Multiview 2D/3D Rigid Registration via a Point-Of-Interest Network for Tracking and Triangulation

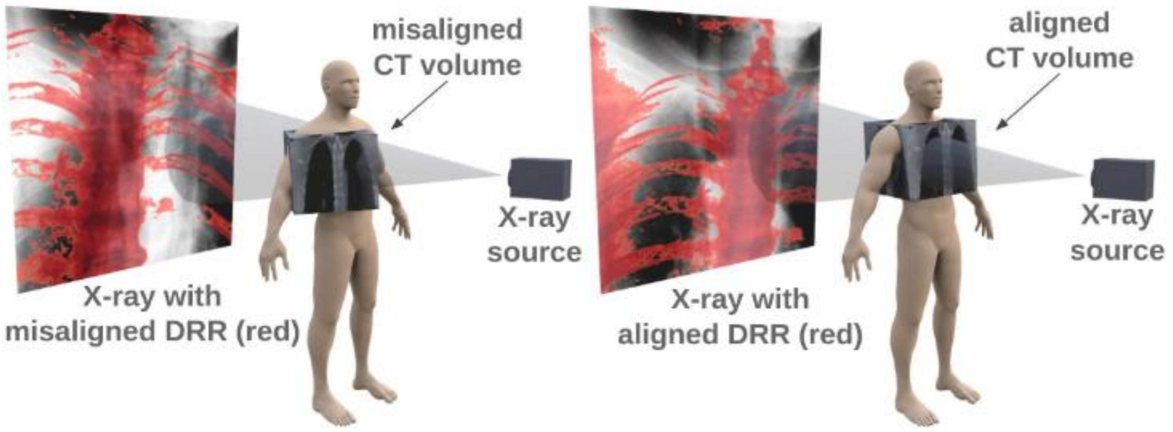
Haofu Liao[1]*, Wei-An Lin[2]*, Jiarui Zhang[3], Jingdan Zhang[4], Jiebo Luo[1], S. Kevin Zhou[5,6]
[1]University of Rochester [2]University of Maryland, College Park [3]Rutgers University
[4]Z2W Corporation [5]Chinese Academy of Sciences [6]Peng Cheng Laboratory, Shenzhen

Contents

1. Motivation and introduction
2. Architecture
3. Experiments

Motivation

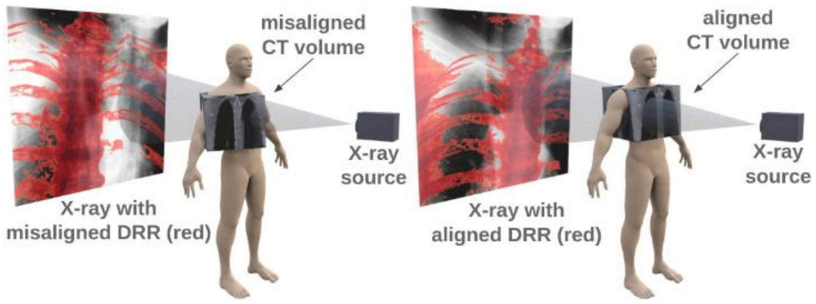
1. In 2D/3D rigid registration for intervention, the goal is to find a rigid pose of a pre-intervention 3D data, e.g., computed tomography (CT), such that it aligns with a 2D intra-intervention image of a patient, e.g., fluoroscopy.
2. To establish 2D point-to-point correspondences between the pre- and intra-intervention images by tracking a set of point- of-interests (POIs).



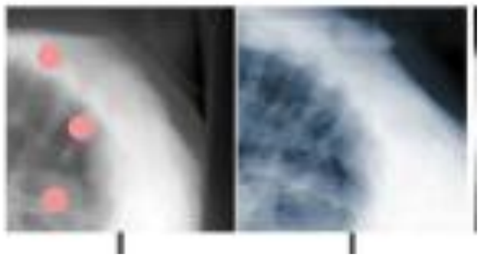
(a) before registration (b) after registration

Digitally Reconstructed Radiographs (DRRs)

1. Digitally reconstructed radiographs (DRRs) can be produced from CT using ray casting [21].
2. The generation of DRRs simulates how an X-ray is captured, which makes them visually similar to the X-rays.
3. are leveraged to facilitate the 2D/3D registration as we can observe the misalignment between the CT and patient by directly comparing the intra-intervention X-ray and the generated DRR.



