# Unified Cross-Field Multi-Contrast MRI Translation and Harmonization

Yulin Wang[1]*, Ziyi Chen[1]*, and Dinggang Shen[1,2,3](✉)

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
Dinggang.Shen@gmail.com

**Abstract.** Magnetic resonance imaging (MRI) is an indispensable non-invasive modality for neurological diagnosis and research. However, MRI appearance, quality, and acquisition cost vary substantially across magnetic field strengths. These variations hinder the joint analysis of multi-center MRI datasets, and resource-constrained regions often have limited access to high-quality imaging. We present a unified framework for cross-field multi-contrast MRI translation and harmonization across five field strengths and three contrasts. Specifically, a task-conditioned 3D translator synthesizes target images from arbitrary source scans under field- and contrast-specified domains. The framework adopts a gated adapter to selectively integrate global contrast and local anatomical information from same-subject auxiliary contrasts into the primary translation pathway. To address the scarcity of paired multi-field data and improve robustness to heterogeneous acquisitions, we introduce an anatomy-anchored, field-aware stochastic simulation strategy that generates diverse pseudo-paired samples from real target images for model training. On the validation set, preliminary experiments show strong synthesized image quality and structural fidelity, achieving average SSIM, nRMSE, and LPIPS scores of 0.908, 0.230, and 0.105, respectively.

**Keywords:** MRI synthesis · Field-strength translation · Multi-contrast MRI · Low-field MRI · Image harmonization

## 1 Introduction

Magnetic resonance imaging (MRI) is a core tool for neuroscience research and clinical assessment, because it provides non-invasive, radiation-free imaging with rich soft-tissue contrast. However, differences in magnetic field strength introduce substantial heterogeneity in signal-to-noise ratio (SNR), spatial resolution, and artifact characteristics. Low-field systems suffer from reduced SNR and effective resolution [5,1], but are lower-cost, more compact, and increasingly portable. In contrast, high-field MRI provides higher signal efficiency and finer anatomical detail, yet requires costly infrastructure and can be more sensitive to susceptibility and intensity inhomogeneity [1]. These field-dependent differences reduce

image comparability across hospitals and studies, and pose a major barrier to the clinical translation of AI-based MRI models [7,14]. Meanwhile, resource-constrained regions that often rely on low-field MRI systems have limited access to high-quality imaging.

Generative image translation offers a solution by synthesizing the appearance of a specified target field strength and contrast from an observed source scan. Early paired studies reconstructed 7T-like images from 3T MRI using coupled patch representations [2], and subsequent deep models combined spatial and wavelet-domain information for 3T-to-7T synthesis [11]. More recently, adversarial diffusion has been explored for translating paired 64mT acquisitions to 3T MRI [6]. In parallel, image-level harmonization methods such as DeepHarmony and HACA3 have addressed scanner, protocol, and site-induced appearance shifts [7,14]. However, existing cross-field synthesis methods are typically optimized for a single translation direction, limiting their ability to accommodate heterogeneous clinical demands.

A practical solution requires a unified model that supports arbitrary cross-field translation without maintaining a separate network for every paired field setting. However, the obstacle is the scarcity of same-subject acquisitions across multiple field strengths. Unpaired translation methods reduce this dependence on aligned data. CycleGAN constrains adversarial distribution matching through an inverse cycle [13], while CUT preserves input–output correspondence using patch-wise contrastive learning [9]. Nevertheless, such distribution-level objectives remain under-constrained, leading to anatomical drift and inconsistent performance across translation directions. An alternative is to construct pseudo-paired samples for data augmentation, thereby converting unpaired learning into a supervised translation problem. SynthSeg [3] and SynthSR [8] randomize contrast and resolution from anatomical label maps for robust image segmentation and synthesis model training, respectively, highlighting the value of anatomy-anchored domain randomization for supervised model training.

In this study, we propose a unified task-conditioned 3D framework for brain MRI translation and harmonization across five field strengths (0.1T, 1.5T, 3T, 5T, and 7T) and three contrasts (T1w, T2w, T2-FLAIR). The image translator uses target field and contrast conditions to guide synthesis of the specified target domain from any acquired source scans, enabling shared modeling across diverse field translation directions for multiple contrasts. To exploit complementary information from same-subject auxiliary contrasts, we design a gated adapter to selectively incorporate global contrast relevance and local anatomical support from auxiliary contrasts into the primary translation pathway. To reduce dependence on scarce paired multi-field acquisitions, we further introduce an anatomy-anchored, field-aware stochastic simulation strategy to generate diverse pseudo-source/real-target pairs from real target images and corresponding brain-tissue segmentations. Preliminary validation results demonstrate strong cross-field synthesis performance, suggesting the potential of the proposed framework for multi-center MRI field harmonization and for facilitating downstream joint image use and analysis.

