Toward Real-time 3D Vascular State Estimation from Monocular Fluoroscopy for Endovascular Robotic Surgery

Tushar Nayak*[1], Rishi Basdeo*[2], Edward G. Andrews[3], Jacob Biehl[4], Kenji Shimada[2]

[1] Department of Biomedical Engineering, Carnegie Mellon University, [2] Department of Mechanical Engineering, Carnegie Mellon University
[3] School of Medicine, University of Pittsburgh Medical Center, [4] School of Computing and Information, University of Pittsburgh

Clinical Need

The Challenge

Static Anatomical Roadmaps: Pre-operative CTA provides detailed 3D vascular anatomy, but guidewire and catheter interactions can deform vessels during intervention.

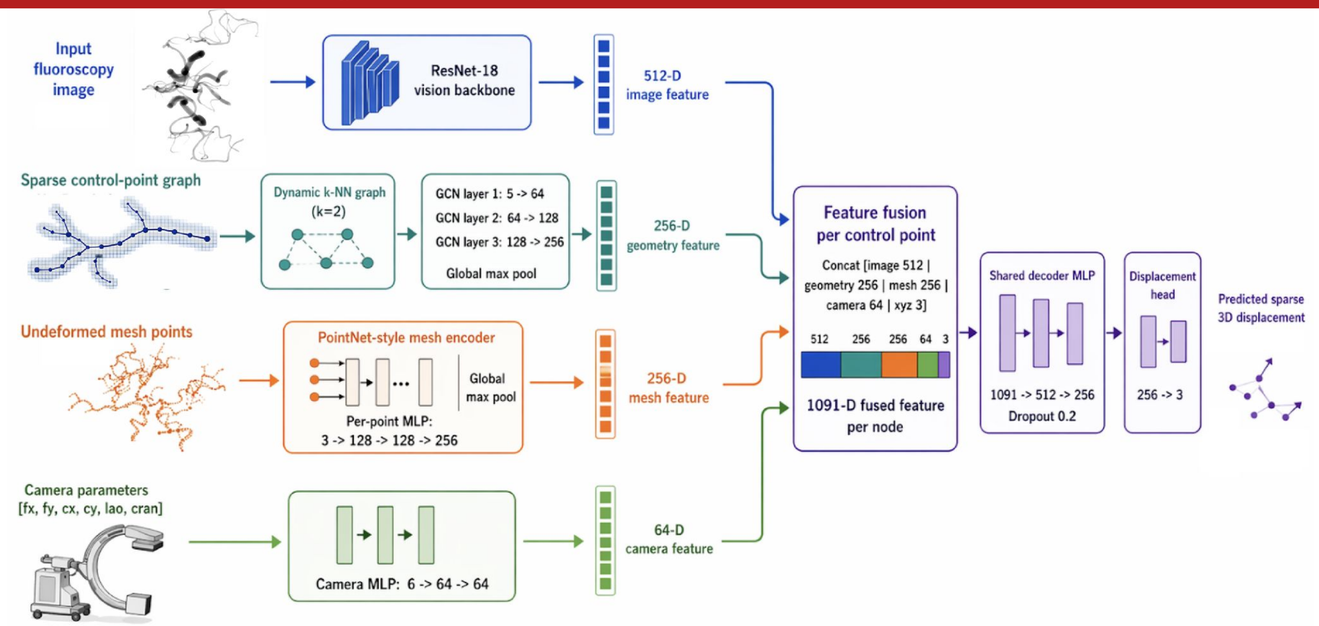
Limited intraoperative visibility: Monocular fluoroscopy provides only 2D projections, making out-of-plane deformation inherently ambiguous.

Real-time state estimation: Future robotic navigation systems require continuously updated 3D vascular geometry without computationally expensive biomechanical simulations.

Our Objective

Develop a real-time, geometry-guided framework that estimates sparse 3D vascular deformation from monocular fluoroscopy and preoperative vascular anatomy

Fluoroscopy Graph-Regularized Intraoperative Deformation Estimator (FluoroGRID) architecture



Regularized Loss Formulation with Geometric Priors

$$\mathcal{L} = \lambda_{\text{disp}}\mathcal{L}_{\text{disp}} + \lambda_{\text{proj}}\mathcal{L}_{\text{proj}} + \lambda_{\text{bend}}\mathcal{L}_{\text{bend}} + \lambda_{\text{stretch}}\mathcal{L}_{\text{stretch}} + \lambda_{\text{anchor}}\mathcal{L}_{\text{anchor}}$$

Displacement Supervision & Reprojection Consistency

$$\mathcal{L}_{\text{disp}} = \frac{1}{N}\sum_i \|\hat{x}_i - x_i^{gt}\|_2^2 \qquad \mathcal{L}_{proj} = \frac{1}{N}\sum_i \|\Pi(\hat{x}_i) - \Pi(x_i^{gt})\|_1$$

- $\mathcal{L}_{\text{disp}}$ is a SmoothL1 loss on the predicted displacement field, tells the network where each node should move in 3D.
- $\mathcal{L}_{\text{proj}}$ is the reprojection consistency term using the projected positions of the predicted and ground-truth geometry. So, it is not an image reconstruction loss.

