

# 1 Scan Matching Performance - Architecture, Extended Tables & Figures

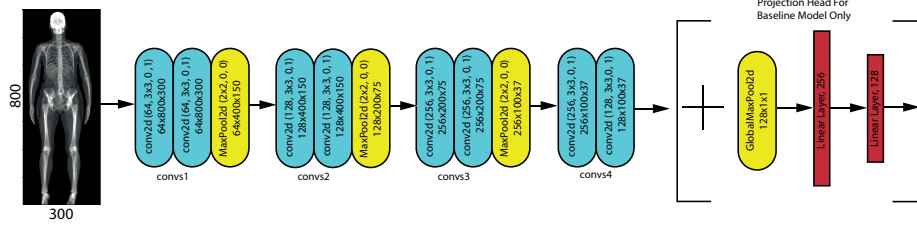


Fig. 1: The simple spatial encoder used in the contrastive framework. Both the MR and DXA spatial encoders use this architecture. For the baseline model, the output spatial features are max pooled and the projection head shown on the right is appended. Each convolutional & linear layer except the final one uses ReLU activations, followed by BatchNorm.

Input Scans	% Recall			AUC	Mean Rank	Equal Error Rate	TPR@ FPR=1%
	Top-1	Top-5	Top-10				
Bone DXA + Fat MRI	89.41	99.11	99.41	0.9992	2.106	0.0077	0.9955
Bone DXA + 2 MRI	87.72	98.66	99.36	<b>0.9993</b>	2.079	0.0084	0.9935
Tissue DXA + 2 MRI	83.12	97.03	98.42	0.9986	3.013	0.0114	0.9891
2 DXA + Fat MRI	85.84	97.57	98.66	0.9989	2.569	0.0098	0.9925
2 DXA + Water MRI	90.05	<b>99.21</b>	99.41	0.9993	<b>1.920</b>	0.0070	<b>0.9960</b>
2 DXA + 2 MRI	<b>90.69</b>	<b>99.21</b>	<b>99.46</b>	0.9992	2.526	<b>0.0060</b>	<b>0.9960</b>

Table 1: An extended table of performance metrics for varying scan-input in contrastive training including true positive rate at a false positive rate of 1% (TPR@FPR=1%) and top-5 recall.

Temperature, $\tau$	% Recall			AUC	Mean Rank	Equal Error Rate	TPR@ FPR=1%
	Top-1	Top-5	Top-10				
$\tau=0.1$	89.46	98.86	99.41	0.9991	2.148	0.0095	0.9921
$\tau=0.05$	91.14	99.16	99.45	0.9991	<b>2.140</b>	0.0080	0.9946
$\tau=0.01$	90.07	99.21	99.45	0.9992	2.527	0.0060	0.9960
$\tau=0.005$	<b>91.18</b>	<b>99.50</b>	<b>99.60</b>	<b>0.9994</b>	2.214	<b>0.0049</b>	<b>0.9975</b>
$\tau=0.001$	74.41	92.97	96.28	0.9979	3.534	0.0197	<b>0.9629</b>

Table 2: Scan-matching performance metrics for configurations with varying softmax temperature,  $\tau$ .

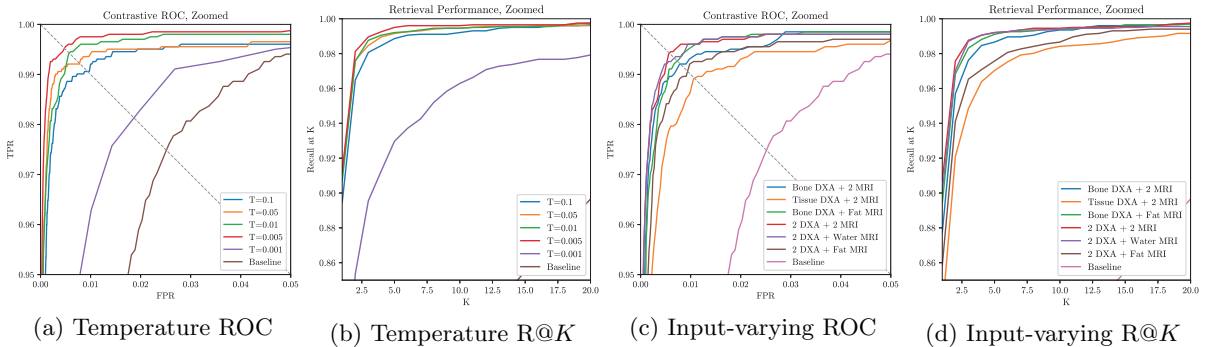


Fig. 2: ROC and Recall at  $K$  (R@K) curves for varying scan-input and temperature parameter  $\tau$ .

## 2 Unsupervised Registration

### 2.1 Lowe's Nearest Neighbours Ratio Test

1. For pixel in source feature map  $s_1$ , find the top-two most correlating pixels in the target feature map,  $t_1$  and  $t_2$  respectively.
2. If  $\tau \cdot \text{sim}(s_1, t_1) < \text{sim}(s_1, t_2)$  save the pair  $(s_1, t_1)$ , where  $\tau$  is some threshold between 0 and 1 and  $\text{sim}$  is the cosine similarity.
3. Repeat this for each pixel in the source feature map to obtain a set of candidate matches between the feature maps.
4. Apply RANSAC to remove spurious correlations from these candidates
5. Use LMEDS to get the best rigid transform between remaining inlying points.

