ESPA: An Unsupervised Harmonization Framework via Enhanced Structure Preserving Augmentation

Mahbaneh Eshaghzadeh Torbati[1] Davneet S. Minhas[2], Ahmad P. Tafti[3], Charles S. DeCarli[5], Dana L. Tudorascu[4], and Seong Jae Hwang[6]

[1] Intelligent Systems Program, University of Pittsburgh, USA
[2] Department of Radiology, University of Pittsburgh, USA
[3] Department of Health Information Management, University of Pittsburgh, USA
[4] Department of Psychiatry, University of Pittsburgh, USA
{mae82, dam148, tafti.ahmad, dlt30}@pitt.edu
[5] Department of Neurology, University of California Davis, USA
cdecarli@ucdavis.edu
[6] Department of Artificial Intelligence, Yonsei University, Republic of Korea
seongjae@yonsei.ac.kr

Abstract. The rising interest in pooling neuroimaging data from various sources presents challenges regarding scanner variability, known as *scanner effects*. While numerous harmonization methods aim to tackle these effects, they face issues with model robustness, brain structural modifications, and over-correction. To combat these issues, we propose a novel harmonization approach centered on simulating scanner effects through augmentation methods. This strategy enhances model robustness by providing extensive simulated matched data, comprising sets of images with similar brain but varying scanner effects. Our proposed method, *ESPA*, is an unsupervised harmonization framework via Enhanced Structure Preserving Augmentation. Additionally, we introduce two domain-adaptation augmentation: tissue-type contrast augmentation and GAN-based residual augmentation, both focusing on appearance-based changes to address structural modifications. While the former adapts images to the tissue-type contrast distribution of a target scanner, the latter generates residuals added to the original image for more complex scanner adaptation. These augmentations assist ESPA in mitigating over-correction through data stratification or population matching strategies during augmentation configuration. Notably, we leverage our unique in-house matched dataset as a benchmark to compare ESPA against supervised and unsupervised state-of-the-art (SOTA) harmonization methods. Our study marks the first attempt, to the best of our knowledge, to address harmonization by simulating scanner effects. Our results demonstrate the successful simulation of scanner effects, with ESPA outperforming SOTA methods using this harmonization approach.

1 Introduction

Interest is rising in pooling neuroimaging data from multiple sites to make larger datasets. Although this boosts statistical power, it meets challenges like *scanner effects* [1]. These effects, from variations within or among sites, can bias neuroimaging measures and affect how clinical signals are understood [2]. To tackle this, different harmonization methods have been suggested. Harmonization methods tackle scanner effects from various perspectives. One view focuses on harmonization of images [3–17] or image-derived measures [1, 18–24]. The latter adjusts measure distributions across scanners better but lacks harmonization accuracy information compared to the former [25]. Another perspective views harmonization as either a standalone preprocessing step or an integrated part of methods targeting specific tasks. These are known as task-agnostic [1, 3–15, 18–24] and task-specific [16, 17, 23, 24] harmonizations, respectively. While task-specific methods benefit from task-related information, they may lack generalizability. Despite these views, two main harmonization approaches prevail: (1) removing scanner effects from data, and (2) adapting data to a scanner domain. In the former, scanner effects are treated as estimable variability to be eliminated [5,18–20]. In the latter, scanner effects are viewed as causing domain shift, with harmonization achieved by adapting data to either a scanner-middle-ground domain [3, 4], the domain specific to a *target scanner* [6–11], or the scanner-variant component of data for a *target individual* [12–14].