(a) before registration (b) after registration



X-Ray DRR

Methodology

1. 3D data: a CT or CBCT volume, which is the most accessible and allows the generation of DRR.
2. 2D data: we use X-rays.

$$\mathbf{T} = \begin{bmatrix} \mathbf{R}(\boldsymbol{\theta}) & \mathbf{t} \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{t} = (t_x, t_y, t_z)^T$$

$$\boldsymbol{\theta} = (\theta_x, \theta_y, \theta_z)^T$$

$$\mathbf{I}^{\mathrm{D}}(\mathbf{x}) = \int_{\mathbf{p} \in \boldsymbol{l}(\mathbf{x})} \mathbf{V}(\mathbf{T}^{-1}\mathbf{p}) d\mathbf{p},$$

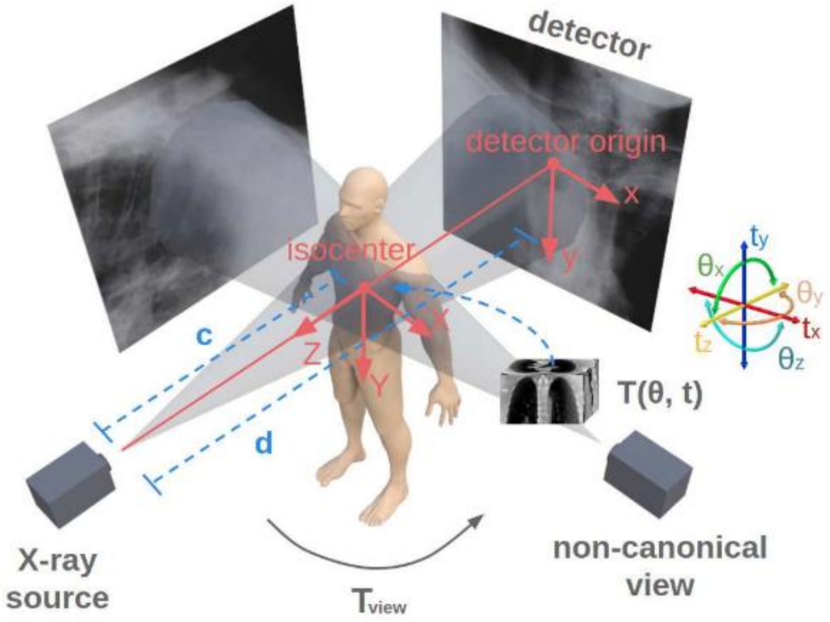


Figure 2: The X-ray imaging model of the canonical-view (bottom-left to upper-right) and a non-canonical view (bottom-right to upper-left).

Methodology

for a point $\mathbf{X} = (X, Y, Z)^T$ in the isocenter coordinate, its projection $\mathbf{x}$ on the detector is given by

$$\mathbf{x}' = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{h}\end{bmatrix}\begin{pmatrix}\mathbf{X}\\1\end{pmatrix},$$

where

$$\mathbf{K} = \begin{bmatrix}-d & 0 & 0\\0 & -d & 0\\0 & 0 & 1\end{bmatrix}, \mathbf{h} = \begin{pmatrix}0\\0\\-c\end{pmatrix}.$$

projection $$\mathbf{x} = (x, y) = (x'/z', y'/z').$$

$$\mathbf{x}' = \mathbf{K}\begin{bmatrix}\mathbf{R}_{\text{view}} & \mathbf{t}_{\text{view}} + \mathbf{h}\end{bmatrix}\begin{pmatrix}\mathbf{X}\\1\end{pmatrix}, \quad (4)$$

$$\mathbf{I}^{\text{D}}_{\text{view}}(\mathbf{x}) = \int_{\mathbf{p}\in\boldsymbol{l}(\mathbf{x})} \mathbf{V}(\mathbf{T}^{-1}\mathbf{T}^{-1}_{\text{view}}\mathbf{p})d\mathbf{p}. \quad (5)$$