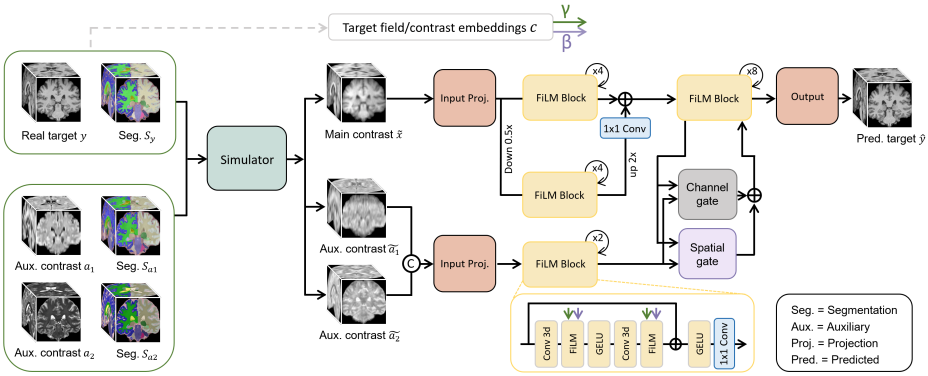


**Fig. 1.** Overview of the unified cross-field translation framework. The model receives a source MRI image and available same-subject auxiliary contrasts, together with a task condition specifying the desired field strength and contrast, and synthesizes the corresponding target-domain image. During training, source images are simulated from real target images to provide paired supervision.

## 2 Methods

### 2.1 Overview

Let $y$ denote a real image of primary contrast $m$ acquired at target field $f_t$, where

$$\mathcal{F} = \{0.1\mathrm{T}, 1.5\mathrm{T}, 3\mathrm{T}, 5\mathrm{T}, 7\mathrm{T}\} \quad \text{and} \quad \mathcal{M} = \{\mathrm{T1w}, \mathrm{T2w}, \mathrm{T2\text{-}FLAIR}\}.$$

We formulate the task as same-contrast cross-field translation. Given a source image $\tilde{x}$ of contrast $m$ simulated at source field $f_s$, the goal is to synthesize its appearance at the target field $f_t$ while preserving the underlying anatomy. Same-subject, same-field images from the remaining contrasts are used only as auxiliary evidence, providing complementary anatomical and structural cues rather than serving as direct cross-contrast translation targets. The shared target-conditioned translator is written as

$$\hat{y} = G_\theta\big(\tilde{x}, \{(\tilde{a}_k, q_k)\}_{k=1}^{2}; f_t, m\big)\,,$$

where $\tilde{a}_k$ denotes the $k$-th auxiliary contrast and $q_k$ is an explicit auxiliary-validity mask indicating whether the corresponding auxiliary input is available. Fig. 1 summarizes the complete pipeline.

### 2.2 Anatomy-Anchored Data Augmentation

Because paired multi-field MRI acquisitions are limited, training the unified cross-field translator relies primarily on simulated pseudo-paired samples. To increase training diversity and expose the model to a broad distribution of acquisition conditions, we adopt an anatomy-anchored, field-aware stochastic input simulation strategy. Anatomical labels are first generated with SynthSeg and used as

structural anchors for subsequent simulation. During training, the real target image, its segmentation map, and all available auxiliary contrasts undergo the same randomly sampled 3D spatial transformation to augment anatomy while preserving cross-modal correspondence. This transformation combines affine scaling, rotation, shear, translation, left–right flipping, and moderate non-rigid displacement. After this spatial augmentation, the transformed anatomical labels provide tissue-level anchors for appearance simulation, enabling label-specific intensity variation and tissue-dependent contrast changes. The input image is then further diversified using two anatomy-anchored strategies.