In practice, harmonization methods encounter two main challenges: (1) over-correction, where biological variables may be inadvertently removed alongside scanner effects [13], and (2) brain structural modifications, resulting in alterations to the brain's structure [3]. To mitigate these issues, certain methods confine harmonization to image contrast or style. For instance, CALAMITI [14] aims to harmonize images by adapting them to the contrast of images from a target scanner. However, this risks over-correction when significant biological differences exist across scanner domains. Style-transfer harmonization [13] adjusts images to match the style of an *individual* target image, thus addressing population-wide over-correction concerns. Still, this presents a potential risk of conveying biological information, such as white matter hyperintensity, through image style [26]. Using *matched data* is a key strategy to tackle these challenges, as *matched images* portray a biologically similar brain with differences solely attributed to scanner effects [3]. These voxel-wise differences are used as supervision for the task of harmonization. Harmonization methods utilizing matched data, classified as supervised methods, offer lower susceptibility to over-correction and brain modifications, as they directly address scanner effect [4]. However, their applicability is limited to datasets with available matched data, potentially affecting their robustness due to the small size of matched data [3].

A more straightforward harmonization perspective involves simulating scanner effects through augmentation methods, a direction explored in this study. These scanner-specific augmentation methods can be applied within self-supervised augmentation-based frameworks [27] to generate scanner-free pretext or used to simulate matched data for pretraining harmonization methods. Such applications

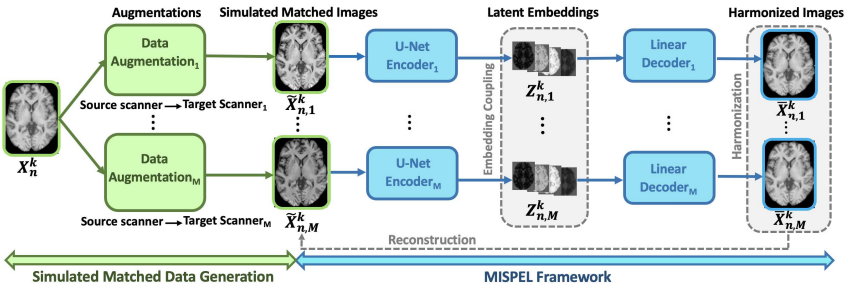

Fig. 1: Illustration of ESPA

contribute to addressing potential robustness issues in supervised harmonization models. Our main **contributions** are summarized as follows:

1. Proposal of ESPA, an unsupervised harmonization framework via **E**nhanced **S**tructure **P**reserving **A**ugmentation.
2. Introducing two novel augmentation methods to simulate scanner effects, effectively addressing brain structural changes and over-correction issues.
3. Utilization of our unique in-house matched dataset as a benchmark to compare ESPA against both supervised and unsupervised SOTA harmonization methods. Our code is available at https://github.com/Mahbaneh/ESPA.git.

2 Method

ESPA (Fig. 1), crafted as an unsupervised task-agnostic image-harmonization framework, adapts images to a scanner-middle-ground domain. Using the supervised harmonization method MISPEL [3] as our harmonization framework, we made a key modification: opting for simultaneously generating and using simulated matched images in training rather than using matched data. These simulated matched images are generated using our augmentation methods.

Notations and Assumptions. We refer to the data targeted for harmonization as *multi-scanner data*. This data contains images of M scanners. We consider another set of data with images of one arbitrary scanner and refer to it as *source data*. Throughout the manuscript, we refer to scanners of the source and multi-scanner data as the *source scanner* and *target scanners*, respectively. Source data, $X_{n=1:N}$, consists of N images with a total of $X^{k=1:K}_{n=1:N}$ slices where K is the number of axial slices of an image. Our goal is to design augmentation methods to adapt images of the *source data* to those of the M scanners in the *multi-scanner data*. These methods can then be applied to the slices in *source data*, $X^{k=1:K}_{n=1:N}$, to generate our desired simulated matched data. We refer to this simulated set as $\tilde{X}^{k=1:K}_{n=1:N,m=1:M}$ in which $\tilde{X}^{k}_{n,m=1:M}$ are matched slices for X^k_n. ESPA uses the augmented methods to sample *variations* of such data during its training to learn generating their harmonized images $\bar{X}^{k=1:K}_{n=1:N,m=1:M}$, where $\bar{X}^k_{n,1} \approx \cdots \approx \bar{X}^k_{n,m} \approx \cdots \approx \bar{X}^k_{n,M}$ (for all n samples and k axial slices).