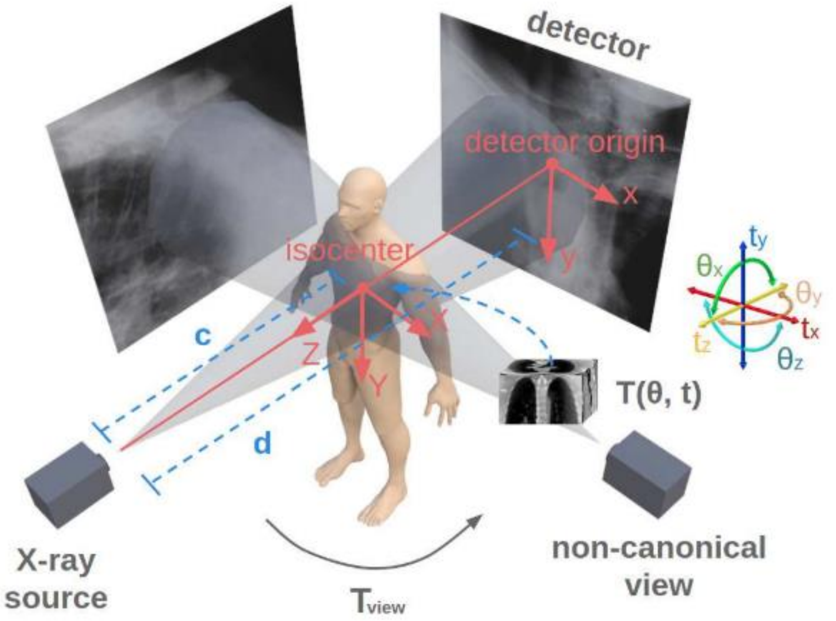


Figure 2: The X-ray imaging model of the canonical-view (bottom-left to upper-right) and a non-canonical view (bottom-right to upper-left).

Constructing a transformation matrix

1. Translation along 3 directions: tX, tY, tZ
2. Rotation along 3 directions: angleX, angleY, angleZ

$$\mathcal{R}_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix}$$

$$\mathcal{R}_y(\theta_y) = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix}$$

$$\mathcal{R}_z(\theta_z) = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathcal{M}(\alpha,\beta,\gamma) = \mathcal{R}_z(\gamma)\mathcal{R}_y(\beta)\mathcal{R}_x(\alpha)$$

$$\mathcal{M}(\alpha,\beta,\gamma) = \begin{bmatrix} \cos\alpha\cos\gamma - \cos\beta\sin\alpha\sin\gamma & -\cos\beta\cos\gamma\sin\alpha - \cos\alpha\sin\gamma & \sin\alpha\sin\beta \\ \cos\gamma\sin\alpha + \cos\alpha\cos\beta\sin\gamma & \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\sin\beta \\ \sin\beta\sin\gamma & \cos\gamma\sin\beta & \cos\beta \end{bmatrix}$$

$$\mathcal{M}(\alpha,\beta,\gamma) = \begin{bmatrix} \cos\alpha\cos\gamma - \cos\beta\sin\alpha\sin\gamma & -\cos\beta\cos\gamma\sin\alpha - \cos\alpha\sin\gamma & \sin\alpha\sin\beta \\ \cos\gamma\sin\alpha + \cos\alpha\cos\beta\sin\gamma & \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\sin\beta \\ \sin\beta\sin\gamma & \cos\gamma\sin\beta & \cos\beta \end{bmatrix}$$

$$\text{Matrix} = \begin{matrix} \cos\alpha\cos\gamma - \cos\beta\sin\alpha\sin\gamma & -\cos\beta\cos\gamma\sin\alpha - \cos\alpha\sin\gamma & \sin\alpha\sin\beta & tX \\ \cos\gamma\sin\alpha + \cos\alpha\cos\beta\sin\gamma & \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\sin\beta & tY \\ \sin\beta\sin\gamma & \cos\gamma\sin\beta & \cos\beta & tZ \\ 0 & 0 & 0 & 1 \end{matrix}$$

Case0005 groundtruth transformation

0.999800	0.018406	0.007847	14.714686
-0.019786	0.967802	0.250932	11.056741
-0.002976	-0.251037	0.967973	-11.534423
0.000000	0.000000	0.000000	1.000000