Physics Regularization Terms

$$\mathcal{L}_{stretch} = \frac{1}{N-1}\sum_i \left(\|\hat{x}_{i+1} - \hat{x}_i\| - \|x_{i+1}^0 - x_i^0\|\right)^2 \quad \mathcal{L}_{bend} = \frac{1}{N-2}\sum_i \left(\Delta^2\hat{x}_i - \Delta^2 x_i^0\right)^2$$

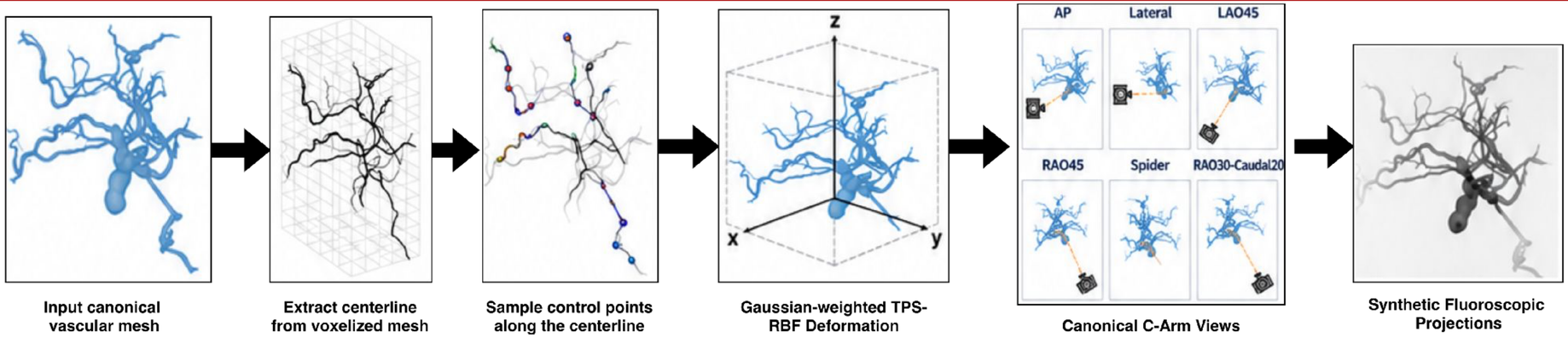
- $\mathcal{L}_{\text{bend}}$ is a curvature regularizer that computes a discrete second derivative, comparing predicted Laplacian to the rest-state Laplacian.
- $\mathcal{L}_{\text{stretch}}$ is a rest-length preservation regularizer that penalizes unrealistic elongation or compression.

Anchor Term

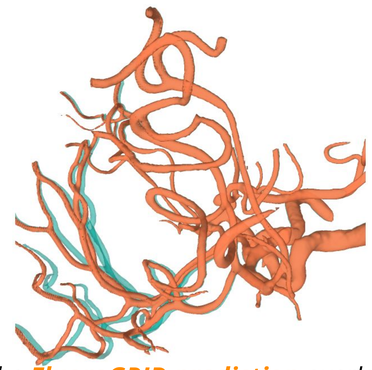
$$\mathcal{L}_{\text{anchor}} = \frac{1}{|A|}\sum_{i\in A}\left\|(\mathbf{x}_i^{rest} + \hat{\mathbf{d}}_i) - \mathbf{x}_i^{rest}\right\|^2$$

$\mathcal{L}_{\text{anchor}}$ is an anchor-point stability regularizer that penalizes displacement of designated anchor nodes from their undeformed state.

Synthetic Data Generation Pipeline



Preliminary Results



*The **FluoroGRID prediction** overlaid on the **ground-truth deformed vascular mesh** for a held-out AP-view sample, illustrating recovery of the global vascular configuration and major deformed branches.*

Metric	ConvNet	FluoroGRID
3D TRE (mm)	0.5242 (0.3051-0.6727)	**0.5184** (0.2978-0.6679)
In-Plane TRE (mm)	0.4034 (0.2081-0.5372)	**0.3956** (0.1994-0.5304)
Depth TRE (mm)	0.2790 (0.1076-0.3815)	**0.2774** (0.1055-0.3804)
CPU Latency	18.10 ms (~55 fps)	**15.02 ms (~67 fps)**
GPU Latency	**2.00 ms (~500 fps)**	2.91 ms (~343 fps)

Conclusion

- Monocular fluoroscopy can support sparse 3D vascular state estimation.
- FluoroGRID achieves sub-mm TRE with real-time inference.
- Structural and geometric priors help constrain implausible deformation estimation.

Future Work

- Move from weak geometric priors toward stronger mechanics-based residuals.
- Test temporal consistency across sequential fluoroscopy frames.
- Validate framework using live intraoperative fluoroscopy videos and pre-op roadmaps.

Acknowledgements

Supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE2140739. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.