Harmonization-enriched domain adaptation with light fine-tuning for multiple sclerosis lesion segmentation

Jinwei Zhang[a], Lianrui Zuo[a], Blake E. Dewey[b], Samuel W. Remedios[c,d], Savannah P. Hays[a], Dzung L. Pham[e], Jerry L. Prince[a,c], and Aaron Carass[a]

[a]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218
[b]Department of Neurology, Johns Hopkins University, Baltimore, MD 21287
[c]Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218
[d]Department of Radiology and Imaging Sciences, National Institutes of Health, Bethesda, MD 20892, USA
[e]Center for Neuroscience and Regenerative Medicine, Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20817

ABSTRACT

Deep learning algorithms using magnetic resonance (MR) images have demonstrated state-of-the-art performance in the automated segmentation of multiple sclerosis (MS) lesions. Despite their success, these algorithms may fail to generalize across sites or scanners, leading to domain generalization errors. Few-shot or one-shot domain adaptation is an option to reduce the generalization error using limited labeled data from the target domain. However, this approach may not yield satisfactory performance due to the limited data available for adaptation. In this paper, we aim to address this issue by integrating one-shot adaptation data with harmonized training data that includes labels. Our method synthesizes new training data with a contrast similar to that of the test domain, through a process referred to as "contrast harmonization" in MRI. Our experiments show that combining one-shot adaptation data with harmonized training data outperformed the use of either one of the data sources alone. Domain adaptation using only harmonized training data achieved comparable or even better performance compared to one-shot adaptation. In addition, all adaptations only required light fine-tuning of 2 to 5 epochs for convergence.

Keywords: Multiple Sclerosis, Lesion Segmentation, Domain Adaptation, Synthesis-based Harmonization

1. INTRODUCTION

Multiple sclerosis (MS) is a central nervous system disorder characterized by inflammatory demyelination and axonal and neuronal degeneration.[1] T2-weighted (T2w) magnetic resonance imaging (MRI) using the fluid-attenuated inversion recovery (FLAIR) pulse sequence is routinely used for clinical diagnosis of MS lesions because it provides high lesion-to-brain contrast while simultaneously suppressing hyperintense cerebrospinal fluid (CSF) signals, which can cause partial-volume artifacts in T2w images.[2] Extensive manual editing is required for accurate delineation of MS lesions, though the task can be quite subjective. Therefore, automatic detection and segmentation of MS lesions is desired for better efficiency and reproducibility.

State-of-the-art methods[3–14] employ deep learning (DL) to automate MS lesion segmentation using multi-contrast MRI scans, including FLAIR. However, these algorithms frequently face challenges in achieving consistent performance across different MRI scanners and imaging protocols. This has led to increased research interest in domain adaptation techniques, such as one-shot domain adaptation,[5] spatially adaptive sub-networks,[6] domain-invariant latent feature learning,[7–9] contrast-adaptive generative modeling,[10] and domain randomness via synthesis,[11] to name just a few.

Further author information: (Send correspondence to Jinwei Zhang)
Jinwei Zhang: E-mail: jwzhang@jhu.edu