*Domain-aware field simulation.* To simulate a pseudo-source field from a real target image, we first estimate 15 modality–field profiles using training volumes only. For each domain, we measure foreground intensity, spatial gradients, high-frequency energy, coarse-scale inhomogeneity, and SynthSeg-based tissue statistics. The 15th–85th percentiles across subjects define robust appearance ranges, while domain medians normalized by pooled training medians are mapped to bounded degradation intervals. For example, lower relative high-frequency energy produce stronger blur, downsampling, and frequency attenuation, whereas coarse-scale variations controlled noise and bias. At each iteration, simulation parameters are uniformly sampled from these domain-specific intervals. Polynomial bias, $k$-space motion, and undersampling are fixed a priori as physics-informed augmentations. When the pseudo-source field is lower than the target field, the simulator increases noise and bias, reduces isotropic and through-plane resolution, applies 2D $k$-space undersampling, lowers SNR, and strengthens frequency attenuation according to the field profile. Conversely, when a higher-field pseudo-source is simulated from a lower-field anchor, a thresholded two-band Laplacian decomposition then adds bounded, source-supported fine and mid-frequency detail rather than unconstrained texture.

*Label-conditioned pseudo-source synthesis.* To further broaden the pseudo-source distribution, we additionally synthesize the pseudo-source directly from segmentation maps with probability $Ps$. This follows the label-conditioned pseudo-source image generation strategy used in SynthSeg and SynthSR, where deformed label maps are converted into images by sampling label-wise intensities from randomized Gaussian mixture models, followed by acquisition-like perturbations such as multiplicative bias fields, noise, contrast transformations, and resolution changes through blurring, partial-volume effects, and subsampling.

### 2.3 Unified Cross-field Multi-contrast Image Translator

The main translator is a fully convolutional 3D residual network. A 3×3×3 input projection followed by four residual blocks maps the single-channel source patch to 256 features. In parallel, a half-resolution copy is processed by another four residual blocks to provide a larger effective receptive field for modeling coarse anatomical context and low-frequency field-dependent appearance at lower memory cost. The half-resolution features are upsampled trilinearly, projected by a

$1\times1\times1$ convolution, and added to the full-resolution features before eight main residual blocks. Each block applies convolution, feature-wise modulation, GELU, a second convolution and modulation, and residual addition. A final GELU and $1\times1\times1$ convolution followed by a sigmoid produce $\hat{y} \in [0, 1]$.

The requested output domain is made explicit through a task-condition encoder. Learned 32-dimensional field and 16-dimensional contrast embeddings are concatenated and mapped by a two-layer MLP to a 128-dimensional vector $c$. This vector modulates both convolutions in every low- and full-resolution residual block, following FiLM [10]:

$$\mathrm{FiLM}(h, c) = h \odot [1 + \gamma(c)] + \beta(c).$$

The same condition also controls the auxiliary encoders, aligning auxiliary evidence with the requested target field and contrast.

Auxiliary contrasts represent same-subject, same-source-field contrasts that differ from the primary contrast. If an auxiliary contrast is missing, the corresponding slot is zero-filled; during pretraining, each valid slot is independently dropped with probability 0.2. A shared projection followed by two target-conditioned residual blocks encodes valid auxiliary image(s) into 64 channels. At each of the eight main residual blocks, an auxiliary adapter, including a channel gate and a spatial gate, compares the current main feature with each auxiliary feature. The channel gate is predicted from globally pooled main and auxiliary features, and therefore estimates which auxiliary feature channels, or contrast-dependent feature types, are relevant to the requested target domain. In contrast, the spatial gate is predicted from high-pass representations of the main and auxiliary images and features, together with their absolute differences. It therefore estimates where the auxiliary contrast provides useful local anatomical support. At block $\ell$, the auxiliary channel and spatial information is added to the main features to selectively contribute to the main feature update.

### 2.4 Training Strategy

Training is performed in two stages. In Stage I, the model is trained only on unpaired data. Available auxiliary contrasts are used when present, but may be randomly omitted to improve robustness to missing auxiliary inputs. Real pairs are used only for validation after every epoch. In Stage II, the model is fine-tuned using the available real paired multi-field data, and we adopt subject-level three-fold leave-one-subject-out cross-validation. During this stage, all available auxiliary contrasts are used without random auxiliary dropout.

