This ACCV 2024 paper, provided here by the Computer Vision Foundation, is the author-created version.
The content of this paper is identical to the content of the officially published ACCV 2024
LNCS version of the paper as available on SpringerLink: https://link.springer.com/conference/accv

Guide3D: A Bi-planar X-ray Dataset for 3D Shape Reconstruction

Tudor Jianu[1], Baoru Huang[2], Hoan Nguyen[3], Binod Bhattarai[4], Tuong Do[1], Erman Tjiputra[5], Quang Tran[6], Pierre Berthet-Rayne[1], Ngan Le[7], Sebastiano Fichera[1], and Anh Nguyen[1]

[1] University of Liverpool, Liverpool, United Kingdom
{t.jianu,sebastiano.fichera,anh.nguyen}@liverpool.ac.uk
[2] Imperial College London
[3] University of Information Technology - VNUHCM
[4] University of Aberdeen
[5] AIOZ Singapore
[6] University of Arkansas llf

Abstract. Endovascular surgical tool reconstruction represents an important factor in advancing endovascular tool navigation, which is an important step in endovascular surgery. However, the lack of publicly available datasets significantly restricts the development and validation of novel machine learning approaches. Moreover, due to the need for specialized equipment such as biplanar scanners, most of the previous research employs monoplanar fluoroscopic technologies, hence only capturing the data from a single view and significantly limiting the reconstruction accuracy. To bridge this gap, we introduce Guide3D, a bi-planar X-ray dataset for 3D reconstruction. The dataset represents a collection of high resolution bi-planar, manually annotated fluoroscopic videos, captured in real-world settings. Validating our dataset within a simulated environment reflective of clinical settings confirms its applicability for real-world applications. Furthermore, we propose a new benchmark for guidewrite shape prediction, serving as a strong baseline for future work. Guide3D not only addresses an essential need by offering a platform for advancing segmentation and 3D reconstruction techniques but also aids the development of more accurate and efficient endovascular surgery interventions. Our code and dataset will be made publicly available to encourage further studies.

Keywords: Endovascular Dataset · 3D Reconstruction

1 Introduction

Minimally invasive surgery has revolutionized endovascular intervention, offering less intrusive options with expedited recovery periods [27]. These procedures' success relies on the precise navigation and manipulation of instruments such as *guidewires* and *catheters*. The procedure primarily relies on 2D visualization methods for guidance, with *monoplanar fluoroscopy* being the most prominent

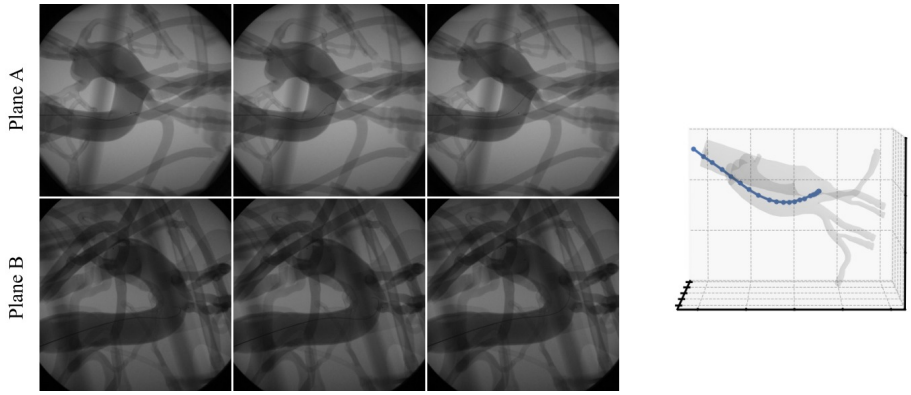


Fig. 1: Dataset Overview: Guide3D contains 8,746 manually annotated frames from two views for 3D reconstruction (left), from which the reconstruction is derived (right).

due to its minimal interference with surgical workflows and reasonable financial cost setup [28]. However, despite their widespread use, these conventional imaging techniques present significant limitations. Among these, one primary issue is depth perception, which poses challenges in accurately visualizing surgical instruments [18,19]. This limitation increases the risk of excessive contact between the surgical instruments and arterial walls, potentially endangering patient safety and diminishing the procedure's efficacy.