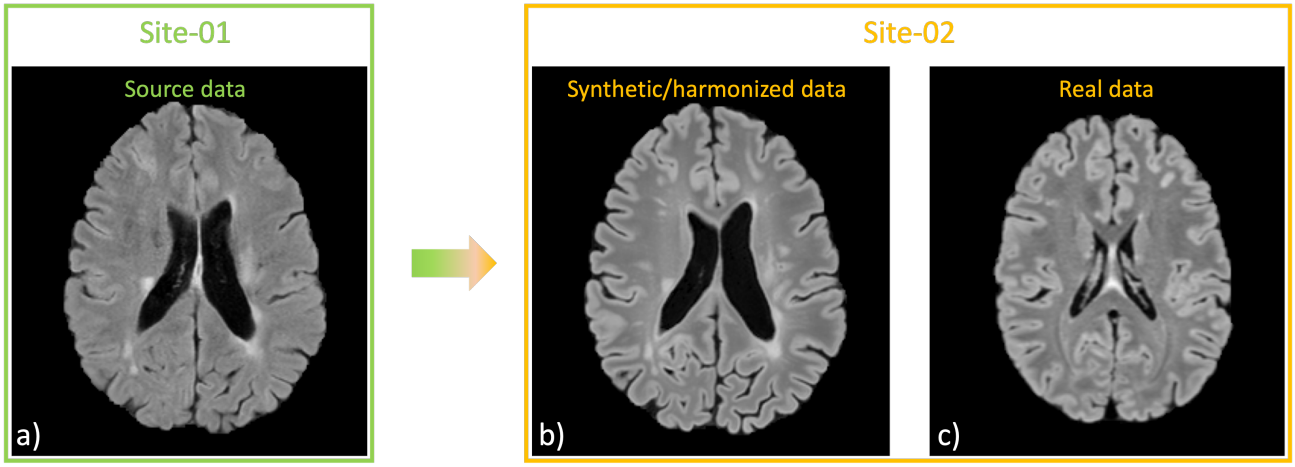

Figure 1. **Synthesis-based harmonization of one FLAIR axial slice.** (a) Source FLAIR slice from Site-01 (ISBI public training set). (b) Harmonized FLAIR slice from Site-01 matching the contrast of Site-02 (in-house test set) using HACA3.[22] (c) Real FLAIR slice from Site-02.

An alternative to domain adaptation is image harmonization, which reduces inter-site variation to aid downstream comparisons and analysis of images across imaging sites, scanners, and over time.[15–17] Synthesis-based multi-contrast MRI harmonization has shown remarkable progress in recent years.[18–21] In this paper, we use HACA3,[22] a new synthesis-based multi-site MRI harmonization approach, to enhance one-shot and even "zero-shot" domain adaptation performance for MS lesion segmentation.

2. METHOD

Dataset. We use the training data from the ISBI longitudinal dataset consisting of five people with MS (PwMS)[23] comprising 42 delineations (2 delineations for each of the 21 imaging sessions) and an in-house dataset including ten PwMS comprising 10 delineations (1 delineation for each of the 10 imaging sessions). We first pre-trained a segmentation network using four of the five training subjects from the ISBI dataset, with the remaining subject being used for validation. We then applied the pre-trained segmentation network to our in-house dataset, which comes from a different domain than the training and validation data.

Network. We implemented a modified 3D UNet following the design choices of nnUNet.[24] For the convolutional building block in the UNet, we chose the "Conv+InstanceNorm+ReLU" configuration with 2 blocks for each level of the encoding or decoding path of the UNet and 4 downsampling/upsampling operations. The numbers of channels in each convolutional block at all levels along the encoding path were 32, 64, 128, 256, and 512. Three dimensional patches were cropped from skull-stripped[25] and white matter intensity normalized[26] T1-weighted (T1w) and FLAIR images, and these patches were concatenated along the channel dimension to be used as input. We generate the binary prediction of segmentation by thresholding the sigmoid of the output of UNet's last convolutional layer.

Training. A batch size of two was used and the 3D patch size was set to $112 \times 112 \times 112$ to fully utilize GPU memory during backpropagation. Heavy augmentations were employed on the fly, including random cropping, axis permutation, intensity shifts, as well as affine and elastic deformations. The loss function was the mean of the Dice similarity coefficient (DSC) and binary cross-entropy. The Adam optimizer was used with an initial learning rate of 10^{-4} for 100 epochs, where each epoch involved the application of eight random augmentations to every training subject.

Harmonization-based domain adaptation. For domain adaptation after training, all T1w and FLAIR images from the ISBI training dataset were transformed to match the contrast of our in-house test dataset using the synthesis-based multi-site harmonization of HACA3.[22] HACA3 was trained on diverse MR datasets acquired from 21 sites, which included white matter lesion multi-contrast images with varying field strengths, scanner platforms, and acquisition protocols. An example of such image harmonization is shown in Fig. 1, where the

FLAIR "source data" from "Site-01" in the ISBI training set was harmonized to match the contrast of "Site-02" of our in-house test set. Consistent gray and white matter contrast was observed between "synthetic/harmonized data" and "real data" for "Site-02".