### 2.5 Objective Function

The objective combines complementary measures of intensity and anatomical fidelity. $\mathcal{L}_1$ is the mean absolute voxel error and constrains global intensity accuracy. $\mathcal{L}_{\mathrm{SSIM}} = 1 - \mathrm{SSIM}$ [12] compares local means, variances, and covariance,

emphasizing luminance, contrast, and structural agreement. $\mathcal{L}_{\nabla}$ averages the absolute discrepancy between first-order finite differences along the three spatial axes and therefore preserves tissue interfaces. $\mathcal{L}_{\text{wave}}$ is the absolute difference between high-pass residuals obtained by subtracting a $3 \times 3 \times 3$ local average, encouraging fine anatomical texture. The total loss is

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_{\text{ssim}} \mathcal{L}_{\text{ssim}} + \lambda_{\nabla} \mathcal{L}_{\nabla} + \lambda_{\text{wave}} \mathcal{L}_{\text{wave}}.$$

## 3 Results

### 3.1 Experimental Settings

**Dataset** The official dataset covers five field strengths and three contrasts. Its retrospective portion contains 1939 unpaired multicenter volumes: 703 T1w, 587 T2w, and 649 T2-FLAIR scans, with nonuniform coverage across field strengths. The prospective portion comprises three travelling volunteers, each scanned at all five fields with all three contrasts, yielding 45 paired volumes. We use the retrospective data in Stage I, whereas subject-level splits of the prospective cohort support Stage II fine-tuning.

**Implementation Details** The model is implemented in PyTorch. Both training stages use a patch size of $96^3$, a batch size of 1, and the AdamW optimizer with a peak learning rate of $10^{-5}$. Stage I is trained for 50 epochs, followed by Stage II fine-tuning for 25 epochs; the complete training process takes approximately 7 days on a single NVIDIA A100 80GB GPU. For simulated input generation, low-to-high and high-to-low field translations are sampled with probabilities of 0.7 and 0.3, respectively. Label-conditioned simulated image synthesis is applied with probability $P_s = 0.35$. The training objective uses $\lambda_1 = 0.8$, $\lambda_{\text{ssim}} = 0.2$, $\lambda_{\nabla} = 0.1$, and $\lambda_{\text{wave}} = 0.15$. During inference, whole-volume prediction is performed using sliding-window patch aggregation with 75% overlap.

### 3.2 Quantitative Results

Quantitative performance is evaluated using structural similarity (SSIM), normalized root-mean-square error (nRMSE), and learned perceptual image patch similarity (LPIPS), which respectively measure structural fidelity, normalized voxel-wise error, and perceptual image discrepancy. We compare the proposed method with the official challenge baseline, StarGAN v2 [4] (Baseline), and three ablations of our model: Uni_Contrast using only the primary contrast without auxiliary input; Multi_Com., which directly concatenates the primary and auxiliary contrasts at the input level; and Multi_Com.+Label Syn, which adds label-conditioned pseudo-source synthesis to Multi_Com. while keeping the same direct concatenation strategy. As shown in Table 1, the proposed model achieves the highest mean SSIM and the lowest mean nRMSE and LPIPS across all MRI contrasts, with scores of 0.908, 0.230, and 0.105, respectively.

**Table 1.** Quantitative comparison on cross-field MRI translation. Higher SSIM and lower LPIPS/nRMSE indicate better performance. The best result for each metric is shown in bold.

| Contrast | Metric | Baseline | Uni_contrast | Multi_com. | Multi_com. + Label Syn | Ours |
|---|---|---|---|---|---|---|
| T1w | SSIM ↑ | 0.718 | 0.876 | 0.891 | 0.894 | **0.907** |
| | LPIPS ↓ | 0.156 | 0.137 | 0.131 | 0.111 | **0.110** |
| | nRMSE ↓ | 0.358 | 0.325 | 0.306 | 0.302 | **0.243** |
| T2w | SSIM ↑ | 0.649 | 0.883 | 0.902 | 0.902 | **0.915** |
| | LPIPS ↓ | 0.250 | 0.122 | 0.118 | 0.112 | **0.095** |
| | nRMSE ↓ | 0.715 | 0.275 | 0.263 | 0.253 | **0.243** |
| T2-FLAIR | SSIM ↑ | 0.686 | 0.856 | 0.888 | 0.889 | **0.901** |
| | LPIPS ↓ | 0.244 | 0.138 | 0.120 | 0.128 | **0.109** |
| | nRMSE ↓ | 0.432 | 0.249 | 0.206 | 0.205 | **0.204** |
| Average | SSIM ↑ | 0.684 | 0.872 | 0.894 | 0.895 | **0.908** |
| | LPIPS ↓ | 0.217 | 0.132 | 0.123 | 0.117 | **0.105** |
| | nRMSE ↓ | 0.502 | 0.283 | 0.258 | 0.253 | **0.230** |

