```
Dorent R, Torio E, Haouchine N, et al. Patient-Specific Real-Time
Segmentation in Trackerless Brain Ultrasound. MICCAI 2024 (Early
Accept). <https://arxiv.org/abs/2405.09959>
```

Contents

1 One-Sentence Verdict

Deep read. Core paper for the "synthetic data" direction — generates synthetic iUS from pre-operative MRI to train patient-specific segmentation models without tracking systems, achieving DSC surpassing both expert and navigation systems. MICCAI 2024 Early Accept, code public, directly aligned with supervisor's requirements.

2 Research Question & Background Gap

Research question: Can synthetic iUS generated from pre-operative MRI be used to train a patient-specific model for trackerless real-time 2D iUS brain tumor segmentation?

Background gap:

- Real iUS annotations are extremely scarce and annotation protocols are inconsistent (BraTS annotations ≠ surgeon's surgical target definition)
- Tracking systems (e.g., Brainlab Curve) are expensive and overlay fails due to intraoperative brain shift
- Existing non-patient-specific models cannot adapt to different surgeons' definitions of "surgical target"

3 Methods & Data

1. 3.1 Three-Step Framework

Step	Content	Technical Details
Step 1: Virtual sweep simulation	Simulate US probe trajectory in MRI space	Randomly select geometric parameters from a reference sweep, place probe on MRI brain surface near tumor, align principal direction with tumor's first principal component
Step 2: MHVAE synthesis	Generate synthetic iUS from 2D MRI slices	Pre-trained hierarchical VAE (MHVAE, Dorent et al. MICCAI 2023), supports incomplete MRI input (any combination of ceT1/T2/FLAIR)
Step 3: Patient-specific training	Train 2D UNet with synthetic iUS + MRI annotations	100 epochs, batch ~100 2D images, SGD + Nesterov, Dice loss + deep supervision

2. 3.2 Data Variability Strategy

- **MRI sequence combinations**: 6 (power set of ceT1/T2/FLAIR minus empty set)
- **Sampling temperature**: $\tau \in \{0.3, 0.5, 0.7, 1.0\}$, controlling synthetic speckle intensity
- **Virtual sweep count**: K=10 reference sweeps × 6 combinations × 4 temperatures = 240 synthetic data groups

3. 3.3 Experimental Design

- **Dataset**: ReMIND 7 cases (excluded from MHVAE training), ~70 2D iUS slices per case
- **Resolution**: 0.5mm isotropic, 192×192

- **Annotation protocols**: BraTS automatic annotation, N1 (surgeon 1, ReMIND public annotation), N2 (surgeon 2, manual iUS annotation)
- **Evaluation ground truth**: MRI annotations propagated to iUS space via NiftyReg affine registration (**not true iUS annotations**)
- **Metrics**: DSC, ASSD, Wilcoxon signed-rank test ($p<0.01$)
- **Hardware**: NVIDIA A5000 24GB

4 Key Evidence

4. 4.1 Table 1: Effect of Virtual Sweep Count K

K=1 → DSC 79.7%, K=10 → DSC 84.2%, K=15 → DSC 84.9%. K=10 requires ~210 min (3.5h) development time, K=15 needs 309 min with only 0.7pp gain. **Diversity of synthetic data (not sheer volume) is what matters** — 10 sweeps already reach the diminishing returns point. (Table 1, Section 3)

5. 4.2 Table 2: Method Comparison (Core)

Method	BraTS DSC	N1 DSC	N2 DSC
RESECT UNet (trained on real iUS annotations)	71.7%	59.8%	58.5%
BraTS UNet (synthetic iUS, non-patient-specific)	76.8%	56.1%	53.8%
Tracking System (Brainlab navigation)	85.4%	80.0%	80.1%
Expert manual segmentation (N2)	—	—	67.4%
Ours (K=10)	**87.2%	**84.2%	**84.2%

Key findings:

- **Patient-specificity is critical**: BraTS UNet (same synthetic data but non-patient-specific) drops to ~55% on N1/N2 annotations because it only learned the BraTS protocol. Ours remains stable at 84–87% across all three protocols
- **Surpasses tracking system**: Ours 87.2% vs Tracking 85.4% (BraTS), without expensive equipment
- **Expert iUS segmentation is extremely difficult**: N2 manual achieves only 67.4%, far below automatic methods
- Ours vs RESECT UNet: +15.5pp (BraTS), +24.4pp (N1) — substantial gap

6. 4.3 Figure 3 (Based on Text Description)

Qualitative comparison shows RESECT/BraTS UNet producing false positives in non-target regions, while Ours aligns better with actual surgical targets. The tracking system shows overlay misalignment due to brain shift (Figure 1, red arrows).