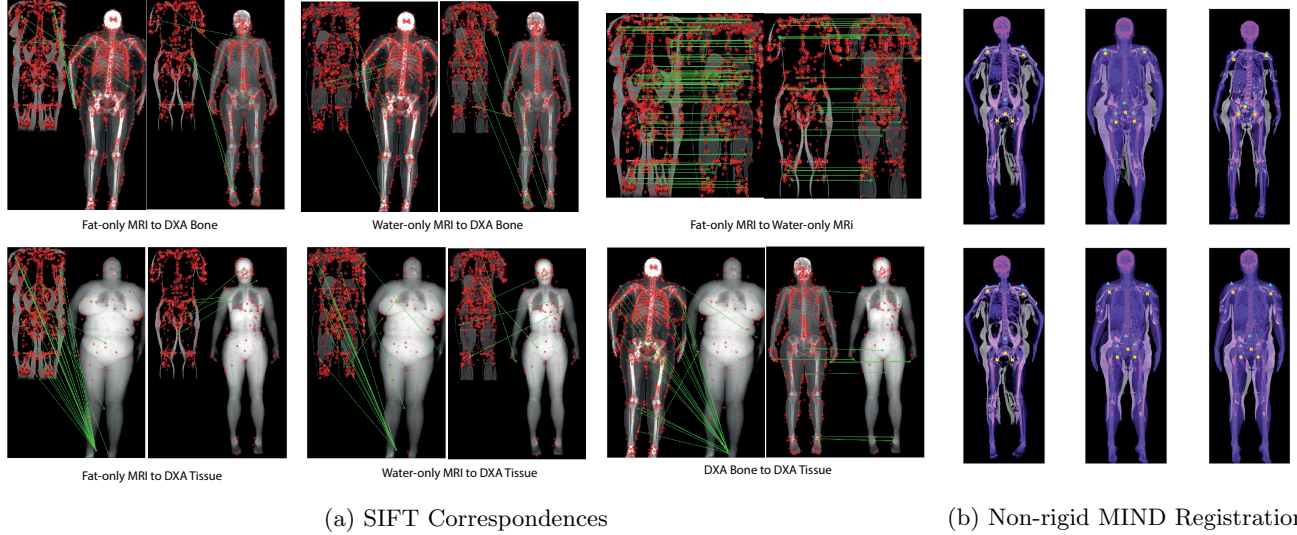


Fig. 3: **Attempted registration using SIFT & MIND features for the varying modalities.** a) SIFT features (shown by red circles) were calculated in both the original image and a negative version. They are then matched across modalities by brute-force matching and RANSAC is applied to find the best affine transform between the images. The in-lying matches are shown in green. This approach only succeeds finding correspondences between the already aligned MR sequences and to, some extent, the DXA images.

b) Results from Gauss-Newton optimised non-rigid MIND registration results as implemented at <https://github.com/cmifin/BBR>.

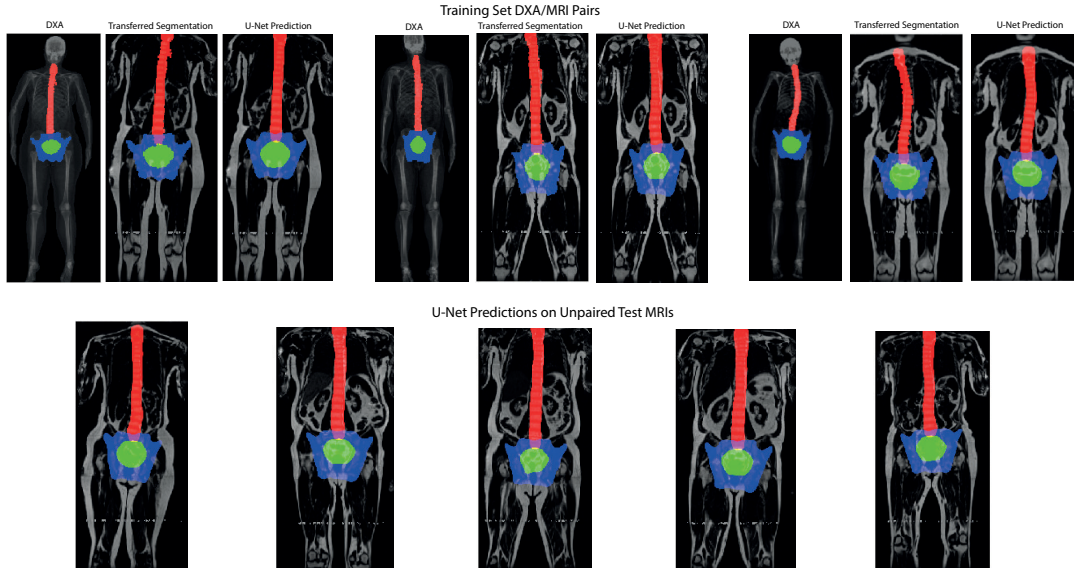


Fig. 4: **Predicted segmentations of the spine, pelvis and pelvic cavity in MR scans by a U-Net trained with DXA annotations.** Structures are segmented in DXA scans and transferred to the corresponding MR scan by the refinement registration method. A model is trained on the transferred segmentations which can then be applied to unpaired MR scans.