After harmonization, three domain adaptation strategies were evaluated:

One-Shot Strategy Fine-tune (FT) the pre-trained network with only one of the ten subjects in the test domain and evaluate on the remaining nine test subjects.

Zero-Shot Strategy FT with only the harmonized ISBI data and evaluate on the ten test subjects.

Harmonization-enriched One-Shot Strategy FT with a combination of all harmonized ISBI training data and one of the test domain subjects, and evaluate on the remaining nine test subjects.

For the One-Shot and Harmonization-enriched One-Shot strategies, ten-fold cross-validation (CV) was employed to evaluate on all ten test subjects from the in-house data, where each of the ten subjects in the test domain was included in the training set for one fold. All FTs were conducted for 20 epochs, and the models were tested after every epoch. For comparison, two-fold CV FT with 4/1/5 training/validation/test subjects per fold was also employed for 100 epochs. This CV served as a "normal" performance estimation, assuming enough labeled data in new domains was provided for network adaptation.

Evaluation metrics. The DSC, positive predictive value (PPV), true positive rate (TPR), volumetric F1 score (V-F1), lesion-wise true positive rate (LTPR), lesion-wise false positive rate (LFPR), lesion-wise F1 score (L-F1), and Pearson's correlation coefficient of the lesion volumes between the ground truth and the prediction (VC) were utilized as the segmentation performance evaluation metrics in our experiment.

3. RESULTS

Quantitative comparison. Figure 2 shows DSC, PPV, TPR, V-F1, LTPR, LFPR, L-F1, and VC scores of the three domain adaptation strategies after each FT epoch. First, the performance of all three strategies converged within just 2 to 5 epochs, exhibiting noticeable improvement over the pre-trained results (dashed purple lines) in terms of DSC, TPR, V-F1, LTPR, and L-F1. However, noticeable degradations in PPV and LFPR were observed in both normal CV (dashed red line) and the three adaptation strategies (solid colored lines). In contrast, a degradation in VC was observed in normal CV and One-Shot strategy (solid orange line), but was not observed for the Zero-Shot (solid blue line) and Harmonization-enriched One-Shot (solid green line) strategies. Second, the Harmonization-enriched One-Shot strategy consistently outperformed the other two strategies in terms of DSC, V-F1, and L-F1, and performed similarly well with the Zero-Shot strategy for LFPR and VC. Notably, the Harmonization-enriched One-Shot strategy achieved a DSC score of above 0.6 after convergence, approaching inter-rater consistency.[27] Third, Zero-Shot outperformed One-Shot strategy in terms of DSC and VC, while these two strategies performed similarly for L-F1.

Qualitative comparison. Figure 3 shows segmentation predictions with the corresponding ground-truth label and axial FLAIR slice (left column). The pre-trained prediction (Fig. 3a) exhibited false negative predictions throughout the entire image, which was addressed by the target site CV (Fig. 3b). False positive lesions (indicated by red arrows) were observed in the One-Shot strategy after 2 (Fig. 3c-1) and 20 (Fig. 3c-2) fine-tuning epochs, which were not observed in the Zero-Shot strategy (Fig. 3d) or the Harmonization-enriched One-Shot strategy (Fig. 3e) after 2 epochs (Fig. 3d-1 and Fig. 3e-1, respectively) or 20-epochs (Fig. 3d-2 and Fig. 3e-2) of fine-tuning. False negative lesions (indicated by yellow arrows) missed by the Zero-Shot strategy (Fig. 3d) were still captured by the One-Shot and Harmonization-enriched One-Shot strategies (Fig. 3c and Fig. 3e, respectively). No significant differences were observed between the 2 epoch (Fig. 3d-1 or Fig. 3e-1) and 20 epoch (Fig. 3d-2 or Fig. 3e-2) fine-tuning results of the Zero-Shot or Harmonization-enriched One-Shot strategies.