In endovascular intervention, depth perception crucially relies on multi-view imaging systems, such as biplanar scanners, which enable shape reconstruction by integrating multiple angular views [2,8,19,35] and employing epipolar geometry based reconstruction [5]. However, two main challenges impede the wider adoption and efficacy of these systems [29]: *i)* the difficulty of accurately segmenting images for successful shape reconstruction, increased by the lack of datasets to evaluate segmentation methods, and *ii)* the requirement for specialized biplanar scanners, which are not widely available in clinical settings due to expensive financial cost [8]. These issues highlight the *critical need for comprehensive datasets* that can improve segmentation algorithm precision and guidewire reconstruction methodologies, accelerating the development of more versatile imaging technologies.

In this paper, we introduce **Guide3D**, a dataset tailored for advancing *3D reconstruction* within endovascular navigation. Guide3D establishes a standardized platform for the development and evaluation of algorithms. By providing a uniform and extensive dataset, complete with *manual annotations* for segmentation and tools for effective 3D visualization, Guide3D is designed to develop innovation and enhancement in endovascular intervention. Additionally, the inclusion of *video-based bi-planar* fluoroscopic data enables the exploration of temporal dynamics, such as employing optical flow networks [25]. Guide3D aims to bridge the gap between research innovations and clinical application, targeting essential challenges in endovascular procedures.

2 Related Work

Endovascular Datasets. Datasets are pivotal for the progression of endovascular navigation, serving as a crucial resource for developing, evaluating, and enhancing algorithms. These datasets, sourced from modalities like mono X-ray [2, 4, 25, 37], 3D Echo [36], and 3D MRI [24], encompass both real and synthetic imagery. However, the limited availability of comprehensive, public datasets for tool segmentation and 3D reconstruction remains a considerable challenge [2, 4, 12, 24, 25]. Table 1 reveals the predominance of mono X-ray datasets for endovascular interventions, which lack the necessary detail for precise 3D reconstruction in previous works.

Catheter and Guidewire Segmentation. The segmentation of endovascular tools, notably guidewires and catheters, represents an advancing field contingent upon dataset availability and quality. Table 1 illustrates that prior studies have employed synthetic and semi-synthetic data to mitigate the challenges of limited real-world datasets. Works by Barbu *et al.* [4], Ambrosini *et al.* [2], and Mastmeyer *et al.* [24] have utilized manually annotated datasets from 2D X-ray and 3D MRI modalities for training segmentation models. Conversely, Nguyen *et al.* [25] and Danilov *et al.* [12] have demonstrated the utility of synthetic datasets in enhancing model efficiency. The introduction of deep learning techniques, particularly U-Net architectures [30], has substantially improved the precision in segmentation and tracking of these surgical instruments. This progress enables fully automated segmentation frameworks that leverage extensively annotated data and employ unsupervised techniques, such as optical flow. Nonetheless, the lack of a public, standardized dataset for method comparison continues to hinder the progression and evaluation of scientific advancements in this area.

3D Reconstruction. Enhancing 3D reconstruction accuracy in endovascular procedures contributes to superior clinical outcomes by facilitating catheter navigation via advanced visualization and precise tracking. Progress in fluoroscopic imaging technology has enabled more accurate positioning of devices. Wagner *et al.* [35] developed an algorithm that leverages elastic grid registration and epipolar geometry for 3D reconstruction from biplane angiography. Hoffmann *et al.* [17–19] employed automatic catheter detection and 3D reconstruction methods utilizing triangulation and graph-search algorithms in electrophysiology studies. Delmas *et al.* [13] and Petković *et al.* [26] demonstrated the significance of accurate 3D models for navigation in both complex and single-view vascular architectures, emphasizing the utility of biplanar data. However, the scarcity of comprehensive, publicly accessible datasets for algorithm development and validation in 3D reconstruction represents a significant barrier to both technological advancement and clinical application. This situation highlights the essential need for specialized datasets to foster continued innovation in endovascular tool reconstruction.

