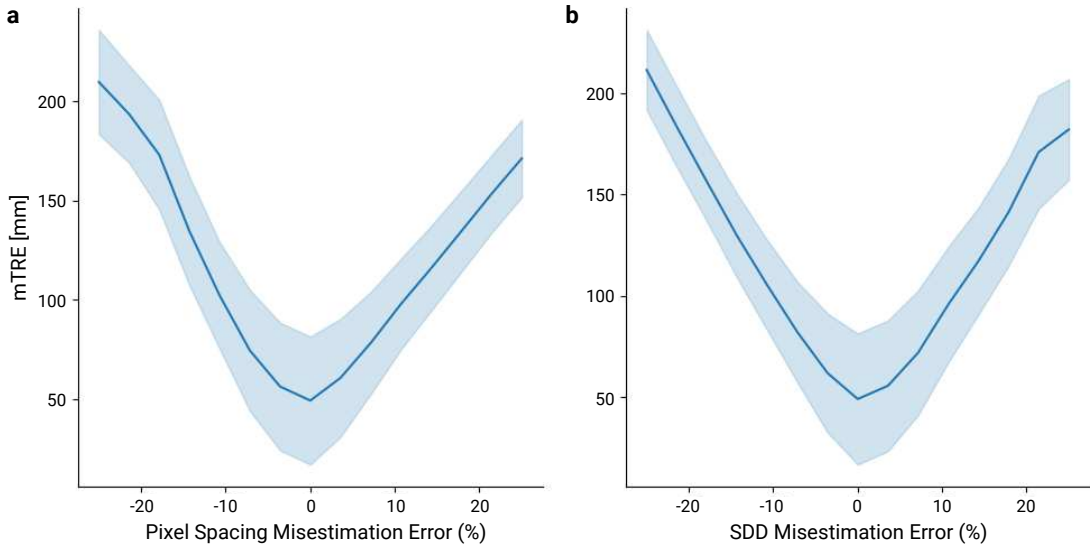

Supplementary information

**Rapid patient-specific neural networks for
X-ray to volume registration**

In the format provided by the
authors and unedited

A Sensitivity to C-arm intrinsics

2D/3D registration can be sensitive to the accuracy of C-arm intrinsic parameters. To quantify this sensitivity, we performed a simulation study in which we systematically perturbed either the pixel spacing or the source-to-detector distance (SDD) provided as input to the pose regression neural network. We evaluated our patient-specific models on the DeepFluoro dataset while varying both parameters across a range of $\pm 25\%$ error. For pixel spacing, the tested range was 0.1940 ± 0.0485 mm; for SDD, the tested range was 1020 ± 255 mm. As shown in Supplementary Fig. 1, misestimation of C-arm intrinsic parameters induces substantial errors in pose estimation. The model exhibits particularly high sensitivity to pixel spacing errors, where even a hundredth of a millimeter deviation increased the mTRE by tens of millimeters, with mTRE reaching approximately 50 mm at the optimal value and exceeding 175 mm at $\pm 25\%$ error (Supplementary Fig. 1a). In contrast, the model was more robust to absolute misestimation error in the SDD, producing comparably accurate pose estimates at 50 mm perturbations (Supplementary Fig. 1b). This finding is clinically relevant, as pixel spacing can typically be determined with high precision from the known specifications of the X-ray detector (sensor dimensions and pixel array size), whereas SDD may vary more substantially across different imaging configurations.



Supplementary Figure 1. Effect of misestimated intrinsics on initial pose estimation error. Mean target registration error (mTRE) as a function of systematic error in C-arm intrinsic parameters for patient-specific models evaluated on the DeepFluoro dataset. **a**, Pixel spacing misestimation across 0.1940 ± 0.0485 mm ($\pm 25\%$ of nominal value). **b**, Source-to-detector distance (SDD) misestimation across 1020 ± 255 mm ($\pm 25\%$ of nominal value). Shaded regions represent 95% confidence intervals.

B Additional baseline comparisons

To further contextualize the performance of our method, we compared **xvr** against the reported empirical evaluations of two previous 2D/3D registration methods: (i) Gao *et al.*¹⁹ train a neural network to approximate a convex image similarity function for iterative optimization, and (ii) Grimm *et al.*⁶ propose a semi-supervised method that transfers manually annotated keypoints from a training corpus to novel patients. Both papers evaluate their methods on real pelvic X-rays, with Gao *et al.*¹⁹ also using the DeepFluoro dataset. However, neither framework provides code, models, or data splits. Therefore, we emphasize that these are indirect comparisons requiring nuanced interpretation. We evaluated **xvr** using the same metrics as in the original publications (Supplementary Table 1).

	Transl. error [mm]	Rot. error [degrees]	mTRE [mm]	Transl. error ≤ 30 mm	mTRE ≤ 10 mm
Fixed init. (xReg) ²⁰	102.2 $\pm$ 78.1	32.7 $\pm$ 22.8	352.2 $\pm$ 239.3	22.9%	20.7%
Landmark reg. (xReg) ⁵	38.7 $\pm$ 79.3	5.5 $\pm$ 11.1	82.1 $\pm$ 162.0	75.1%	70.2%
Landmark init. (xReg) ⁵	28.6 $\pm$ 127.1	7.3 $\pm$ 22.6	79.6 $\pm$ 238.7	86.7%	80.1%
Fixed init. (xvr) ¹⁸	101.1 $\pm$ 66.5	26.3 $\pm$ 13.5	340.5 $\pm$ 170.1	15.3%	10.8%
Gao <i>et al.</i> , 2023 ^{19*}	33.9 $\pm$ 70.1	13.9 $\pm$ 30.5	45.0 $\pm$ 88.1	Not reported	73.2%
Grimm <i>et al.</i> , 2021 ^{6*}	4.2 $\pm$ 4.0	1.1 $\pm$ 1.3	4.2 $\pm$ 3.9	86.8%	Not reported
<i>De novo</i> model (xvr)	1.7 $\pm$ 8.1	0.2 $\pm$ 0.5	3.3 $\pm$ 10.1	99.4%	94.5%
Finetuned model (xvr)	1.5 $\pm$ 4.8	0.2 $\pm$ 0.8	3.2 $\pm$ 9.4	99.2%	94.8%

Supplementary Table 1. Benchmark evaluations of 2D/3D registration methods on pelvic X-ray datasets. All metrics report the mean and standard deviation. Asterisks (*) indicate results reported in the original publications on similar but not identical data, necessitating nuanced interpretation: Gao *et al.*¹⁹ also evaluate on the DeepFluoro dataset, but calculate mTRE using points on the surface of 3D bone segmentations, while Grimm *et al.*⁶ evaluate a different, non-public pelvic dataset.