

Contrastive Anatomy-Contrast Disentanglement: A Domain-General MRI Harmonization Method

Daniel Scholz^{1,2,3}, Ayhan Can Erdur^{3,4}, Robbie Holland⁶, Viktoria Ehm^{2,5},
Jan C. Peeken^{4,7,8}, Benedikt Wiestler^{1,2,*}, and Daniel Rueckert^{2,3,*}

¹ Chair for AI for Image-Guided Diagnosis and Therapy, Technical University of Munich (TUM) and TUM University Hospital, Munich, Germany

² Munich Center for Machine Learning (MCML), Munich, Germany

³ Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM) and TUM University Hospital, Munich, Germany

⁴ Department of Radiation Oncology, TUM University Hospital, Munich, Germany

⁵ Chair for Computer Vision and Artificial Intelligence, Technical University of Munich (TUM), Munich, Germany

⁶ Stanford Center for Artificial Intelligence in Medicine and Imaging, Stanford University, Stanford, CA, USA

⁷ Deutsches Konsortium für Translationale Krebsforschung (DKTK), Partner Site Munich, Munich, Germany

⁸ Institute of Radiation Medicine (IRM), Department of Radiation Sciences (DRS), Helmholtz Center Munich, Munich, Germany

*contributed equally as senior authors

daniel.scholz@mri.tum.de

Abstract. Magnetic resonance imaging (MRI) is an invaluable tool for clinical and research applications. Yet, variations in scanners and acquisition parameters cause inconsistencies in image contrast, hindering data comparability and reproducibility across datasets and clinical studies. Existing scanner harmonization methods, designed to address this challenge face limitations, such as requiring traveling subjects or struggling to generalize to unseen domains. We propose a novel approach using a conditioned diffusion autoencoder with a contrastive loss and domain-agnostic contrast augmentation to harmonize MR images across scanners while preserving subject-specific anatomy. Our method enables brain MRI synthesis from a single reference image. It outperforms baseline techniques, achieving a +7% PSNR improvement on a traveling subjects dataset and +18% improvement on age regression in unseen scanners. Our model provides robust, effective harmonization of brain MRIs to target scanners without requiring fine-tuning. This advancement promises to enhance comparability, reproducibility, and generalizability in multi-site and longitudinal clinical studies, ultimately contributing to improved healthcare outcomes.¹

Keywords: Scanner Harmonization · Domain Generalization · Disentanglement · Contrastive Learning.

¹ The code is publicly available at: <https://github.com/daniel-scholz/cacd>.

1 Introduction

Magnetic resonance imaging (MRI) is a versatile technique for brain studies, offering multiple tissue contrasts in a single session and enabling non-invasive insights into the brain’s structure and function. However, its flexibility leads to a lack of standardization across imaging studies. Variations in pulse sequences, acquisition parameters, and scanner hardware can cause undesired contrast variations, particularly in multi-site and longitudinal studies [15,9]. These variations create a domain shift problem, introducing underlying bias into the models and limiting the generalizability of quantitative analyses. Addressing these challenges is crucial for advancing multi-center MRI research and improving diagnostic reliability across healthcare settings.

By using traveling subjects, i.e., scans of the same subject in different scanners serving as ground truth, harmonization can be achieved by learning direct mappings between scanners [4,24]. However, traveling subjects require significant effort and hence occur only seldom in clinical reality, limiting the application of such methods to a narrow context. Many methods alleviate the necessity for traveling subjects through inter-domain cycle consistency [18,26]. These methods can usually translate between only two scanners by their pair-wise design, thus lacking the generalizability to harmonize multiple sources into a common space. Furthermore, the absence of anatomical constraints can lead to unrealistic transformations or even hallucinations.

Recent methods [28,2] tackle scanner harmonization by disentangling contrast from anatomy, i.e., keeping subject-specific anatomical aspects consistent while learning contrast changes that originate from scanners to generalize to more than two domains. To this end, Zuo et al. [28] rely on paired MRI sequences, such as T1w, T2w, and FLAIR, of the same subject, which are more common in clinical practice, to learn anatomical consistency. Models that can achieve disentanglement from only a single sequence also exist [2]. However, due to the lack of domain generalization, these models are limited to synthesizing images similar to their training domain. In particular, introducing new domains unseen during training often fails but is a clinical reality.