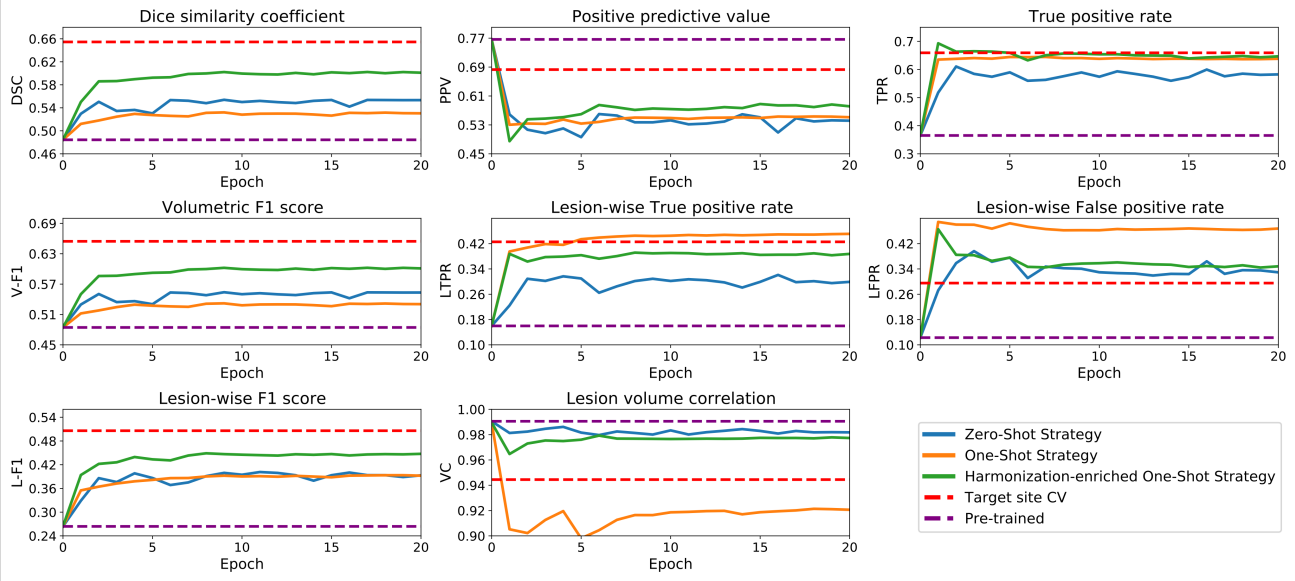

Figure 2. **Quantitative metrics assessed after each epoch of domain adaptation.** The performance of all three strategies converges after only two to five epochs, with noticeable improvement from pre-trained results (dashed purple lines) in DSC, TPR, V-F1, LTPR, and L-F1. Harmonization-enriched One-Shot adaptation data (solid green lines) consistently outperformed the use of either the One-Shot (solid orange lines) or Zero-Shot (solid blue lines) strategies in terms of DSC, V-F1, and L-F1. Notably, the Harmonization-enriched One-Shot strategy achieved a DSC score of above 0.6 after convergence, which is close to inter-rater consistency.[27]

4. DISCUSSION AND CONCLUSION

We demonstrate the feasibility of leveraging synthesis-based MRI harmonization to enhance domain adaptation performance in MS lesion segmentation. Our experiments demonstrate that our Zero-Shot domain adaptation, utilizing solely public data synthesized to the target contrast, yields comparable or superior performance than a One-Shot strategy on the target domain. More notably, the combination of One-Shot and Zero-Shot adaptation, which we coin as Harmonization-enriched One-Shot domain adaptation, achieved DSC results approaching inter-rater performance. Additionally, only light fine-tuning of between 2 and 5 epochs was enough for an adequate adaptation of the pre-trained network.

In Figure 2, when applying domain adaptation to the pre-trained networks, noticeable degradations in PPV and LFPR were observed in both normal CV and the three adaptation strategies, implying that the fine-tuned models tended to segment more lesions at the expense of more false positive predictions. Such over-segmentation after fine-tuning was also observed qualitatively in Figure 3, where more lesions were captured after three adaptation strategies, with both true and false positive predictions of new lesions.