2.1 MISPEL

MISPEL (Fig. 1) specializes in harmonizing images of scanners with matched data. It uses encoder-decoder units for each scanner, translating input slices into latent embeddings using a U-Net [30] encoder. Linear decoding combines these embeddings to reconstruct the input image, ensuring similarity between embeddings and reconstructed images across scanners for harmonization. Also, MISPEL maintains brain structure by ensuring similarity between reconstructed and original images. These were respectively referred to as Embedding Coupling, Harmonization, and Reconstruction in MISPEL [3] and our Fig. 1.

2.2 Tissue-type contrast augmentation

Scanner effects can alter brain tissue contrast [21]. Therefore, we employ a three-step augmentation method to adapt tissue contrast from a source scanner to *one* target scanner while preserving brain structure. This involves modifying an augmentation method initially designed for brain segmentation [28].

Step 1: Estimating the distributions of tissue types. In this step, we apply the Gaussian Mixture Model [29] to the intensity values of the brain voxels in source image X_n. The intensity set $\{v_1, \ldots, v_P\}$, where P is the total number of brain voxels in the image, is modeled as $p(v_p) = \Sigma_{t=1}^{T=3} \pi_t \mathcal{N}(v_p|\mu_t, \sigma_t^2)$, with t denoting brain tissue types, $\mathcal{N}(\mu_t, \sigma_t^2)$ representing a Gaussian distribution with mean μ_t and variance σ_t^2, and π_t as the mixing coefficient. Using Bayes' rule, we compute the probability of each class label C as $p(C = t|v) = \frac{\pi_t \mathcal{N}(v|\mu_t, \sigma_t^2)}{\Sigma_{t'=1}^{3} \pi_{t'} \mathcal{N}(v|\mu_{t'}, \sigma_{t'}^2)}$.

Step 2: Modifying tissue type distributions. We adapt this method step to align images from our *source data* with those of a *single* target scanner in our *multi-scanner data*. To achieve this, we adjust the tissue type distributions of images in the source data by sampling from estimated normal distributions capturing directional differences in tissue-type parameters between the source data images and those of the target scanner. The desired modified tissue type distribution of the source image X_n is determined as $\mathcal{N}(\mu_t^{'}, \sigma_t^{'2}) = (\mu_t - q_{\mu_t}, \sigma_t^2 - q_{\sigma_t^2})$, where q_{μ_t} and $q_{\sigma_t^2}$ are adaptation terms sampled from the determined distributions of differences. To calculate these terms, we first compute directional differences of distribution parameters from all source images to all target images. One instance of such differences is denoted as $D_{\mu_t}^f = \mu_{n,t} - \mu_{l,t}$ and $D_{\sigma_t^2}^f = \sigma_{n,t}^2 - \sigma_{l,t}^2$, where $(\mu_{n,t}, \sigma_{n,t}^2)$ and $(\mu_{l,t}, \sigma_{l,t}^2)$ are distribution parameter pairs for images n and l in the source data and target scanner, respectively, and f denotes the f^{th} difference in a total of F calculated differences. Finally, we compute the adaptation terms as $q_{\mu_t} = Mean(D_{\mu_t}^{f=1:F}) + r_\mu$ and $q_{\sigma_t^2} = Mean(D_{\sigma_t^2}^{f=1:F}) + r_\sigma$, where r_μ and r_σ are sampled from the uniform distributions $U(-Std(D_{\mu_t}^{f=1:F}), Std(D_{\mu_t}^{f=1:F}))$ and $U(-Std(D_{\sigma_t^2}^{f=1:F}), Std(D_{\sigma_t^2}^{f=1:F}))$, and $Mean(\cdot)$ and $Std(\cdot)$ denote the mean and standard deviation functions.