Compared with 2D StarGAN v2, the improvement suggests the benefit of 3D translation for preserving inter-slice consistency and anatomical structure. The comparison with Uni_Contrast demonstrates that auxiliary contrasts provide complementary information for improving synthesis accuracy and image quality. In addition, the improvement over Multi_Com. and Multi_Com.+Label Syn supports the effectiveness of the proposed auxiliary fusion strategy, where the model selectively incorporates useful channel-level features and spatial anatomical evidence from auxiliary contrasts. This design further improves SSIM by 0.013 and reduces nRMSE and LPIPS by 0.023 and 0.012, respectively.

### 3.3 Qualitative Results

To assess perceptual quality and anatomical fidelity, we visualize representative translations from different methods, including three low-to-high field and three high-to-low field cases. As shown in Fig. 2, the proposed method yields more anatomically plausible structures and target-like tissue contrast than the baseline and ablation models, with fewer visually apparent artifacts. Relative to StarGAN v2, training with simulated pseudo-paired images better preserves subject-specific anatomy than the CycleGAN-based translation strategy. In addition, 3D patch-based generation with overlapping sliding-window inference substantially reduces the inter-slice inconsistency observed in 2D synthesis. The comparison with Uni_Contrast indicates that auxiliary contrasts help recover fine anatomical details, while the improvement over Multi_Com. shows that the proposed auxiliary adapter further enhances overall image quality, particularly for T2w and T2-FLAIR synthesis. Since these contrasts often exhibit stronger noise, artifacts, and lower spatial resolution in low-field acquisitions, the proposed adapter helps suppress the amplification of low-field degradation while preserving more realistic anatomical detail in the synthesized high-field images.

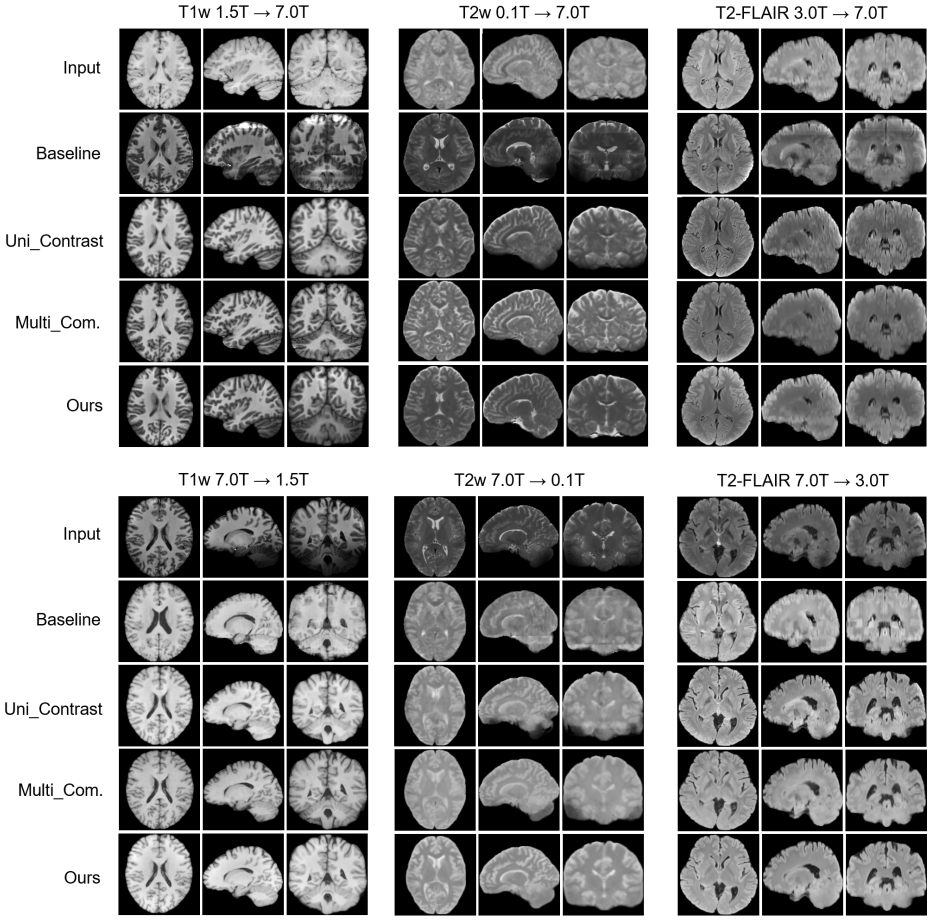


**Fig. 2.** Qualitative comparison of synthesized images using different methods. Representative low-to-high and high-to-low translations are shown for T1w, T2w, and T2-FLAIR, including 1.5T→7T and 7T→1.5T for T1w, 0.1T→7T and 7T→0.1T for T2w, and 3T→7T and 7T→3T for T2-FLAIR.