Domain adaptation through network fine-tuning has been successfully applied to different imaging problems in MRI, including under-sampled k-space reconstruction,[28] biophysical inversion[29] with uncertainty quantification,[30] and contrast translation.[31] For lesion segmentation, it is possible to improve the generalization ability of pre-trained models using only synthetic data, as demonstrated in this work. However, fine-tuning is still required for improved generalization. Future work should include synthesizing input images of various target contrasts during training to eliminate the need for additional adaption steps.

ACKNOWLEDGMENTS

This material is partially supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1746891 (Remedios). This work also received support from National Multiple Sclerosis Society RG-1907-34570 (Pham), CDMRP W81XWH2010912 (Prince), and the Department of Defense in the Center for Neuroscience and Regenerative Medicine. The opinions and assertions expressed herein are those of the authors

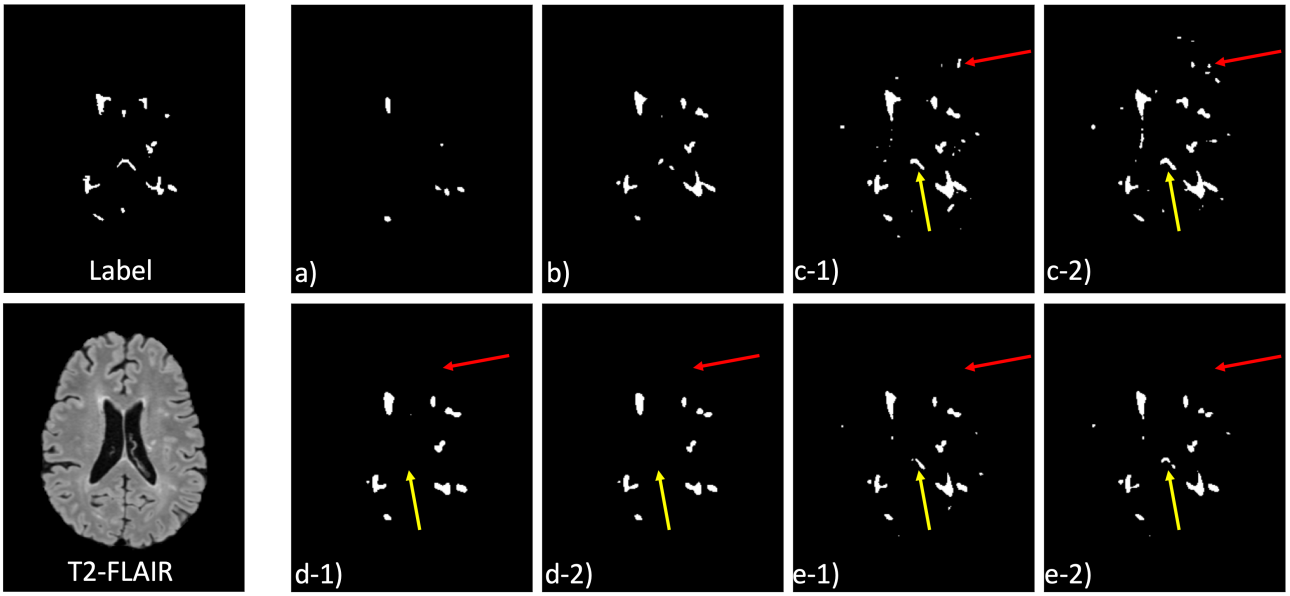

Figure 3. **Segmentation predictions with ground-truth label and axial FLAIR slice.** (a) Pre-trained prediction. (b) Target site CV. (c) One-shot adaptation after 2 (c-1) and 20 epochs (c-2). (d) Our Zero-Shot adaptation after 2 (d-1) and 20 epochs (d-2). (e) Harmonization-enriched One-Shot adaptation after 2 (e-1) and 20 epochs (e-2). False positive lesions (red arrows) in (c) were avoided in the Zero-Shot (d) and Harmonization-enriched One-Shot (e) strategies. False negative lesions are denoted with the yellow arrows.

and do not reflect the official policy or position of the Uniformed Services University of the Health Sciences or the Department of Defense.