In our work, we develop an anatomy-consistent approach that harmonizes between an arbitrary number of unseen scanners outside the training domain. Additionally, the proposed method overcomes the reliance on paired subjects or multiple MRI sequences per subject. We summarize our contributions as follows:

1. We present a domain-general scanner harmonization algorithm capable of **harmonizing arbitrary unseen scanners** in brain MRI by controlling contrast and maintaining anatomy.
2. To this end, we introduce **a novel anatomy-contrast-disentanglement module**, including an elegant loss formulation with contrastive losses and a domain-agnostic contrast augmentation suite.
3. We **advance MRI scanner harmonization in unseen scanners** in terms of traveling subjects harmonization and age regression.

2 Related Works

Numerous methods have been proposed for MRI harmonization, which are discussed in detail in these two recent reviews [7,10], so that we focus on discussing approaches most relevant to our work.

Learning-based Methods Many existing deep learning models for scanner harmonization require traveling subjects [4,24]. Their reliance on full supervision, i.e., datasets with images of the same patient from multiple sites, limits their applicability drastically. In contrast, the unsupervised CycleGAN method [27] uses a cycle consistency loss to learn mappings between two domains without paired examples. While adaptations of CycleGAN have been explored for MRI harmonization [18,26], this approach can naturally only handle two sites, and anatomical consistency during harmonization is not ensured.

Latent-Space Disentanglement Representation disentanglement has been a significant focus of the computer vision community, e.g. for style transfer [6,14] or medical image segmentation [8,11]. Some works accomplish scanner harmonization with disentangled latent spaces to maintain anatomical consistency while learning intensity-level translation [2,28]. HACA3 [28] extract the anatomy embedding through shape constraints from paired MRI sequences while the contrast is learned as the remaining information. To overcome the need for paired MRI sequences, [2] introduce Gamma augmentations to create different contrasts while [25,29] utilize energy- or mutual-information-based models. These approaches, however, are limited to the contrast differences covered by Gamma augmentations, which do not account for the entire variation in real-world data, and relies on patches for training. These limitations highlight the need for a framework capable of processing full volumes while leveraging a more generalizable image synthesis and latent-space disentanglement.

3 Method

3.1 Diffusion Autoencoder

The foundation of our work is the diffusion autoencoder (DiffAE) [19]. It consists of a conditional DDIM [21] and a separate semantic encoder $E(x)$ that projects the input image x to a latent feature vector z_{sem} , which is then used to condition the DDIM. The diffusion autoencoder architecture allows control of the image editing process by altering the semantic feature vector to contain the semantic information desired in the edited image. However, the feature space needs to be disentangled to make specific changes that leave other imaging features constant.

3.2 Disentanglement Module

In this work, we aim to disentangle anatomy and contrast to enable altering intensity while maintaining the anatomy in the image (Figure 1). To this end,