POINT

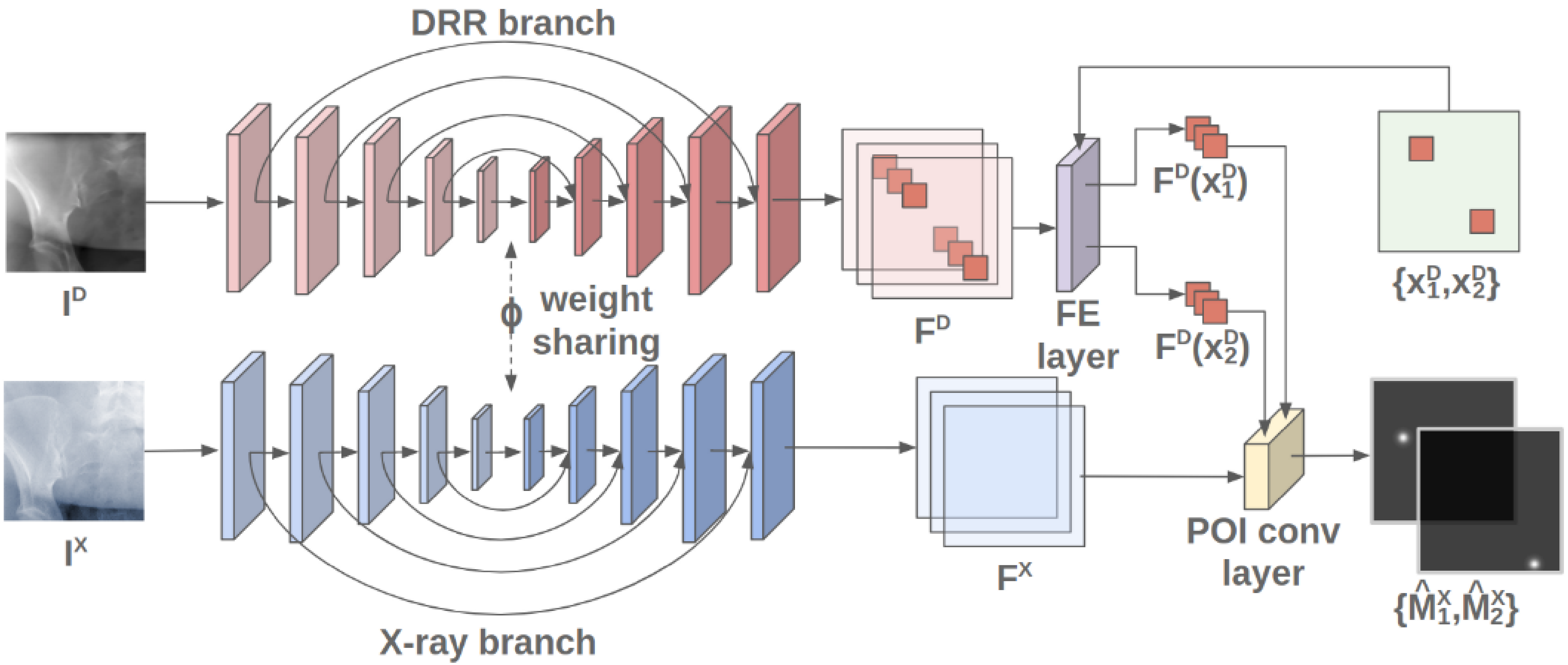


Figure 4: The architecture of the POINT network.

$$\hat{\mathbf{M}}_i^X = \mathbf{F}^X * (\mathbf{W} \odot \mathbf{F}^D(\mathbf{x}_i^D)),$$