Step 3: Reconstructing augmented image. For reconstructing the augmented image for source image X_n, we adapt each voxel value v_p for each tissue type as $v'_{p,t} = \mu'_t + d_{pt}\sigma'_t$, where $d_{pt} = (v_p - \mu_t)/\sigma_t$ maintains the original relative

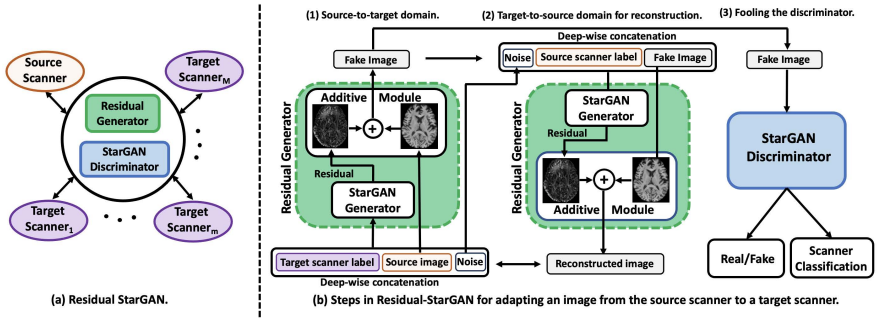

Fig. 2: Illustration of Residual StarGAN

distance of voxel intensity from the mean intensity of tissue type t in the images, preserving structural brain information. We then compute the augmented intensity voxel v'_p as $v'_p = \Sigma_{t=1}^{T=3} p(C = t|v_p)v'_{p,t}$. After calculating all v'_p for $p \in \{1, \ldots, P\}$, we obtain the adapted image of X_n aligned with the tissue-type distribution of our single target scanner.

2.3 GAN-based residual augmentation

Scanner effects can be more intricate than tissue-type modifications. Thus, we develop a GAN-based augmentation method to generate and sample scanner effects as images (*residuals*) added to the original images. By applying scanner effects as additive components to images, we consider brain structure during augmentation. For this purpose, we introduce Residual StarGAN, which performs image-to-image translation between all pairs of our scanner domains (source and target scanners) using a single generator and discriminator pair (Fig. 2(a)). Residual StarGAN is a modification of StarGAN [31], where we replace the generator with a *Residual Generator*, and include noise as input to this generator for sampling. The Residual Generator comprises the StarGAN Generator followed by the *Additive Module* (Fig. 2(b)). The StarGAN Generator generates the residuals to be added to the image in the Additive Module for domain adaptation. The process of adapting an image from the source scanner to the domain of a target scanner in Residual StarGAN is depicted in Fig. 2(b). These steps mirror those outlined in StarGAN, with details provided in its original paper [31]. We utilize the trained Residual Generator as our residual augmentation method.

3 Experiments and Results

Scanners and Datasets. We utilized an in-house matched dataset of 3T T1 images across four scanners: General Electrics (GE), Philips, Siemens Prisma (SiemP), and Siemens Trio (SiemT). We reported more details on specification of scanners in this data in Supp. Table 1. This data was collected from 18 subjects with a median age of 72 years (range 51-78), 44% of whom were male, all

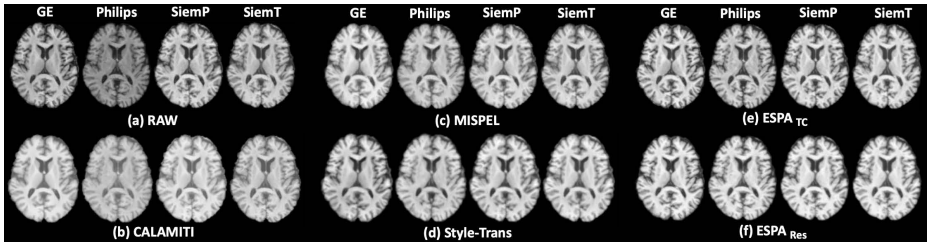

Fig. 3: Visual assessment of scanner effects and harmonization across matched images.

