



```
<h3>Learning multi-site harmonization of magnetic  
resonance images without traveling human phantoms ·  
Concept Sketch</h3>  
<p>Neural dynamics lens highlighting connectivity  
vs. representation trade-offs.</p>
```

# Learning multi-site harmonization of magnetic resonance images without traveling human phantoms

**Authors:** Siyuan Liu, Pew-Thian Yap

**Year:** 2024

**Venue:** Communications Engineering (Nature)

# 1. Classification

- **Domain Category:**

- Brain MRI harmonization for multi-site neuroimaging. The work focuses on structural brain MRI (T1/T2) and methods to remove site/scanner artifacts while preserving anatomy.

- **FM Usage Type:**

- Not a foundation model per se; instead, a specialized deep generative harmonization framework that can serve as front-end infrastructure for brain FMs and downstream predictive models.

- **Key Modalities:**

- Structural T1-weighted and T2-weighted brain MRI volumes from large multi-site cohorts (e.g., ABCD scanners from GE, Philips, Siemens).
- 

## 2. Executive Summary

This paper introduces **MURD (Multi-site Unsupervised Representation Disentangler)**, a deep learning framework for harmonizing structural MRI images collected across many scanners and sites **without** requiring traveling human phantoms (the same subject scanned at multiple sites). Instead of learning a separate mapping for each site pair, MURD learns a single model that decomposes each image into **site-invariant anatomical content** and **site-specific appearance style** (contrast, intensity, scanner artifacts). By recombining content and style codes, the network can generate harmonized images for any target site, or retain appearance while testing identity mappings. Trained on more than 6,000 multi-site T1/T2 volumes from the ABCD study, MURD produces images that closely match target-site appearance while maintaining fine anatomical details. Quantitative comparisons against strong baselines (e.g., DRIT++, StarGAN-v2) show that MURD yields better FID/KID scores, improves segmentation consistency, reduces site differences in

volumetric measurements, and even supports cross-resolution and continuous-style harmonization. Overall, the method offers a practical way to retrospectively harmonize large, existing multi-site MRI datasets for more reliable downstream analysis and model training.

---

### 3. Problem Setup and Motivation

#### Scientific / practical problem

- Modern MRI studies increasingly pool data from many scanners and sites (e.g., ABCD, ADNI, AIBL) to achieve the sample sizes needed for robust statistical analysis.
- Multi-site acquisition introduces **non-biological variability** from scanner vendor, hardware, protocol differences, and resolution that can confound downstream analyses and machine learning models.
- Prospective protocol harmonization helps but is expensive, needs to be designed before data collection, and cannot fix previously acquired datasets.

#### Why this is hard

- **Scanner and protocol variation:** Different vendors and sequence settings produce very different contrast and noise characteristics, even when nominal parameters are matched.
- **Retrospective harmonization:** Correcting already-collected data requires methods that can separate anatomical information from acquisition-induced appearance differences.
- **Existing methods trade-offs:**
  - Statistics-based approaches (intensity normalization, ComBat-style batch correction) usually operate on global intensity summaries and struggle with spatially varying artifacts.
  - Pairwise deep learning methods learn mappings between specific site pairs; for (N) sites, they may require (N(N-1)) mappings, which

does not scale and wastes shared information.

- **Risk of anatomical distortion:** Any harmonization method must avoid hallucinating or altering subtle anatomical features that matter for clinical or research questions.

## 4. Data and Modalities

- **Datasets used**

- Multi-site **ABCD** study: more than 6,000 T1-weighted and T2-weighted brain MRI volumes from children aged roughly 9–10 years, acquired on GE, Philips, and Siemens scanners.
- Data are grouped into three “virtual sites” corresponding to vendor families, each with matched acquisition protocols but noticeable appearance differences.

- **Modalities**

- Structural MRI only:
  - T1-weighted images (anatomical structure, gray/white contrast).
  - T2-weighted images (complementary tissue contrast).

- **Preprocessing / representation**

- Volumes are aligned and preprocessed using standard neuroimaging tools (e.g., ANTs-based registration) to ensure anatomical correspondence across subjects and scanners.
- Central axial slices or 2.5D slices (three adjacent slices stacked) are extracted and fed into the network for training and evaluation.
- Images are grouped by nominal site and modality to define domains for harmonization (e.g., GE-T1, Philips-T1, Siemens-T1).

