```
Dorent R, Haouchine N, Kogl F, et al. Unified Brain MR-Ultrasound
Synthesis using Multi-Modal Hierarchical Representations. MICCAI 2023.
<https://arxiv.org/abs/2309.08747>
```

Contents

1 One-Sentence Verdict

Deep read. The core synthesis engine behind Dorent 2024's patient-specific segmentation — the SOTA method for generating synthetic iUS from MRI. Pre-trained model publicly available; foundational component for the "synthetic data" route in my project.

2 Research Question & Background Gap

Research question: Can a single unified model achieve bidirectional multi-modal (MRI↔iUS) image synthesis while handling incomplete MRI inputs?

Background gap:

- Standard Multi-modal VAE (MVAE) latent space is too low-dimensional, producing blurry synthesis

- Conditional GANs (Pix2Pix/SPADE) are not unified models — each synthesis direction requires separate training and cannot handle partially missing inputs
- Transformer methods (ResViT) perform well but have massive parameter counts (293M vs MHVAE's 10M)

3 Methods & Data

1. 3.1 MHVAE Architecture

Component	Details
Latent levels	L=7, from z1 (192×192×8) to z7 (1×1×256)
Encoder	U-Net structure, MobileNetV2 residual cells + SE + Swish
Fusion	Product-of-Experts (PoE), independent fusion at each level
Decoder	5 ResNet blocks per modality
Adversarial learning	PatchGAN discriminator; first 800 epochs λ_GAN=0, last 200 epochs λ_GAN=1 (staged strategy: let VAE stabilize reconstruction and KL regularization first, then introduce adversarial signal for perceptual quality, avoiding early GAN interference causing mode collapse or training instability. MMHVAE TPAMI version adjusts this threshold to epoch 790)
Model size	Paper claims ~10M parameters, 8G MACs (1/30 and 1/60 of ResViT). However, MMHVAE TPAMI states MMHVAE ~14M with "only 4% increase," implying MHVAE should be ~13.5M. The 10M figure likely counts only the main network excluding the GAN discriminator

The core innovation is **PoE fusion at every latent level**, not just the deepest. Shallow levels (z1) capture local textures (speckle etc.), deep levels (z7) capture global structure (anatomical layout). This ensures synthesized images maintain both anatomical correctness and local realism. (Section 3.1–3.2)

2. 3.2 Data & Training

- **Data**: ReMIND 66 cases, T2-SPACE + pre-dura iUS (affine-registered via NiftyReg), 0.5mm isotropic, 192×192
- **Training**: 1000 epochs, batch size 16, 56 training / 10 test cases
- **Loss**: L1 + KL divergence + PatchGAN (last 200 epochs)

4 Key Evidence

3. 4.1 Table 1: Synthesis Quality Comparison (Section 4)

T2→iUS direction (most relevant to my project):

Method	PSNR↑	SSIM↑	LPIPS↓
Pix2Pix	20.31 dB	70.2%	19.8%
SPADE	20.30 dB	70.1%	21.5%
MVAE	21.21 dB	73.5%	26.9%
ResViT	21.22 dB	**75.2%**	24.0%
MHVAE (no GAN)	**21.87 dB**	74.9%	24.2%
MHVAE (with GAN)	21.26 dB	71.9%	**19.0%**

MHVAE with GAN achieves the best **perceptual quality** (LPIPS), which may matter more than PSNR for downstream segmentation. All methods achieve T2→iUS PSNR in the 20–22 dB range — iUS synthesis is inherently difficult due to speckle and artifacts making pixel-level reconstruction challenging.

4. 4.2 Ablation Studies (Section 4)

- **MVAE vs MHVAE**: Hierarchical structure yields significant improvement (MVAE blurry vs MHVAE sharp)
- **GAN loss effect**: Metric differences are modest, but GAN version produces visually more realistic speckle textures
- **Reconstruction quality** (input = target): iUS PSNR 33.15 dB, T2 PSNR 36.38 dB → latent space retains complete information

5. 4.3 Fig 4 (Appendix): BraTS Generalization

Synthetic iUS generated from BraTS dataset T2 images — MHVAE **generalizes to MRI data outside the training set**. This is the basis for Dorent 2024's BraTS UNet trained on 611 cases. (Qualitative only, no quantitative evaluation)

5 Author Claims & My Critical Assessment

6. 5.1 What the Paper Explicitly States

- MHVAE outperforms MVAE, Pix2Pix, SPADE, and ResViT on MRI-iUS synthesis
- Hierarchical latent is the key innovation
- Model is lightweight (10M parameters) and handles incomplete inputs
- Discussion explicitly mentions using "BraTS annotations to generate synthetic iUS for training segmentation networks"

7. 5.2 What Can Be Reasonably Inferred

T2→iUS synthesis quality is moderate (PSNR ~21 dB, SSIM ~72%) — not photorealistic but possibly sufficient for segmentation tasks. Dorent 2024 confirms this (model trained on synthetic data achieves DSC 87%).
MHVAE training relies on 66 paired ReMIND cases (single-center BWH) — synthesis quality may be biased toward that center's equipment and scanning protocol. However, the Appendix

BraTS generalization examples suggest the bias is manageable.
The ~10–14M parameters + 8G MACs means fast inference, suitable for generating large volumes of synthetic data.

8. 5.3 What Remains Uncertain

- **Quantitative** quality of BraTS→synthetic iUS — Appendix only shows images
- Synthesis quality for low-grade gliomas (diffuse boundaries, low contrast)
- Generalization to multi-center MRI (different equipment/protocols)
- Which synthesis defects propagate to segmentation errors when synthetic iUS is used for training?

6 Relevance to My Project

The MHVAE 2-modality code is available at GitHub ReubenDo/MHVAE. It can generate synthetic iUS from T2/ceT1/FLAIR. The temperature parameter τ controls synthesis variability (higher τ = more speckle), and incomplete inputs are supported — not all MRI sequences need to be available.
MHVAE is pre-trained on ReMIND data. If my project's MRI data differs significantly in resolution or sequences, synthesis quality may degrade.
Faanes takes the "registration-propagated MRI annotations" route (no synthetic image generation), while MHVAE/Dorent takes the "generate synthetic iUS" route. Faanes' route propagates MRI annotations to iUS space via registration to form pseudo labels, then trains directly on real iUS. Dorent's route generates synthetic iUS from MRI plus MRI annotations, trains on synthetic iUS, and infers on real iUS.

7 My Questions & Ideas

What is the relationship between synthesis quality and downstream performance? Dorent 2024 trained a DSC 87% model on synthetic data with only ~21 dB PSNR, suggesting segmentation networks have high tolerance for synthesis imperfections. Could this be systematically studied (degradation study)? Could newer generative models (diffusion models) replace MHVAE? Potentially improving synthesis quality but requiring retraining.
MHVAE's BraTS generalization capability hints at an important route: generate synthetic iUS from thousands of BraTS MRI cases, then train a generic segmentation model. This directly tests the value of open data.

8 Key References

- Wu & Goodman 2018 — MVAE original paper (theoretical foundation of MHVAE)
- Vahdat & Kautz 2020 — NVAE (architectural inspiration for hierarchical VAE)
- Dorent et al. 2024 — Patient-Specific Segmentation (downstream application of MHVAE, **already deep-read**)

#image-synthesis #cross-modal #MRI-to-US-synthesis #HierarchicalVAE #ProductOfExperts #adversarial-learning #high-priority