cognitively unimpaired, with 10 individuals exhibiting a high degree of small vessel disease (SVD). We selected this data as our *multi-scanner* data and did not use its supervision (matched aspect) for multi-scanner data during ESPA training. We only used its matched aspect for evaluating harmonization. We chose *source data* of 192 T1 images from a 3T Siemens Trio scanner in the OAISIS-3 dataset [32]. We aligned the demographics of source data to that of multi-scanner data to address over-correction during configuring augmentations. We applied preprocessing to both datasets, including non-linear registration to a T1 atlas [33], N4 bias correction [34], skull-stripping through brain masking, and image scaling by dividing images by their mean intensity. We refer to the preprocessed matched data as RAW.

Baselines and Training Setup. For SOTA, we employed style-transfer harmonization (referred to as Style-Trans) [13] as unsupervised, and CALAMITI [14] and MISPEL [3] as supervised methods. We slightly modified CALAMITI for training it as a supervised harmonization method. For Style-Trans, we directly applied their released pre-trained model to RAW. For CALAMITI and MISPEL, we conducted 6-fold cross-validation for RAW, splitting subjects into 12/3/3 for train/validation/test sets. ESPA follows a 3-step process, using the same subject-level cross-validation splits for the preprocessed multi-scanner data. It uses two sets from source data: 12/20/20 and 100/20/20 splits of train/validation/test sets for its first two steps, respectively. **(1)** In the initial step, two augmentation methods are configured individually to adapt images of 12 training source images to 12 training images within each of the 4 scanners in the multi-scanner data. **(2)** The second step involves training ESPA by creating variations of simulated matched data, individually applying augmentations to the 100 source training images considered for this step. Separate sets of ESPA models, referred to as ESPA_{TC} and ESPA_{Res}, are trained for each augmentation. **(3)** In the final step, these models are individually applied to images of 3 test subjects in the multi-scanner data. These three steps were repeated for each cross-validated folds. Harmonized test sets are then combined across folds as one set of harmonized multi-scanner data for evaluation. Model training and hyper-parameter tuning were conducted on NVIDIA RTX5000 within suggested ranges from the original papers: [3] for CALAMITI and MISPEL, and [3, 28, 31] for ESPA.

3.1 Results

Validation on domain adaptation in augmentation. To evaluate this, we tested scanner classification performance on augmented images. For each fold of the multi-scanner data, we trained, optimized, and evaluated a multi-class scanner classifier using the train/validation/test images from both the multi-scanner data as well as the source data designated to the first step of ESPA. To ensure balanced classification, we used images of solely 3 subjects from each of the source validation and test sets. For the classifier, we used the discriminator network in [8]. The classifiers' cross-fold accuracy averaged $78.6 \pm 1.9\%$, with accuracies of [85.2±5.7, 81.±2.5, 73.9±4.1, 74.4±4.2]% for the target scanner set: [GE, Philips, SiemP, SiemT], respectively. For the augmented images, we applied the configured augmentations of each fold individually to our 20 source test images considered for the first step in ESPA. The classifiers were then applied to the augmented images, resulting in an average cross-fold accuracy of $88.2 \pm 3.9\%$ for tissue-type contrast augmentation, with accuracies of [86.1 ± 6.9, 86.5 ± 7.2, 91.3 ± 2.8, 89.0 ± 3.4]% for target scanners. Similarly, for residual augmentation, the accuracy averaged $88.1 \pm 3.9\%$, with accuracies of [86.4 ± 5.7, 84.3 ± 2.5, 92.3 ± 4.1, 89.4 ± 4.2]% for the target scanners. Despite the classifier's limited performance due to the small training image size, these results highlight the effectiveness of our augmentation methods in domain adaptation.