POINT2

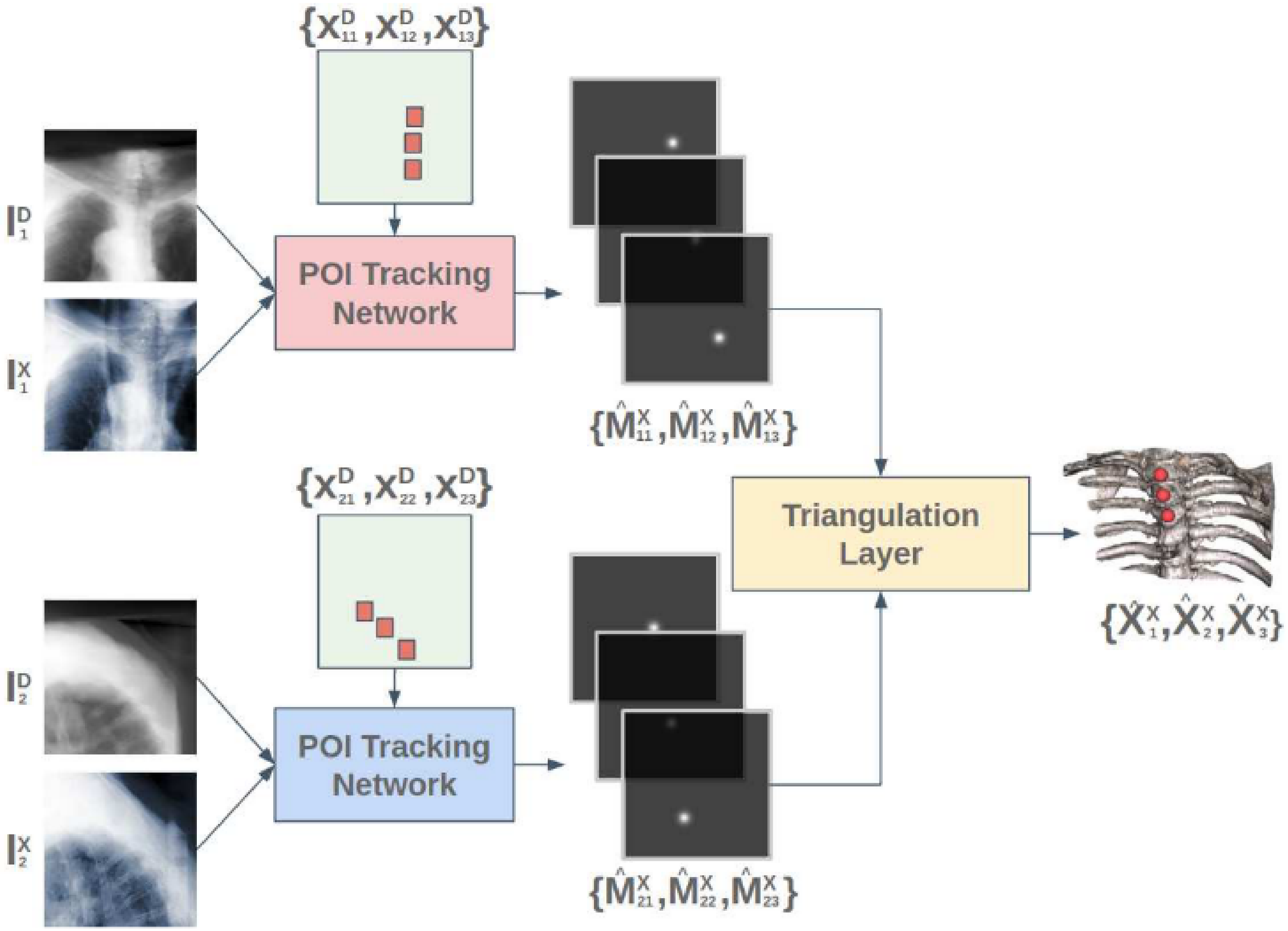


Figure 5: The overall framework of POINT^2.

Triangulation

from the i-th view, and we obtain the 2D X-ray POI by

$$\hat{\mathbf{x}}_{ij}^{\mathrm{X}} = \frac{1}{\sum_{\mathbf{x}} \hat{\mathbf{M}}_{ij}^{\mathrm{X}}(\mathbf{x})} \sum_{\mathbf{x}} \hat{\mathbf{M}}_{ij}^{\mathrm{X}}(\mathbf{x})\mathbf{x}. \tag{7}$$

Next, we rewrite Equation (4) as

$$\mathbf{D}(\mathbf{x})\mathbf{R}_{\text{view}}\mathbf{X} = c\mathbf{x} - \mathbf{D}(\mathbf{x})\mathbf{t}_{\text{view}}, \tag{8}$$

where

$$\mathbf{D}(\mathbf{x}) = \begin{bmatrix} d & 0 & \\ 0 & d & \mathbf{x} \end{bmatrix}.$$