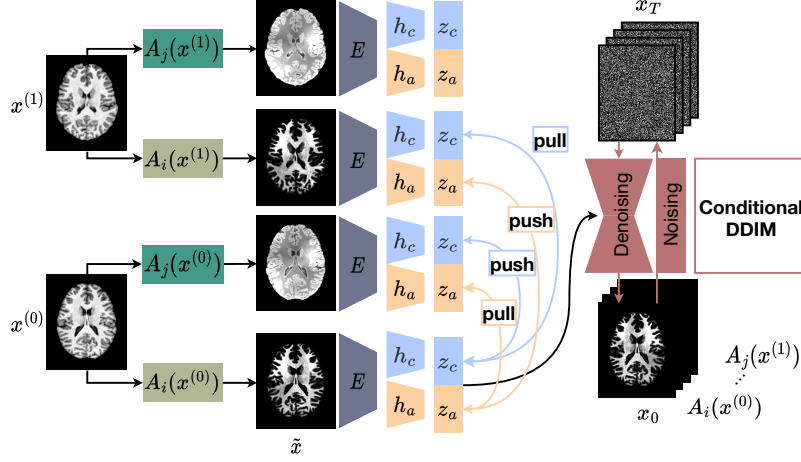


Fig. 1: **Our novel anatomy-content-disentanglement module.** Two images from different subjects and arbitrary (possibly even the same) scanners $x^{(0)}$ and $x^{(1)}$ are modified through two augmentations $A_j(x)$ and $A_i(x)$, yielding a view \tilde{x} each. An anatomy-contrast disentanglement is learned through deliberate pushing and pulling (see contrastive loss, Eq. 1) of the anatomy $z_a = h_a(E(\tilde{x}))$ and contrast $z_c = h_c(E(\tilde{x}))$ embeddings (Section 3.2). Note that in this example, push and pull are shown only for a single view - in practice, all views are considered. The resulting feature vectors z_a and z_c are used to condition the DDIM to synthesize \tilde{x} by denoising the noise map x_T to the input view denoted as $x_0 \hat{=} \tilde{x}$.

we split the feature vector $z_{\text{sem}} \in \mathbb{R}^d$ into two feature vectors $z_a \in \mathbb{R}^{d_a}$ and $z_c \in \mathbb{R}^{d_c}$ for anatomy and contrast, respectively. To achieve this, we add two separate network heads such that $h_a(E(x)) = z_a$ and $h_c(E(x)) = z_c$.

To disentangle the learned anatomical and contrast vectors, we introduce a novel loss formulation based on a contrastive loss [3,13]. First, we sample a set of N_{aug} contrast augmentations $\{A_j(x)\}_{j \in \{0, \dots, N_{\text{aug}}\}}$, which model a variety of intensities, but, crucially, preserve the anatomy (Section 3.3). We then apply each augmentation to each image $x^{(i)}$ in a batch of images such that we end up with augmented versions of the input images, termed views, $\tilde{x}^{(i)}$ that have greatly varying contrasts but maintain the original anatomy of $x^{(i)}$. The total number of views generated is $B \cdot N_{\text{aug}}$, where B is the original number of input images. Finally, each augmented view $A_j(x^{(i)})$ is projected into an anatomy $z_a^{(i)}$ and contrast vector $z_c^{(i)}$ using the encoder and respective network heads. For simplicity, we denote $z_a^{(i)} = z_a$ and $z_c^{(i)} = z_c$.

Next, we utilize the supervised contrastive loss [13], an established modification of SimCLR [3] (Eq. 1). This formulation includes the temperature-scaled cosine similarity sim , with temperature τ . This loss is minimized by requiring the feature vectors of positive pairs in $\mathcal{P}^{(i)}$ of a view $z^{(i)} \in \mathbb{R}^d$ to be similar

(*pull*) and negative pairs $\mathcal{N}^{(i)}$ to be dissimilar (*push*):

$$\ell(z^{(i)}, \mathcal{P}^{(i)}, \mathcal{N}^{(i)}) = -\log \frac{\sum_{k \in \mathcal{P}^{(i)}} \exp(\text{sim}(z^{(i)}, z^{(k)})/\tau)}{\sum_{k \in (\mathcal{P}^{(i)} \cup \mathcal{N}^{(i)})} \exp(\text{sim}(z^{(i)}, z^{(k)})/\tau)} \quad (1)$$

We enforce anatomy-contrast disentanglement by carefully defining positive and negative pairs. To this end, we choose the positive pairs to be all pairs of views originating from the *same* subject but were transformed with *different* augmentations. The index set of these positive pair candidates for the i -th view in a batch is denoted as $\mathcal{P}_a^{(i)}$. The negative pairs are built from all views originating from a *different* subject that were transformed with the *same* augmentations. Correspondingly, the index set of negative view candidates for the i -th image in a batch is denoted as $\mathcal{N}_a^{(i)}$. This design forces the network to find similar representations for the intensity-augmented views, leading to only anatomy information encoded in z_a and, hence, anatomy-contrast disentanglement. In preliminary experiments, we found the low-dimensional vector z_a guiding the reconstruction of global anatomical structures, while fine-grained details are preserved by the initial noise map x_T . This noise map is determined via DDIM inversion to encode high-frequency information from the input [19].