Validation on augmentation removal in ESPA. We assessed augmentation removal for our cross-validated models for ESPA_{TC} and ESPA_{Res}. Removal aimed to decrease dissimilarity between augmented images of a source image. Mean Average Error (MAE) and Jensen–Shannon Divergence (JD) metrics were used to assess this dissimilarity, reported as mean ± standard deviation (SD) for images of all scanner pairs and folds. Initially, we augmented images of the source test set designated for the second step in ESPA, using the configured augmentation of each fold. Then, the trained ESPA models of each fold were applied to their corresponding augmented image sets to obtain augmented-free (harmonized) images. For ESPA_{TC}, MAE and JD decreased from 0.071 ± 0.037 to $\mathbf{0.030 \pm 0.009}$, and 0.023 ± 0.030 to $\mathbf{0.012 \pm 0.015}$ before and after harmonization, respectively. Similarly, for the ESPA_{Res}, MAE values decreased from 0.403 ± 0.107 to $\mathbf{0.135 \pm 0.023}$, and JD values decreased from 0.012 ± 0.009 to $\mathbf{0.007 \pm 0.006}$. All changes were statistically significant (paired t-test, $p < 0.05$), indicating successful augmentation removal from images.

Validation on harmonization. A robust harmonization method effectively addresses scanner effects while preserving or enhancing biological signals. We assessed scanner effects and harmonization through *dissimilarity* and *increased similarity* within matched images, respectively. Accordingly, *vxisual*, *structural*, and *biological* similarities were examined for both RAW and harmonized RAW. Scanner effects were ***visually*** evident in Fig. 3 as cross-scanner contrast dissimilarity, which reduced after harmonization. CALAMITI disturbed image contrast, while MISPEL and Style-Trans slightly smoothed images, and ESPA_{TC} and ESPA_{Res} provided better visual quality. Structural similarity increased significantly, from 0.81 ± 0.05 for RAW to 0.83 ± 0.04, $\mathbf{0.87 \pm 0.04}$, $\mathbf{0.87 \pm 0.05}$,

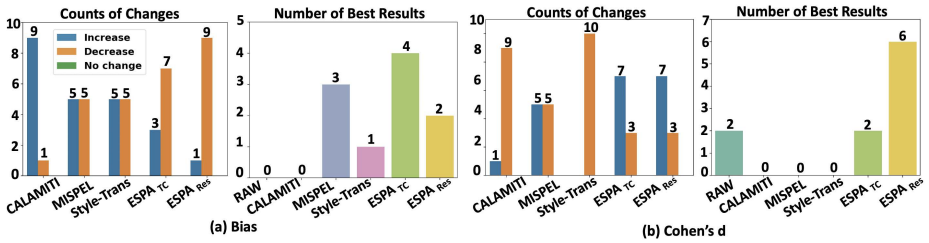

Fig. 4: Detailed statistics on bias and Cohen's d for biomarkers of AD.

0.83 ± 0.05, and 0.85 ± 0.05 for CALAMITI, Style-Trans, MISPEL, ESPA_{TC}, and ESPA_{Res}, respectively, with MISPEL and Style-Trans having the largest increase. All the increases compared to RAW were statistically significant where paired t-test ($p < 0.05$) was used.

We assessed the ***biological similarity*** of the top 10 FreeSurfer-derived [35] AD biomarkers [36]. Bias, computed as mean absolute differences across scanners, focused on cortical thickness and volumes detailed in Supp. Table 2. Harmonization was confirmed if bias decreased compared to RAW. Results (Fig. 4) revealed CALAMITI worsened bias for 9 biomarkers. ESPA_{TC} and ESPA_{Res} outperformed MISPEL and Style-Trans, reducing bias for **7** and **9** biomarkers compared to 5 for MISPEL and Style-Trans. ESPA_{TC} had the largest decreases for **4** cases, while MISPEL, ESPA_{Res}, and Style-Trans had 3, 2, and none. Paired t-tests ($p < 0.05$) showed significant decreases in 5, 5, and 4 for ESPA_{TC}, ESPA_{Res}, and MISPEL, respectively. None were significant for Style-Trans.