5 Author Claims & My Critical Assessment

7. 5.1 What the Paper Explicitly States

- Patient-specific synthetic data training significantly outperforms non-patient-specific methods on 7 cases
- Inference runs at 200 FPS in real time
- No tracking system needed — only pre-operative MRI + annotations required

8. 5.2 What Can Be Reasonably Inferred

The evaluation has circular bias. The ground truth itself comes from MRI→iUS registration propagation, inherently favoring methods that use MRI information (both Ours and Tracking System use MRI; RESECT UNet does not). This may inflate Ours' advantage over RESECT UNet. **Quantitative evidence** (added after reading MMHVAE TPAMI 2025): when the same team evaluated on RESECT-SEG's independent manual annotations, DSC dropped from the circular evaluation's 0.87 to **0.74** (a 13 percentage point decline). 0.74 still matches the fully-supervised baseline (0.73), confirming the method works, but 0.87 should not be used as a performance expectation.

MHVAE does not truly eliminate the need for iUS data. Pre-training MHVAE requires paired MRI-iUS data (ReMIND) — it just doesn't need iUS **annotations**. So the "synthetic data" direction still depends on existing paired datasets.

BraTS UNet was trained on 611 UPenn-GBM cases — far more data than Ours' single-patient synthetic data — yet patient-specificity still wins. This suggests that in this task, **data relevance > data scale**.

9. 5.3 What Remains Uncertain

- MHVAE synthesis quality on low-grade gliomas (diffuse boundaries)
- Performance during/post-resection (this paper only uses pre-dura iUS)
- Whether 3.5h development time is clinically acceptable (pre-operative preparation time is typically limited)
- Whether 7-case statistical significance is sufficiently robust

6 Relevance to My Project

Reusable: The synthetic data generation pipeline design is complete (sweep simulation → MHVAE → training), and the data variability strategy (MRI sequence combinations × temperature × multiple sweeps) provides concrete implementation reference.

The paper's cited `MHVAE-Seg` repository (GitHub: ReubenDo/MHVAE-Seg) is a **placeholder** (README only). The actual MHVAE 2-modality code is at `ReubenDo/MHVAE`, but pre-trained weights are not publicly available. MMHVAE 4-modality TPAMI version code has not been released. Virtual sweep simulation code must be self-implemented.

Not directly reusable: Dorent's approach is patient-specific (separate training per patient), while my project leans population-level (one generic model for all patients). The patient-specific concept needs adaptation. The evaluation method has circular bias — my project should use RESECT-SEG's real iUS annotations for fairer evaluation.

Consistent settings: Both use ReMIND data, 2D iUS segmentation, and MRI as training signal source.

Warning

Inconsistent settings: Dorent uses pre-dura 3D→2D slices; Faanes uses before-resection iUS with a different pipeline. Architecture differences require careful attention in cross-paper comparisons — Dorent uses a simple UNet with deep supervision (Dice loss, SGD), while Faanes and MMHVAE TPAMI use nnU-Net (auto-configured architecture/preprocessing/augmentation). nnU-Net typically outperforms manually-tuned UNet, so Dorent's DSC 0.87 vs Faanes' DSC 0.62 cannot be entirely attributed to patient-specific vs population-level; architecture differences are a confounding factor.

Caution: Synthetic iUS quality is the bottleneck of the entire method — if MHVAE synthesis quality is poor, all downstream training fails. Patient-specific models cannot be pre-evaluated; you don't know the outcome before surgery.

7 My Questions & Ideas

The core idea is whether Dorent's patient-specific approach and Faanes' population-level MRI pseudo-label approach can be combined: first train a generic baseline with MRI pseudo labels, then fine-tune with MHVAE synthetic data in a patient-specific manner.
Dorent's synthetic route (MRI → synthetic iUS → training) and Faanes' pseudo-label route (MRI annotation → registration → iUS pseudo labels → training) are two parallel paths. Which is better? Can an experiment be designed for direct comparison? Evaluation fairness is the key issue — how to design an evaluation scheme independent of MRI registration? RESECT-SEG's real annotations are currently the only option. Could the 3.5h/patient development time be reduced through pre-training + fine-tuning?

8 Key References

- Dorent et al. 2023 — MHVAE: Unified Brain MR-Ultrasound Synthesis (synthesis engine, **must-read**)
- Juvekar et al. 2023/2024 — ReMIND dataset (Dorent is co-author, data source)
- Behboodi et al. 2022 — RESECT-SEG (only public dataset with real iUS annotations, for fair evaluation)
  #segmentation #intraoperative-ultrasound #MRI-to-US-synthesis #MHVAE #patient-specific #cross-modal-synthesis #high-priority