If some dataset or preprocessing details are missing from the extracted text, they are likely described in more depth in the original paper’s methods and supplementary material.

## 5. Model / Foundation Model

Although not framed as a “foundation model”, MURD is a **deep generative architecture** designed for scalable, multi-site MRI harmonization via disentangled representations.

### Model type

- Multi-domain **image-to-image translation** framework with:
  - A **site-shared content encoder** capturing anatomical structure.
  - **Site-specific style encoders and generators** capturing scanner-dependent appearance.
  - A **site-shared decoder/generator** that recombines content and style to synthesize harmonized images.
  - Site-specific discriminators for adversarial learning.

### Key components and innovations

Component	Description
Content encoder	Maps input MR images into a <b>site-invariant content representation</b> encoding anatomy that should be preserved across scanners.
Style encoder	Extracts <b>site-specific style codes</b> capturing contrast, intensity distributions, and scanner/protocol artifacts.
Generator / decoder	Combines content and style codes to synthesize images for a chosen target site, enabling harmonization or identity mappings.
Style generator	

Component	Description
	For each site, generates randomized style codes so that multiple realistic appearances can be sampled for a given anatomical content.
Discriminators	Site-specific discriminators enforce realism of generated images and alignment with each site's appearance distribution.
Losses	A combination of <b>consistency (cycle, content, style)</b> losses, <b>adversarial loss</b> , <b>content alignment</b> , <b>style diversity</b> , and <b>identity</b> losses ensures that anatomy is preserved while appearance is altered appropriately.

### Training setup (high level)

- MURD is trained separately for T1-weighted and T2-weighted volumes.
- Training uses modest labeled data per site (e.g., 20 volumes/vendor for training, plus separate validation/generalization/human phantom test sets).
- The network is optimized so that:
  - Harmonized images match the target site's distribution (via adversarial and style losses).
  - Content representations remain consistent across harmonization and identity mappings.
  - Structural details are preserved, assessed via downstream segmentation and volumetry.

## 6. Multimodal / Integration Aspects (If Applicable)

- This paper focuses on **single-modality structural MRI harmonization** across many scanners and protocols.
  - It does not explicitly integrate multiple biological modalities (e.g., genetics, behavior, multimodal FMs).
  - However, by reducing non-biological site variance in T1/T2 images, MURD is complementary to multimodal integration pipelines and brain foundation models that consume harmonized MRI as input.
- 

## 7. Experiments and Results

### Tasks / benchmarks

- **Visual quality evaluation:** qualitative assessment of harmonized T1/T2 images when mapping between GE, Philips, and Siemens sites, including identity mappings and reference-image-based harmonization.
- **Image quality metrics:** Frechét Inception Distance (FID) and Kernel Inception Distance (KID) comparing harmonized images to real images from target sites.
- **Human phantom evaluation:** traveling human phantom dataset with subjects scanned on multiple vendors to quantify structural preservation and harmonization quality.
- **Segmentation consistency:** brain extraction and tissue segmentation (e.g., BET + FAST) before and after harmonization, measuring Dice similarity coefficients.
- **Volumetric measures:** comparing distributions of GM/WM/CSF volumes across sites before and after harmonization, and examining preservation of biological effects such as gender differences.

- **Resolution and continuous harmonization:** cross-resolution harmonization (1.25 mm → 1 mm) and continuous interpolation of style codes between sites.

## Baselines

- Statistics-based harmonization approaches (e.g., intensity normalization, batch-effect correction) as conceptual background.
- Deep learning baselines: **DRIT++** and **StarGAN-v2**, representing prior unsupervised dual-domain and multi-domain image-to-image translation methods for style transfer and domain adaptation.