Finally, we explored whether harmonization preserved or enhanced biological signals by comparing Cohen's d effect sizes between low and high SVD groups for each AD biomarker. Cohen's d was computed separately for each scanner, and the mean±SD across scanners was reported in Supp. Table 2. Harmonization success was determined by an increase in Cohen's d compared to RAW. Our findings (Fig. 4) revealed CALAMITI and Style-Trans's failure, possibly due to deteriorated contrast and over-correction. ESPA_{TC} and ESPA_{Res} each surpassed MISPEL with **7** increases, while yielding the best Cohen's d values for 2 and **6** biomarkers, respectively, compared to MISPEL's 5 increases.

Ablation study. To demonstrate the efficacy of our augmentation methods, we trained ESPA with random contrast and brightness augmentation [37]. These techniques involve contrast transformation $(X_n^A - E(X_n)) * b + E(X_n)$ and brightness transformation $X_n + c$, where b and c are uniformly sampled from $[0.8, 1.2]$ and $[-0.1, 0.1]$, respectively, with $E(X_n)$ representing the mean brain intensity values in X_n. We repeated experiments for validation on *augmentation removal*, confirming reduction in MAE and JD from 0.164 ± 0.088 and 0.028 ± 0.025 to $\mathbf{0.099 \pm 0.037}$ and $\mathbf{0.013 \pm 0.016}$, respectively. However, our structural similarity analysis for *harmonization* yielded SSIMs similar to that of RAW, suggesting no significant modification and thus no harmonization.

4 Discussion and Conclusion

In this paper, we propose ESPA, an unsupervised image harmonization framework addressing common issues with current methods: model robustness, brain structural modification, and over-correction. ESPA simulates scanner effects on plentiful images using two novel structure-preserving augmentation methods, enabling harmonization through the adaptation of augmented images to a scanner-middle-ground space in which brain structure is preserved. It also addresses over-correction through population matching during simulation. Results indicate that our augmentation methods successfully simulate scanner effects and ESPA performed at least as well as SOTA harmonization methods. However, the performance of our ESPA models is limited to the scanner types on which they were trained. Future work will further explore these augmentation methods on larger multi-site neuroimaging data within a self-supervised framework.

Acknowledgements. This work was supported in part by the National Institutes of Health (NIH) under Grants R01 AG063752, P30 AG10129, and UH3 NS100608. SJH work was supported in part by the IITP 2020-0-01361 (AI Graduate School Program at Yonsei University), NRF RS-2024-00345806, and NRF RS-2023-00219019 funded by Korean Government (MSIT).

Disclosure of Interests. The authors have no competing interests.

References

1. Fortin, J. P., et al.: Harmonization of multi-site diffusion tensor imaging data. Neuroimage. **161**, 149-170 (2017)
2. Shinohara, R.T., et al.: Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. American Journal of Neuroradiology. **38**(8), 1501-1509 (2017)
3. Eshaghzadeh Torbati, M., et al.: MISPEL: A supervised deep learning harmonization method for multi-scanner neuroimaging data. Medical image analysis. **89**, 102926 (2023)
4. Dewey, B.E., et al.: DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. Magnetic resonance imaging. **64**, 160-170 (2019)
5. Fortin, J.P., et al.: Removing inter-subject technical variability in magnetic resonance imaging studies. NeuroImage. **132**, 198-212 (2016)
6. Sederevičius, D., et al.: A robust intensity distribution alignment for harmonization of T1-w intensity values. bioRxiv. 2022-06 (2022)
7. Modanwal, G.: MRI image harmonization using cycle-consistent generative adversarial network. In Medical Imaging 2020: Computer-Aided Diagnosis (Vol. 11314, pp. 259-264). SPIE (2020)
8. Bashyam, V.M., et al.: Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors. Journal of Magnetic Resonance Imaging. **55**(3), 908-916 (2022)