REFERENCES

[1] Haider, L., Zrzavy, T., Hametner, S., Höftberger, R., Bagnato, F., Grabner, G., Trattnig, S., Pfeifenbring, S., Brück, W., and Lassmann, H., "The topograpy of demyelination and neurodegeneration in the multiple sclerosis brain," *Brain* **139**(3), 807–815 (2016).

[2] Rovaris, M., Filippi, M., Minicucci, L., Iannucci, G., Santuccio, G., Possa, F., and Comi, G., "Cortical/subcortical disease burden and cognitive impairment in patients with multiple sclerosis," *Am. J. of Neuroradiology* **21**(2), 402–408 (2000).

[3] Zhang, H., Valcarcel, A. M., Bakshi, R., Chu, R., Bagnato, F., Shinohara, R. T., Hett, K., and Oguz, I., "Multiple sclerosis lesion segmentation with Tiramisu and 2.5 D stacked slices," in [*Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*], 338–346, Springer (2019).

[4] Zhang, H., Zhang, J., Li, C., Sweeney, E. M., Spincemaille, P., Nguyen, T. D., Gauthier, S. A., Wang, Y., and Marcille, M., "ALL-Net: Anatomical information lesion-wise loss function integrated into neural network for multiple sclerosis lesion segmentation," *NeuroImage: Clinical* **32**, 102854 (2021).

[5] Valverde, S., Salem, M., Cabezas, M., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Salvi, J., Oliver, A., and Lladó, X., "One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks," *NeuroImage: Clinical* **21**, 101638 (2019).

[6] Kamraoui, R. A., Ta, V.-T., Tourdias, T., Mansencal, B., Manjon, J. V., and Coupé, P., "DeepLesionBrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation," *Medical Image Analysis* **76**, 102312 (2022).

[7] Ackaouy, A., Courty, N., Vallée, E., Commowick, O., Barillot, C., and Galassi, F., "Unsupervised domain adaptation with optimal transport in multi-site segmentation of multiple sclerosis lesions from MRI data," *Front. Comput. Neurosci.* **14**, 19 (2020).

[8] Zhao, X., Sicilia, A., Minhas, D. S., O'Connor, E. E., Aizenstein, H. J., Klunk, W. E., Tudorascu, D. L., and Hwang, S. J., "Robust white matter hyperintensity segmentation on unseen domain," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 1047–1051, IEEE (2021).

[9] Aslani, S., Murino, V., Dayan, M., Tam, R., Sona, D., and Hamarneh, G., "Scanner invariant multiple sclerosis lesion segmentation from MRI," in [*2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*], 781–785, IEEE (2020).

[10] Cerri, S., Puonti, O., Meier, D. S., Wuerfel, J., Mühlau, M., Siebner, H. R., and Van Leemput, K., "A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis," *NeuroImage* **225**, 117471 (2021).

[11] Billot, B., Cerri, S., Van Leemput, K., Dalca, A. V., and Iglesias, J. E., "Joint segmentation of multiple sclerosis lesions and brain anatomy in MRI scans of any contrast and resolution with CNNs," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 1971–1974, IEEE (2021).

[12] Zhang, H., Zhang, J., Zhang, Q., Kim, J., Zhang, S., Gauthier, S. A., Spincemaille, P., Nguyen, T. D., Sabuncu, M., and Wang, Y., "Rsanet: Recurrent slice-wise attention network for multiple sclerosis lesion segmentation," in [*Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*], 411–419, Springer (2019).

[13] Zhang, H., Zhang, J., Wang, R., Zhang, Q., Spincemaille, P., Nguyen, T. D., and Wang, Y., "Efficient folded attention for medical image reconstruction and segmentation," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **35**(12), 10868–10876 (2021).

[14] Zhang, H., Zhang, J., Wang, R., Zhang, Q., Gauthier, S. A., Spincemaille, P., Nguyen, T. D., and Wang, Y., "Geometric loss for deep multiple sclerosis lesion segmentation," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 24–28, IEEE (2021).