To force the model to extract only contrast information into z_c the definition of positive and negative pairs is “inverted”. All positive pairs, denoted as $\mathcal{P}_c^{(i)}$, are given by the views originating from the same augmentation but different subjects, i.e., the shared feature between these views is the contrast. On the contrary, we consider the views of the same subject that were transformed with different augmentations to be negative, denoted as $\mathcal{N}_c^{(i)}$. Consequently, the differentiating feature between the negative views is also the contrast. Overall, the encoder $E(x)$ and the heads h_a and h_c need to learn disentangled representations to minimize the overall loss per augmented view at index i :

$$\mathcal{L}_{\text{total}}^{(i)} = \mathcal{L}_{\text{DDIM}}^{(i)} + \lambda_c \ell(z_c^{(i)}, \mathcal{P}_c^{(i)}, \mathcal{N}_c^{(i)}) + \lambda_a \ell(z_a^{(i)}, \mathcal{P}_a^{(i)}, \mathcal{N}_a^{(i)}) \quad (2)$$

with $\ell(z^{(i)})$ the loss defined in Eq. 1. $\mathcal{L}_{\text{DDIM}}^{(i)}$ is the standard MSE loss to train the DDIM model [21]. In practice, we set $\lambda_c = \lambda_a = 0.5$.

During inference, we first estimate a target contrast vector \bar{z}_c by projecting and averaging a set of target images to a contrast vector. The target anatomy vectors are discarded. The source image is projected to the latent space where its contrast vector z_c is swapped with \bar{z}_c . The new feature vectors \bar{z}_c, z_a , i.e., the target contrast and source anatomy, serve as a condition to the diffusion model.

3.3 Augmentation Strategy

Our augmentation strategy is designed to cover all kinds of scanner variations while preserving the subject’s anatomy. For our augmentation set, we gather augmentation schemes from different previous works, namely, the global intensity non-linear augmentation (GIN) [17] and the SynthX-framework [12]. We

apply Gamma augmentations with $\gamma = e^u, u \sim \mathcal{U}(0, 0.5)$, and randomly generated biasfield corruptions as in the SynthX-framework. However, we replace the synthetic contrasts generated in SynthX by applying the GIN augmentation [17] to real samples from the training dataset, making our method agnostic to the anatomical region. The augmented image is normalized to the $[-1, 1]$ -range and linearly interpolated with the original input image with a uniformly sampled interpolation factor α . To counteract the blurring that occurs with GIN and deeper network sizes, we linearly upsample the images to obtain sharper images that are down-sampled back to the original resolution after augmentation. Examples are shown in Figure 1. Importantly, the augmented images also serve as targets for the diffusion model, enabling our model to synthesize various intensities.

3.4 Implementation Details

We parametrize the latent space heads $h_{\{c,a\}}$ as linear layers. The latent space dimensions for both z_a and z_c is $d_a = d_c = 256$. We use an Adam optimizer with a learning rate of $1e-4$. We train the model for 2.5 million steps on a single NVIDIA A40 GPU. Our total number of views is 81, derived from $N_{\text{aug}} = 9$ and $B = 9$ subjects. The GIN network [17] consists of two hidden layers with two hidden channels, a kernel size of 3, and bilinear upsampling to 2048 pixels.

4 Results

4.1 Datasets

We use three different datasets for training and evaluation of our models: OASIS3 [16], IXI [1], and OpenNeuro(ON)Harmony [23]. The OASIS3 dataset includes 3231 T1w MR Scans from 1047 subjects from various scanners of multiple vendors and field strengths (for more details, see [16]). Next, the IXI dataset consists of 580 healthy subjects from three sites: Hammersmith Hospital (HH, Philips 3T), Guy’s Hospital (Philips 1.5T), and the Institute of Psychiatry (IOP, GE 1.5T). The ONHarmony dataset (Phase A, Version 1.0.2) consists of 10 healthy subjects who have been scanned at least once in six 3T MR scanners from the major vendors Siemens, Philips, and GE, yielding 80 T1w MRIs. We use 80% of IXI- $\{\text{Guy’s}, \text{HH}\}$, and the full OASIS3 dataset to train our model. The remaining IXI data, i.e., 20% of IXI- $\{\text{Guy’s}, \text{HH}\}$ and the full IXI-IOP, are used to evaluate our model and compare it to other scanner harmonization methods. The ONHarmony dataset is used to evaluate traveling subjects, which poses a particularly hard challenge since the scanners are unseen to all methods.