Dataset	Data Collection	Data Type	#Frames	Data Source	Annotation	Public	Task
Barbu et al. [4]	Mono X-ray	Video	535	Real	Manual	No	Segmentation
Wu et al. [36]	3D Echo	Video	800	Real	Manual	No	
Ambrosini et al. [2]	Mono X-ray	Image	948	Real	Manual	No	
Mastmeyer et al. [24]	3D MRI	Image	101	Real	Manual	No	
Yi et al. [37]	Mono X-ray	Image	2,540	Synthesis	Automatic	No	
Nguyen et al. [25]	Mono X-ray	Image	25,271	Phantom	Semi-Auto	No	
Danilov et al. [12]	3D Ultrasound	Video	225	Synthetic	Manual	No	
Wagner et al. [35]	Mono X-ray	Video	-	Phantom	Automatic	No	3D Reconstruction
Hoffmann et al. [19]	Mono X-ray	Image	176	Phantom	Semi-Auto	No	
Delmas et al. [13]	Mono X-ray	Image	2,289 5 63	Simulated Phantom Clinical	Automatic	No	
Baur et al. [5]	Bi-planar X-ray	Image	70	Canine	Manual	No	
Brost et al. [7]	Mono X-ray	Image	938	Clinical	Semi-Auto	No	
Ma et al. [22]	Mono X-ray, CT	Image	1,048	Clinical	Manual	No	
Hoffman et al. [17]	Bi-planar X-Ray	Video	33	Clinical	Semi-Auto	No	
Guide3D (ours)	Bi-planar X-ray	Video	8,746	Phantom	Manual	Yes	Segmentation 3D Reconstruction

Table 1: Endovascular intervention datasets comparison.

3 The Guide3D Dataset

3.1 Data Collection Setup

X-ray System. Our setup utilized a Bi-planar X-ray system (Fig. 2a) with 60kW Epsilon X-ray Generators from EMD Technologies Ltd. and 16-inch Image Intensifier Tubes by Thales, featuring dual focal spot Varian X-ray tubes for high-definition imaging. The setup included Ralco Automatic Collimators for precise alignment and exposure, with calibration achieved through acrylic mirrors and geometric alignment grids.

Anatomical Models. Our study utilizes the half-body vascular phantom model from Elastrat Sarl Ltd., Switzerland (Fig. 2a). This model is enclosed in a transparent box and integrated into a closed water circuit to simulate blood flow. Constructed from soft silicone and incorporating compact continuous flow pumps with a slippery liquid, it replicates human blood flow dynamics. The design is based on detailed postmortem vascular casts, ensuring anatomical accuracy reflective of human vasculature as documented in seminal works [16, 23]. Our utilization of this anatomically accurate model facilitates the realistic simulation of vascular scenarios.

Surgical Tools. We enhance our dataset by navigating complex vascular structures using two distinct types of guidewires that are widely used in real-world endovascular surgery. The first, the Radifocus™ Guide Wire M Stiff Type (Terumo Ltd.) (Fig. 2c), is made from nitinol and coated with polyurethane-tungsten. It has a diameter of 0.89mm and a length of 260cm, featuring a 3cm angled tip. This guidewire is designed for seeking, dissecting, and crossing lesions. The second, the Nitrex Guidewire (Nitrex Metal Inc.) (Fig. 2b), also made of niti-

(a) (b) (c)

Fig. 2: Materials: *a)* Overall setup & endovascular phantom, *b)* Radifocus (angled) guidewire. and *c)* