```
Rasheed H, Dorent R, Fehrentz M, et al. Learning to Match 2D Keypoints
Across Preoperative MR and Intraoperative Ultrasound. MICCAI 2024.
arXiv:2409.08169
```

Contents

1 One-Sentence Verdict

The Dorent/Golby/Frisken group uses MHVAE-synthesized US to train a cross-modal keypoint descriptor, achieving ~74% average matching precision on ReMIND (paper abstract claims "over 80%" as best-case; 3-case detailed average is 73.6%; traditional methods: 3–5%). **Deep read** — the matching-by-synthesis strategy can directly address our MI registration 13% failure rate problem and shares MHVAE as a prerequisite with Route B.

2 Research Question & Background Gap

How to achieve reliable 2D keypoint matching between pre-operative MR and intraoperative US? MR and US have fundamentally different imaging physics (MR: soft tissue

hydrogen proton relaxation contrast; US: acoustic impedance interface reflection + speckle noise), causing traditional feature descriptors (SIFT, MIND, SuperPoint) to drop to 3–5% precision in cross-modal scenarios.

The gap is that existing multi-modal matching methods primarily target modality pairs with higher similarity (e.g., different MRI sequences), with no dedicated feature learning scheme for the MR-US "completely uncorrelated textures" scenario. LC2 partially addresses this through physics simulation but requires commercial software (ImFusion).

3 Methods & Data

1. 3.1 Four-Stage Pipeline

A. MHVAE Synthesis (§2.2, Fig 2): Use pre-trained MHVAE (Dorent et al. MICCAI 2023) to generate 28 synthetic US images from patient MR. 7 modality combinations (T1/T2/FLAIR individually, pairwise, and all three) × 4 sampling parameters γ. γ controls speckle texture variation, enabling 1-to-many diversity.

B. Training Set Construction (§2.3): SuperPoint detects MR keypoints → force-detect same-location points on 28 synthetic US images (5px tolerance) → retain only points detected in ≥3 images (DBSCAN, 5px radius) → extract 64×64 patches, 256 per slice, ~half retained after clustering.

C. Siamese Descriptor Network (§2.4): Simple Siamese CNN with Triplet Loss (margin=1) + hard negative mining. Anchor = MR patch, Positive = same-location synthetic US patch, Negative = different-location synthetic US patch (hardest in-batch sample by L2 distance). **Patient-specific training**, ADAM lr=1e-3, batch=256, <30min/patient on 10GB GPU.

D. Inference & Matching (§2.4): Detect n=200 keypoints on MR, up to 1500 on US. Single-branch forward pass produces descriptor vectors → KNN Cosine matching + Lowe ratio test + uniqueness filter.

2. 3.2 Data & Evaluation

Item	Value
Dataset	ReMIND, 7 cases
MR modalities	T2-SPACE (primary) + T1 + FLAIR
US type	Pre-dural opening, 3D reconstructed from tracked 2D
Resolution	0.5mm isotropic, 192×192 in-plane
Normalization	[-1, 1]
Ground truth	ReMIND-provided MR-US registration alignment
Evaluation metrics	Precision (correct matches / total matches), MSc (correct matches / total detected points), MP (matched point count)
Correct match criterion	Keypoint position difference ≤ 4px

4 Key Evidence

3. 4.1 SOTA Comparison (Table 2, Fig 3)

Method	Prec (%)	MSc (%)	Avg MP
SIFT+Cosine	~3.3	~2.9	~187
MIND+Cosine	~4.8	~4.8	200
SP+Cosine	~3.7	~2.6	~148
SP+LightGlue	~28.8	~4.9	~18
Ours+Cosine	**~73.6**	**~18.2**	**~49.6**
Ours+LG	**~70.3**	**~9.7**	**~25.8**

(Table 2 reports only 3 cases; above are hand-computed averages. Inter-case variation is notable: Case 1 Prec=81%, Case 3 Prec=66%.)

P.S. **Absolute correct match comparison** (self-computed): MIND 200 matched points × 4.8% precision ≈ **10 correct matches**; Ours 50 matched points × 73.6% precision ≈ **37 correct matches**. Actual improvement is ~3–4×, far less than the 15–20× suggested by precision numbers alone. However, 37 high-confidence matches are more than sufficient for rigid registration (6 DOF), while 10 low-confidence points contaminated with outliers are hard to use.

Fig 3's visual comparison is extremely compelling — traditional methods show almost all red dots (incorrect matches), while Ours shows abundant green lines (correct matches).

4. 4.2 Ablation Study (Table 1)

Input Modalities	Prec (%)	MSc (%)	Avg MP	Area (%)
T2 only	**85.64**	7.06	16.50	25.05
T2+FLAIR	83.01	7.60	18.33	30.73
T2+T1	83.25	12.87	30.92	44.77
T2+T1+FLAIR	81.08	**20.32**	**50.14**	**55.11**

More modalities: precision drops slightly (85.6→81.1%), but matched point count triples (16.5→50.1), and spatial coverage area doubles (25→55%). Different modalities provide complementary information → more matchable anatomical features are detected.

5. 4.3 Slice Retrieval (Fig 5)

Using 200 MR keypoints to retrieve the corresponding slice across the entire US volume yields 1.34mm error (within a 20-slice range). This is essentially a crude 2D-to-3D localization — implying descriptors can be used for slice-to-volume registration initialization.

6. 4.4 Stability (Fig 4)

Precision across slices has std=7.1% with no systematic degradation. After excluding 12 synthesis modes, >50% of keypoints are still repeatedly detected → descriptors genuinely learn texture-invariant features rather than overfitting to specific synthesis patterns.

5 Author Claims & My Critical Assessment

7. 5.1 What the Paper Explicitly States

- Matching precision ~74% across 7 ReMIND cases (3-case detailed average 73.6%; abstract claims 80.35% for all 7 — discrepancy likely because Table 2 only reports 3 cases)
- Patient-specific training <30min/patient on 10GB GPU
- Uses MHVAE pre-trained weights for synthesis (weight source not public)
- 2D matching only

8. 5.2 What Can Be Reasonably Inferred

The matching-by-synthesis strategy effectively bridges the MR-US modality gap. Multi-modal synthesis diversity is crucial for robustness (Table 1 ablation). Patient-specific training circumvents the cross-patient generalization challenge.

9. 5.3 What Remains Uncertain

Warning **No TRE reported.** 80% matching precision ≠ registration accuracy. What is the actual registration error when 43 matched points (~34 correct) are converted to a rigid transformation? The paper does not address this at all.

- 7-case sample size is extremely small with notable inter-case variation (81% vs 66%), limiting statistical significance
- Only pre-dural US evaluated; whether it works post-dural (greater brain shift) is uncertain
- GT relies on ReMIND's Brainlab navigation alignment; the GT's own accuracy is not discussed

- No registration-level comparison with LC2/ImFusion
- Code not public

6 Relevance to My Project

High direct value — matching-by-synthesis is one of the most promising directions for solving our MI registration 13% failure rate.

Factor	Rasheed	Ours	Alignment
Dataset	ReMIND (7 cases)	ReMIND (61 cases)	✅ Same dataset
MR modality	T2-SPACE primary	T1ce primary	⚠️ Needs verification
US type	Pre-dural only	Pre + post dural	⚠️ Paper only pre
Resolution	0.5mm iso, 192²	~1mm, non-isotropic	⚠️ Needs resampling
Target task	2D matching	3D rigid registration	❌ Needs additional conversion

Reusable: MHVAE synthesis pipeline (Route B already in progress), SuperPoint/LightGlue (public pre-trained models), patch-level Siamese training strategy can be directly borrowed. The matching-to-registration conversion does not exist in the paper (they didn't do it) and must be self-designed. Code is not public and needs self-implementation. The key prerequisite is MHVAE pre-trained weights, which are fully tied to Route B's progress.

7 My Questions & Ideas

Could we bypass keypoint learning entirely and use MHVAE-synthesized US for "bridging registration" — synthetic US vs real US with intensity matching (NCC), since they are same-modality?
This "bridging registration" approach is far simpler than fully reproducing the Rasheed pipeline and requires no SuperPoint/Siamese training. It essentially uses MHVAE to convert the cross-modal registration problem into a same-modality one. LC2's principle is similar (simulating US within the optimization loop), just with a different implementation (pre-generation vs online simulation).

P.S. **Cross-validation** (added after reading MMHVAE TPAMI 2025): MMHVAE paper Table 3 actually validates this bridging idea — synthesizing iUS into MRI modality then performing same-modality registration reduces ceT1 TRE from 14.6mm to **3.2mm**, FLAIR from 9.4mm to 4.4mm. This is direct evidence for "bridging registration" effectiveness, no need to experiment from scratch.
Another idea: Rasheed's matched points could directly serve as a QC metric — if a registration result has very few matched points or low precision, the registration likely failed. This is more informative than our current MI improvement + translation/rotation threshold QC.

8 Key References

- **[5] Dorent 2023 MICCAI** — MHVAE core paper, our Route B's key dependency, already deep-read
- **[3] DeTone 2018 SuperPoint** — Keypoint detector, publicly available

- **[15] Lindenberger 2023 LightGlue** — Deep matcher, publicly available
- **[10] Heinrich 2012 MIND** — Modality-agnostic descriptor, widely used in deformable registration, foundation of DeedsSSC in CuRIOUS 2018
- **[13] Juvekar 2023 ReMIND** — Our dataset
  #cross-modal-keypoint-matching #registration #ultrasound #MRI #T2 #T1 #FLAIR #MHVAE-synthesis #contrastive-learning #Siamese-network #SuperPoint #matching-by-synthesis #high-priority