To standardize MRI data, we affinely register all scans to the MNI152 atlas at $1 \times 1 \times 1\text{mm}^3$ resolution using the niftyreg library. We then skull-strip the scans using HD-BET and apply N4 bias field correction. Lastly, the images are normalized to the $[-1, 1]$ -range in the brainmask. To save computational resources, we randomly sample axial slices from the middle section (middle slice ± 10) of the scan for training and use the axial middle slice for evaluation. Our method, however, is not reliant on patches and can thus be readily applied to 3d volumes when the resources are available.

4.2 Traveling Subjects

To evaluate the efficacy of scanner harmonization we need ground truth data, which requires a subject to be scanned in multiple scanners within a short time period. To this end, we use paired scans from traveling subjects in the ONHarmony dataset [23] and translate the scans from all scanners to the GE scanner present in the dataset, as this is visually the most distinct. We compare our method to the current state-of-the-art unsupervised deep learning scanner harmonization model HACA3 [28] with the provided pre-trained weights as well as the standard DiffAE [19] without disentanglement. HACA3 is trained on images with skull, hence, we performed inference on non-skull-stripped images but calculated all metrics solely within the brain mask to ensure a fair comparison.

Our quantitative results in Figure 2a demonstrate a significant improvement ($p < 0.01$, Wilcoxon signed rank test) over all baselines in all metrics, especially in PSNR. The qualitative results in Figure 2b support these findings visually. Our difference maps’ lower intensity indicates a smaller error than the other methods. Importantly, we also observe anatomically faithful reconstructions and well-adjusted contrasts in the harmonized images regardless of the source scanner or subject due to the large variety of intensities our model has seen during training. HACA3 reconstructs the anatomy faithfully but fails to capture the contrast correctly due to the unseen scanners confronting the model with unseen intensities. The standard DiffAE manages to match the intensity profile but significantly changes the anatomy of the subject due to the entanglement of anatomy and contrast by changing ventricle sizes and gyri and sulci shapes.

4.3 Scanner Classification and Age Regression

Table 1: **Scanner classification and age regression results.** We show the effectiveness of our method in harmonizing images by (i) fooling a scanner classifier (*scanner bias*), (ii) achieving high *target fidelity* by translating images to a new domain, and (iii) improving *age regression* performance on an unseen scanner (IOP). Bold values indicate the best performance in each column.

Method	Scanner Bias			Target Fidelity			Age Regression
	F1 (\downarrow)			Acc (\uparrow)			R^2 (\uparrow)
Target Scanner	Guy’s	HH	IOP	Guy’s	HH	IOP	IOP \rightarrow Guy’s
Unharmonized	0.593	0.568	0.543	0.031	0.222	0.021	0.58
CycleGAN	0.0	0.4	n/a	1.0	0.0	n/a	-6.53
HACA3	0.226	0.688	0.451	0.469	0.0	0.383	0.57
Ours	0.187	0.111	0.167	0.75	0.8	0.681	0.69

To evaluate the effectiveness of various harmonization methods in mitigating domain shift, we train a radiomics-based classifier [22] (pyradiomics 3.1.1) on