9. Chang, X., et al.: Self-supervised learning for multi-center magnetic resonance imaging harmonization without traveling phantoms. Physics in Medicine & Biology. **67**(14), 145004 (2022)
10. Fatania, K., et al.: Harmonisation of scanner-dependent contrast variations in magnetic resonance imaging for radiation oncology, using style-blind auto-encoders. Physics and Imaging in Radiation Oncology. **22**, 115-122 (2022)
11. Moyer, D., et al.: Scanner invariant representations for diffusion MRI harmonization. Magnetic resonance in medicine. **84**(4), 2174-2189 (2020)
12. Tian, D., et al.: A deep learning-based multisite neuroimage harmonization framework established with a traveling-subject dataset. NeuroImage,. **257**, 119297 (2022)
13. Liu, M., Zhu, et al.: Style transfer generative adversarial networks to harmonize multisite MRI to a single reference image to avoid overcorrection. Human Brain Mapping. **44**(14), 4875-4892 (2023)
14. Zuo, L., et al.: Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. NeuroImage. **243**, 118569 (2021)
15. Zhao, F., et al.: Disentangling Site Effects with Cycle-Consistent Adversarial Autoencoder for Multi-site Cortical Data Harmonization. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 369-379). Cham: Springer Nature Switzerland (2023)
16. Aslani, S., et al.: Scanner invariant multiple sclerosis lesion segmentation from MRI. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (pp. 781-785). IEEE (2020)
17. Dinsdale, N.K., Jenkinson, M., Namburete, A.I.: Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. NeuroImage. **228**, 117689 (2021)
18. Reynolds, M., et al.: Combat harmonization: Empirical bayes versus fully bayes approaches. NeuroImage: Clinical. **39**, 103472 (2023)
19. Garcia-Dias, R., et al.: Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. Neuroimage. **220** (2020)
20. Chen, A.A., et al.: Privacy-preserving harmonization via distributed ComBat. NeuroImage. **248**, 118822 (2022)
21. Meyer, M.I., et al.: Relevance vector machines for harmonization of MRI brain volumes using image descriptors. In International Workshop on OR 2.0 Context-Aware Operating Theaters (pp. 77-85). Cham: Springer International Publishing (2019)
22. Wang, R., Chaudhari, P., Davatzikos, C.: Harmonization with flow-based causal inference. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France Proceedings, Part III 24 (pp. 181-190). Springer International Publishing (2021)
23. An, L., et al.: Goal-specific brain MRI harmonization. Neuroimage. **263**, 119570 (2022)
24. Bayer, J.M., et al.: Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. NeuroImage. **264**, 119699 (2022)
25. Eshaghzadeh Torbati, M., et al.: A multi-scanner neuroimaging data harmonization using RAVEL and ComBat. Neuroimage. **245**, 118703 (2021)
26. Debette, S., Markus, H.S: The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. Bmj, **341** (2010)

27. Chen, T., et al.: A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR (2020)
28. Meyer, M.I., et al.: A contrast augmentation approach to improve multi-scanner generalization in MRI. Frontiers in neuroscience. **15**, 708196 (2021)
29. Reynolds, D. A. (2009). Gaussian mixture models. Encyclopedia of biometrics, 741(659-663).
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing (2015)
31. Choi, Y., et al.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797), (2018)
32. LaMontagne, P.J., et al.: OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. MedRxiv (2019)
33. Oishi, K., et al.: Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. Neuroimage. **46**(2), 486-499 (2009)
34. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE transactions on medical imaging. **29**(6), 1310-1320 (2010)
35. Fischl, B. (2012). FreeSurfer. Neuroimage, **62**(2), 774-781.
36. Schwarz, C.G., et al.: A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. NeuroImage: Clinical. **11**, 802-812 (2016)
37. Chaitanya, K., et al.: Semi-supervised task-driven data augmentation for medical image segmentation. Medical Image Analysis, **68**, 101934 (2021)