[15] Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., et al., "Harmonization of multi-site diffusion tensor imaging data," *NeuroImage* **161**, 149–170 (2017).

[16] 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., et al., "Harmonization of cortical thickness measurements across scanners and sites," *NeuroImage* **167**, 104–120 (2018).

[17] Hays, S. P., Zuo, L., Carass, A., and Prince, J. L., "Evaluating the impact of MR image contrast on whole brain segmentation," in [*Proceedings of SPIE Medical Imaging (SPIE-MI 2022), San Diego, CA, February 20 – 24, 2022*], **12032**, 122–126 (2022).

[18] Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., et al., "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes," *Mag. Reson. Im.* **64**, 160–170 (2019).

[19] Dewey, B. E., Zuo, L., Carass, A., He, Y., Liu, Y., Mowry, E. M., Newsome, S., Oh, J., Calabresi, P. A., and Prince, J. L., "A disentangled latent space for cross-site MRI harmonization," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 720–729, Springer (2020).

[20] Zuo, L., Dewey, B. E., Liu, Y., He, Y., Newsome, S. D., Mowry, E. M., Resnick, S. M., Prince, J. L., and Carass, A., "Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory," *NeuroImage* **243**, 118569 (2021).

[21] Zuo, L., Liu, Y., Xue, Y., Han, S., Bilgel, M., Resnick, S. M., Prince, J. L., and Carass, A., "Disentangling a single MR modality," in [*MICCAI Workshop on Data Augmentation, Labelling, and Imperfections*], **13567**, 54–63, Springer (2022).

[22] Zuo, L., Liu, Y., Xue, Y., Dewey, B. E., Bilgel, M., Mowry, E. M., Newsome, S. D., Calabresi, P. A., Resnick, S. M., Prince, J. L., and Carass, A., "HACA3: A unified approach for multi-site MR image harmonization," *Computerized Medical Imaging and Graphics* **102285**, *IN PRESS* (2023).

[23] Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C. H., et al., "Longitudinal multiple sclerosis lesion segmentation: Resource and challenge," *NeuroImage* **148**, 77–102 (2017).

[24] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods* **18**(2), 203–211 (2021).

[25] Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., et al., "Automated brain extraction of multisequence MRI using artificial neural networks," *Human Brain Mapping* **40**(17), 4952–4964 (2019).

[26] Reinhold, J. C., Dewey, B. E., Carass, A., and Prince, J. L., "Evaluating the impact of intensity normalization on MR image synthesis," in [*Medical Imaging 2019: Image Processing*], **10949**, 109493H, International Society for Optics and Photonics (2019).

[27] Carass, A., Roy, S., Gherman, A., Reinhold, J. C., Jesson, A., Arbel, T., Maier, O., Handels, H., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Pham, D. L., Crainiceanu, C. M., Calabresi, P. A., Prince, J. L., Gray Roncal, W. R., Shinohara, R. T., and Oguz, I., "Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis," *Nature Scientific Reports* **10**, 8242 (2020).

[28] Zhang, J., Liu, Z., Zhang, S., Zhang, H., Spincemaille, P., Nguyen, T. D., Sabuncu, M. R., and Wang, Y., "Fidelity imposed network edit (FINE) for solving ill-posed image reconstruction," *NeuroImage* **211**, 116579 (2020).

[29] Zhang, J., Zhang, H., Spincemaille, P., Nguyen, T., Sabuncu, M. R., and Wang, Y., "Hybrid optimization between iterative and network fine-tuning reconstructions for fast quantitative susceptibility mapping," in [*Medical Imaging with Deep Learning*], 870–880, PMLR (2021).

[30] Zhang, J., Zhang, H., Sabuncu, M., Spincemaille, P., Nguyen, T., and Wang, Y., "Bayesian learning of probabilistic dipole inversion for quantitative susceptibility mapping," in [*Medical Imaging with Deep Learning*], 892–902, PMLR (2020).

[31] He, Y., Carass, A., Zuo, L., Dewey, B. E., and Prince, J. L., "Autoencoder based self-supervised test-time adaptation for medical image analysis," *Medical Image Analysis* **72**, 102136 (2021).