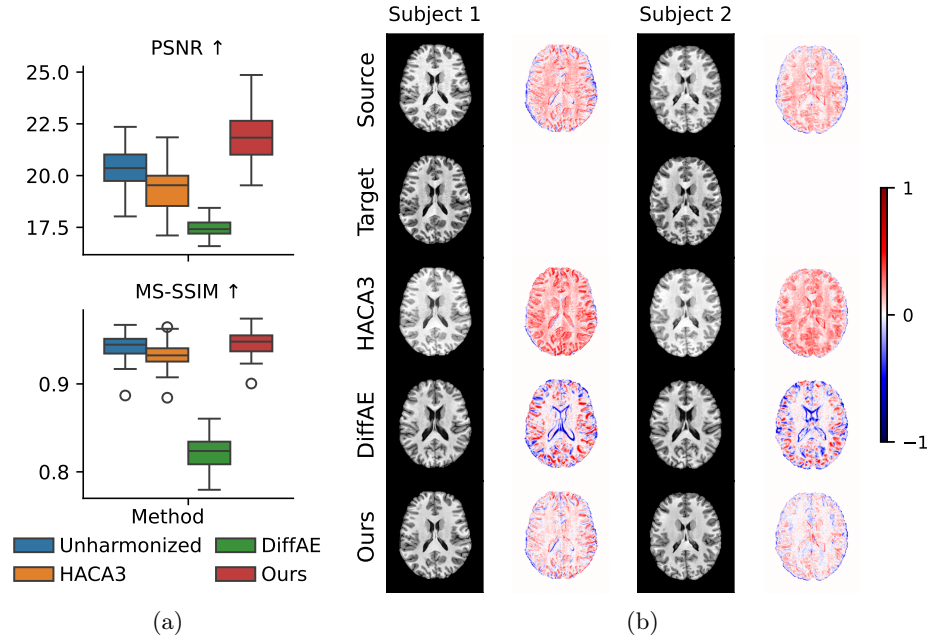


Fig. 2: **(a) Traveling subjects evaluation.** Our method outperforms the other methods in both PSNR and MS-SSIM. **(b) Visual comparison.** The difference maps between the harmonized and target images visualize contrast adjustment while anatomy is maintained (Best viewed magnified on screen).

three subsets of the IXI dataset: Guy’s, HH, and IOP. When images are perfectly harmonized to a specific target scanner, the classifier should classify every sample as stemming from this target scanner. We thus harmonize all images to a specific scanner (column head in Table 1) and evaluate two scenarios: In the *scanner bias* scenario, we compare the classifier’s output to the original scanner labels aiming at minimizing the F1 score (\downarrow), contrary to typical objectives. Conversely, in the *target fidelity* scenario, we compare the classifier’s output to the label of the target scanner itself expecting maximal classification accuracy (\uparrow). We compare our method against CycleGAN [27], trained on Guy’s and HH, and HACA3 [28]. Our model consistently works robustly for all three scanners, even when choosing IOP as the target - a scanner not seen during training - and outperforms HACA3 across all comparisons. CycleGAN performs well for the Guy’s scanner but fails for HH and is not applicable for IOP, highlighting its limitations.

Finally, as a downstream task, we train an age regressor on MRIs from Guy’s and HH, and evaluate it in IOP, either unharmonized or following harmonization to Guy’s domain space (Table 1, "Age Regression"). Our method outperforms all comparisons methods in terms of R^2 (\uparrow) and significantly improves age regression in unseen scanners. Overall, our method shows the best domain generalization to unseen scanners, while not needing to train on target domain data.

5 Discussion and Conclusion

This work presents a novel method for scanner harmonization of brain MRI. Our approach builds upon the diffusion autoencoder [19] and introduces a novel contrastive anatomy-contrast-disentanglement module, including a domain-agnostic intensity augmentation suite. Importantly, our approach does not require paired data, either in the form of multiple scans of the same subject or multiple sequences, underscoring its broad applicability and overcoming a key limitation of earlier approaches [4,24]. Currently, our approach uses 2D slices of the volumes due to the heavy computational load of diffusion models. However, only the synthesis part of our model is challenging to upgrade to 3D, as we do not rely on comparisons between slices, as in [29,2]. A simple remedy to this problem is to use latent or wavelet diffusion models [20,5]. In conclusion, we believe that our work addresses challenges crucial for advancing multi-center MRI research and diagnostic healthcare through generalizable scanner harmonization in the future.

Acknowledgments. This study was supported by the DFG, grant #504320104.