Thus, by applying Equation (8) for each view, we can get

$$\begin{cases} \mathbf{D}(\hat{\mathbf{x}}_{1j}^{\mathrm{X}})\mathbf{R}_1\hat{\mathbf{X}}_j^{\mathrm{X}} & = c\hat{\mathbf{x}}_{1j}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{1j}^{\mathrm{X}})\mathbf{t}_1, \\ \mathbf{D}(\hat{\mathbf{x}}_{2j}^{\mathrm{X}})\mathbf{R}_2\hat{\mathbf{X}}_j^{\mathrm{X}} & = c\hat{\mathbf{x}}_{2j}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{2j}^{\mathrm{X}})\mathbf{t}_2, \\ & \vdots \\ \mathbf{D}(\hat{\mathbf{x}}_{nj}^{\mathrm{X}})\mathbf{R}_n\hat{\mathbf{X}}_j^{\mathrm{X}} & = c\hat{\mathbf{x}}_{nj}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{nj}^{\mathrm{X}})\mathbf{t}_n. \end{cases} \tag{9}$$

Let

$$\mathbf{A} = \begin{bmatrix} \mathbf{D}(\hat{\mathbf{x}}_{1j}^{\mathrm{X}})\mathbf{R}_1 \\ \mathbf{D}(\hat{\mathbf{x}}_{2j}^{\mathrm{X}})\mathbf{R}_2 \\ \vdots \\ \mathbf{D}(\hat{\mathbf{x}}_{nj}^{\mathrm{X}})\mathbf{R}_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} c\hat{\mathbf{x}}_{1j}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{1j}^{\mathrm{X}})\mathbf{t}_1 \\ c\hat{\mathbf{x}}_{2j}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{2j}^{\mathrm{X}})\mathbf{t}_2 \\ \vdots \\ c\hat{\mathbf{x}}_{nj}^{\mathrm{X}} - \mathbf{D}(\hat{\mathbf{x}}_{nj}^{\mathrm{X}})\mathbf{t}_n \end{bmatrix}, \tag{10}$$

then $\hat{\mathbf{X}}_j^{\mathrm{X}}$ is given by

$$\hat{\mathbf{X}}_j^{\mathrm{X}} = \mathbf{A}^{+}\mathbf{b}. \tag{11}$$

The triangulation can be plugged into a loss function that regulates the training of POINT networks of different views.

$$\mathcal{L} = \frac{1}{mn} \sum_i \sum_j \mathrm{BCE}(\sigma(\hat{\mathbf{M}}_{ij}^{\mathrm{X}}), \sigma(\mathbf{M}_{ij}^{\mathrm{X}})) + \frac{w}{n} \sum_j ||\hat{\mathbf{X}}_j^{\mathrm{X}} - \mathbf{X}_j^{\mathrm{X}}||_2, \tag{12}$$

Shape Alignment

Shape Alignment. Let $\mathbf{P}^{\mathrm{D}} = [\mathbf{X}_1^{\mathrm{D}}\ \mathbf{X}_2^{\mathrm{D}}\ \ldots\ \mathbf{X}_m^{\mathrm{D}}]$ be the selected CT POIs and $\mathbf{P}^{\mathrm{X}} = [\hat{\mathbf{X}}_1^{\mathrm{X}}\ \hat{\mathbf{X}}_2^{\mathrm{X}}\ \ldots\ \hat{\mathbf{X}}_m^{\mathrm{X}}]$ be the estimated 3D POIs [3]. The shape alignment finds a transformation matrix $\mathbf{T}^*$ such that the transformed $\mathbf{P}^{\mathrm{D}}$ aligns closely with $\mathbf{P}^{\mathrm{X}}$, i.e.,

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} ||\mathbf{T}\mathbf{P}^{\mathrm{D}} - \mathbf{P}^{\mathrm{X}}||_F, \text{ s.t., } \mathbf{R}\mathbf{R}^T = \mathbf{I} \qquad (13)$$