## 4 Conclusion

In this work, we propose a unified task-conditioned 3D framework for multi-contrast cross-field brain MRI translation and harmonization. By conditioning the translator on the requested field strength and contrast, the model supports diverse low-to-high, high-to-low, and intermediate-field translation directions within a single network. Same-subject auxiliary contrasts are incorporated through gated adapters that selectively use global contrast relevance and local anatomical support, while the anatomy-anchored field-aware simulator generates diverse pseudo-paired samples from real target images and segmentations to reduce dependence on scarce paired acquisitions. Preliminary validation re-

sults suggest that these components can improve cross-field synthesis quality and provide a practical basis for multi-center MRI field harmonization.

The current study is limited by the small paired cohort, and its performance may therefore depend on the quality of the simulated data. The comparative evaluation is also limited to the official baseline and controlled internal ablations. Future work will include broader multi-institutional paired evaluation, comparisons with stronger 3D conditional and diffusion-based baselines, and downstream analyses to assess whether the harmonized images facilitate joint image use and quantitative assessment across heterogeneous MRI datasets.

## References


1. Arnold, T.C., Freeman, C.W., Litt, B., Stein, J.M.: Low-field MRI: Clinical promise and challenges. Journal of Magnetic Resonance Imaging **57**(1), 25–44 (2023). https://doi.org/10.1002/jmri.28408
2. Bahrami, K., Shi, F., Zong, X., Shin, H.W., An, H., Shen, D.: Reconstruction of 7t-like images from 3t MRI. IEEE Transactions on Medical Imaging **35**(9), 2085–2097 (2016). https://doi.org/10.1109/TMI.2016.2549918
3. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (2023)
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. Cooley, C.Z., McDaniel, P.C., Stockmann, J.P., Srinivas, S.A., Cauley, S.F., Sliwiak, M., Sappo, C.R., Vaughn, C.F., Guerin, B., Rosen, M.S., Lev, M.H., Wald, L.L.: A portable scanner for magnetic resonance imaging of the brain. Nature Biomedical Engineering **5**, 229–239 (2021). https://doi.org/10.1038/s41551-020-00641-5
6. Dayarathna, S., Islam, K.T., Chen, Z.: Ultra low-field to high-field MRI translation using adversarial diffusion. In: IEEE International Symposium on Biomedical Imaging (2024). https://doi.org/10.1109/ISBI56570.2024.10635808
7. 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., van Zijl, P.C.M., Prince, J.L.: Deepharmony: A deep learning approach to contrast harmonization across scanner changes. Magnetic Resonance Imaging **64**, 160–170 (2019). https://doi.org/10.1016/j.mri.2019.05.041
8. Iglesias, J.E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gonzalez, R.G., Alexander, D.C., Golland, P., Edlow, B.L., Fischl, B.: 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 (2021). https://doi.org/10.1016/j.neuroimage.2021.118206
9. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision. pp. 319–345 (2020)
10. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)

11. Qu, L., Zhang, Y., Wang, S., Yap, P.T., Shen, D.: Synthesized 7t MRI from 3t MRI via deep learning in spatial and wavelet domains. Medical Image Analysis **62**, 101663 (2020). https://doi.org/10.1016/j.media.2020.101663
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
13. 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. 2242–2251 (2017). https://doi.org/10.1109/ICCV.2017.244
14. Zuo, L., Liu, Y., Xue, Y., Dewey, B.E., Remedios, S.W., Hays, S.P., Bilgel, M., Mowry, E.M., Newsome, S.D., Calabresi, P.A., Resnick, S.M., Prince, J.L., Carass, A.: HACA3: A unified approach for multi-site mr image harmonization. Computerized Medical Imaging and Graphics **109**, 102285 (2023). https://doi.org/10.1016/j.compmedimag.2023.102285