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

References

1. IXI Dataset – Brain Development, <http://brain-development.org/ixi-dataset/>
2. Cackowski, S., Barbier, E.L., Dojat, M., Christen, T.: Imunity: a generalizable vae-gan solution for multicenter mr image harmonization. *Medical Image Analysis* **88**, 102799 (2023)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A Simple Framework for Contrastive Learning of Visual Representations. In: *Proceedings of the 37th International Conference on Machine Learning*. pp. 1597–1607. PMLR (nov 2020)
4. Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., et al.: Deepharmony: A deep learning approach to contrast harmonization across scanner changes. *Magnetic resonance imaging* **64**, 160–170 (2019)
5. Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P.C.: Wdm: 3d wavelet diffusion models for high-resolution medical image synthesis. In: *MICCAI Workshop on Deep Generative Models*. pp. 11–21. Springer (2024)
6. Gatys, L.A., Ecker, A.S., Bethge, M.: Image Style Transfer Using Convolutional Neural Networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2414–2423. IEEE, Las Vegas, NV, USA (jun 2016)
7. Gebre, R.K., Senjem, M.L., Raghavan, S., Schwarz, C.G., Gunter, J.L., et al.: Cross-scanner harmonization methods for structural MRI may need further work: A comparison study. *NeuroImage* **269**, 119912 (apr 2023)
8. Gu, R., Wang, G., Lu, J., Zhang, J., Lei, W., et al.: Cddsa: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation. *Medical Image Analysis* **89**, 102904 (2023)
9. Hedges, E.P., Dimitrov, M., Zahid, U., Brito Vega, B., Si, S., et al.: Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. *NeuroImage* **246**, 118751 (feb 2022)

10. Hu, F., Chen, A.A., Horng, H., Bashyam, V., Davatzikos, C., et al.: Image harmonization: A review of statistical and deep learning methods for removing batch effects and evaluation metrics for effective harmonization. *NeuroImage* **274**, 120125 (jul 2023)
11. Hu, S., Liao, Z., Xia, Y.: Devil is in Channels: Contrastive Single Domain Generalization for Medical Image Segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. pp. 14–23. Springer Nature Switzerland, Cham (2023)
12. Iglesias, J.E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., et al.: Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage* **237**, 118206 (aug 2021)
13. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., et al.: Supervised contrastive learning. *Advances in neural information processing systems* **33**, 18661–18673 (2020)
14. Kotovenko, D., Sanakoyeu, A., Lang, S., Ommer, B.: Content and Style Disentanglement for Artistic Style Transfer. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 4422–4431 (2019)
15. Kushol, R., Parnianpour, P., Wilman, A.H., Kalra, S., Yang, Y.H.: Effects of MRI scanner manufacturers in classification tasks with deep learning models. *Scientific Reports* **13**(1), 16791 (oct 2023)
16. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., et al.: OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *medRxiv : the preprint server for health sciences* pp. 2019–12 (2019)
17. Ouyang, C., Chen, C., Li, S., Li, Z., Qin, C., et al.: Causality-Inspired Single-Source Domain Generalization for Medical Image Segmentation. *IEEE Transactions on Medical Imaging* **42**(4), 1095–1106 (apr 2023)
18. Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Öztürk, Ş., et al.: Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging* (2023)
19. Preechakul, K., Chatthee, N., Wizadwongsa, S., Suwajanakorn, S.: Diffusion Autoencoders: Toward a Meaningful and Decodable Representation. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 10609–10619. IEEE, New Orleans, LA, USA (jun 2022)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 10684–10695 (2022)
21. Song, J., Meng, C., Ermon, S.: Denoising Diffusion Implicit Models (oct 2022)
22. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., et al.: Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* **77**(21), e104–e107 (oct 2017)
23. Warrington, S., Ntata, A., Mougin, O., Campbell, J., Torchi, A., et al.: "ON-Harmony: A resource for development and comparison of multimodal brain 3T MRI harmonisation approaches" (2023)
24. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 35–45. Springer (2022)

25. Wu, M., Zhang, L., Yap, P.T., Zhu, H., Liu, M.: Disentangled latent energy-based style translation: An image-level structural mri harmonization framework. *Neural Networks* **184**, 107039 (2025)
26. Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., et al.: Harmonization of infant cortical thickness using surface-to-surface cycle-consistent adversarial networks. In: *International conference on medical image computing and computer-assisted intervention*. pp. 475–483. Springer (2019)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2223–2232 (2017)
28. Zuo, L., Liu, Y., Xue, Y., Dewey, B.E., Remedios, S.W., et al.: Haca3: A unified approach for multi-site mr image harmonization. *Computerized Medical Imaging and Graphics* **109**, 102285 (2023)
29. Zuo, L., Liu, Y., Xue, Y., Han, S., Bilgel, M., et al.: Disentangling a single mr modality. In: *MICCAI Workshop on Data Augmentation, Labelling, and Imperfections*. pp. 54–63. Springer (2022)