Ablation Study

① **POI selection**

② **POI convolution**

③ **Shift-invariant tracking**

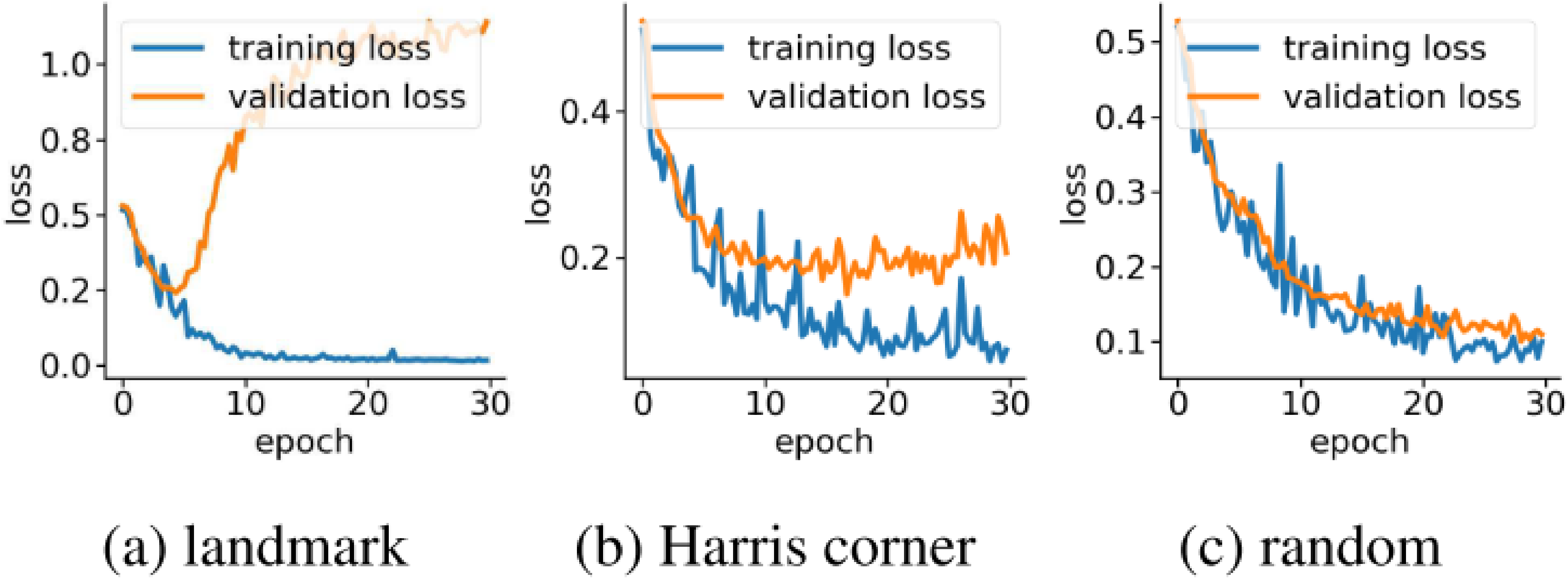


Figure 7: Training and validation losses of different POI selection methods.

Table 1: Ablation study of the proposed POINT network.

#	Kernel size			POI type			Weight		mPD (mm)
	1	3	5	land.	Harris	rand.	w/	w/o	
1	✓					✓	✓		8.46
2		✓				✓	✓		**8.12**
3			✓			✓	✓		9.49
4		✓			✓		✓		9.87
5		✓		✓			✓		12.72
6		✓				✓		✓	11.26