## Key findings (trends)

- MURD produces harmonized images whose appearance closely matches target sites while preserving detailed anatomy, both qualitatively and quantitatively.
- On FID and KID, MURD outperforms DRIT++ and StarGAN-v2, with scores much closer to “reference” values computed between real training and testing images from the same site.
- In the traveling human phantom dataset, MURD yields lower mean absolute error, higher structural similarity (MS-SSIM), and better PSNR than baselines, indicating better structural fidelity.
- Tissue segmentation consistency (Dice scores) improves substantially after harmonization, and identity mappings show that harmonization does not degrade segmentation when source and target sites are identical.
- Volumetric distributions across sites (GM, WM, CSF) become better aligned after harmonization while preserving biologically meaningful gender differences.
- Cross-resolution and continuous-style experiments show that MURD can recover fine details from lower-resolution images and support smooth transitions between site appearances without introducing artifacts.



## 8. Strengths, Limitations, and Open Questions

### Strengths

- Addresses a **practical bottleneck** in multi-site MRI research: harmonizing images without requiring traveling human phantoms or paired scans.
- Scales beyond pairwise mappings to **many sites within a single unified model**, reducing the number of networks that must be trained and maintained.
- Uses a principled **content-style disentanglement** design and rich loss functions to preserve anatomy while adjusting scanner-specific appearance.
- Demonstrates effectiveness across multiple evaluation angles: visual quality, image similarity metrics, segmentation consistency, volumetric statistics, and cross-resolution harmonization.

### Limitations

- Focuses on structural MRI (T1/T2) and healthy/developmental cohorts; applicability to other modalities (e.g., diffusion, fMRI, clinical populations) is not fully explored.
- Training and evaluation are centered on a specific set of vendors and protocols; performance on very different scanners or field strengths is uncertain.
- As a generative model, MURD may still introduce subtle biases or artifacts that are hard to detect without extensive validation.
- The framework requires reasonably good preprocessing and registration; failures there can propagate into harmonized outputs.

### Open questions / future directions

1. How well does MURD generalize to other anatomical regions, body parts, or imaging modalities (e.g., diffusion MRI, whole-body MRI)?

2. Can the learned content representations be repurposed as initializations or inputs for **brain foundation models** and multimodal FMs?
  3. How robust is MURD to severe motion artifacts, rare scanner types, or out-of-distribution protocols?
  4. What formal guarantees or additional validation strategies can be used to ensure that harmonization never alters disease-relevant anatomical signals?
  5. Could lighter-weight or semi-supervised variants of MURD support harmonization in settings with fewer images per site?
- 

## 9. Context and Broader Impact

- **Within MRI harmonization:** MURD extends the move from global intensity corrections and pairwise mappings to **unified, multi-site deep harmonization**, explicitly disentangling anatomy from scanner appearance.
  - **Relation to foundation models:** By cleaning up non-biological variance in structural MRI, MURD can make inputs more homogeneous for **brain FMs and multimodal models**, potentially improving downstream generalization and fairness.
  - **Neuroimaging practice:** The approach is particularly relevant for large-scale consortia (ABCD, ADNI, etc.) and retrospective harmonization of legacy datasets where re-acquisition is impossible.
  - **Ethical considerations:** While harmonization can reduce site bias, it also risks masking systematic differences that relate to demographic or health disparities; careful auditing and transparent reporting are important.
-

## 10. Key Takeaways (Bullet Summary)

- **Problem:** Multi-site MRI studies suffer from scanner- and protocol-induced variability that can overwhelm biological signals and distort downstream analyses.
- **Idea:** Learn a unified deep model (MURD) that **disentangles** site-invariant anatomical content from site-specific appearance style and recombines them to generate harmonized images.
- **Model:** A content encoder, style encoders/generators, site-shared generator, and site-specific discriminators trained with a combination of cycle, content, style, adversarial, and identity losses.
- **Data:** Thousands of T1/T2 images from the ABCD study across GE, Philips, and Siemens scanners, plus a traveling human phantom dataset and large generalization sets.
- **Results:** MURD outperforms DRIT++ and StarGAN-v2 on FID/KID, improves segmentation consistency and volumetric agreement across sites, supports cross-resolution and continuous-style harmonization, and preserves anatomical details.
- **Impact:** Provides a scalable, retrospective harmonization tool that can clean multi-site MRI datasets for more robust statistical analysis and as input to future brain foundation models and multimodal integration pipelines.