in longitudinal studies of brain volume changes. J. Magn. Reson. Imaging **34**(2), 438–444 (2011)
23. Torbati, M.E., et al.: A multi-scanner neuroimaging data harmonization using ravel and combat. Neuroimage **245**, 118703 (2021)
24. Wang, P., Zheng, W., Chen, T., Wang, Z.: Anti-oversmoothing in deep vision transformers via the fourier domain analysis: from theory to practice. arXiv preprint arXiv:2203.05962 (2022)
25. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
26. Zuo, L., et al.: Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. Neuroimage **243**, 118569 (2021)
27. Zuo, X.N., Xu, T., Milham, M.P.: Harnessing reliability for neuroscience research. Nat. Hum. Behav. **3**(8), 768–771 (2019)