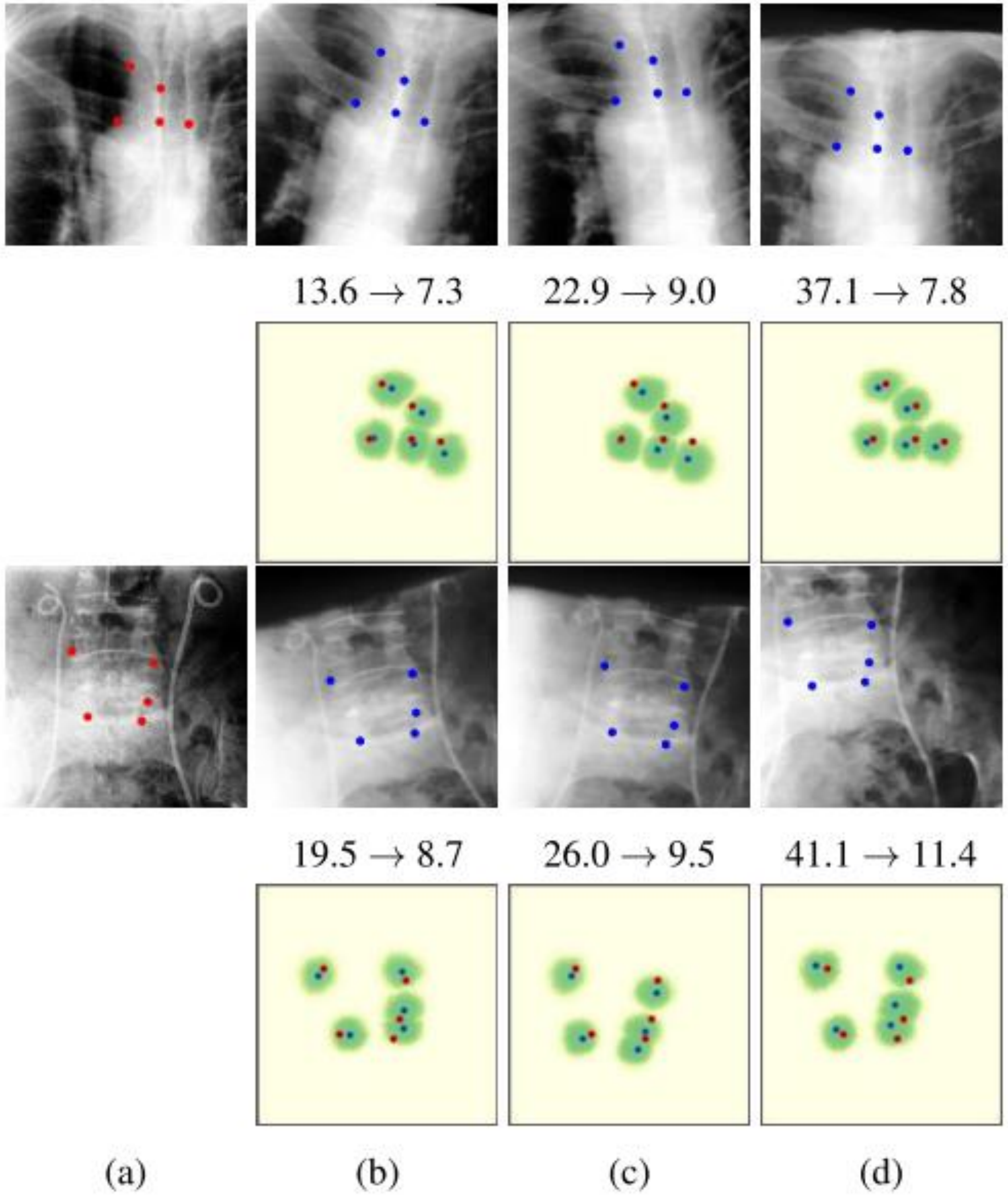


Figure 8: POI tracking results. (a) X-ray image. (b-d) DRR images with different in-plane offsets. The heatmaps of the tracking results are all aligned with the X-ray images and appear similar, showing the shift-invariant property.

Table 2: 2D/3D registration performance comparing with the state-of-the-art results.

	mTRE (mm)			GFR	Reg. time
	50th	75th	95th		
Initial	20.4	24.4	29.7	92.9%	N/A
Opt-NGI [16]	**0.62**	25.2	57.8	40.0%	23.5s
Opt-GO [4]	6.53	23.8	44.7	45.1%	22.8s
Opt-GC [4]	7.40	25.7	56.5	47.7%	22.1s
MDP [13]	5.40	8.62	27.6	<u>16.4%</u>	1.74s
POINT	5.63	<u>7.72</u>	<u>12.8</u>	18.6%	**0.75s**
POINT^2	<u>4.22</u>	**5.70**	**9.84**	**4.9%**	<u>0.78s</u>
MDP [13] + Opt	<u>1.06</u>	<u>2.25</u>	24.6	15.6%	3.21s
POINT + Opt	1.19	4.67	<u>21.8</u>	<u>14.8%</u>	**2.16s**
POINT^2 + Opt	**0.55**	**0.96**	**5.67**	**2.7%**	<u>2.25s</u>

Limitations

1) First, similar to other learning-based approaches, our method requires a considerably large dataset from the targeting medical domain for learning reliable feature representations. When the data is insufficient, the proposed method may fail.
2) Second, although our method alone is quite robust and its accuracy is state-of-the-art through a combination with the optimization-based approach, it is still desirable to come up with a more elegant solution to solve the problem directly.
3) Finally, due to the use of triangulation, our method requires X-rays from at least two views to be available. Hence, for the applications where only a sin- gle view is acceptable, our method will render an estimate of registration parameter with inherent ambiguity.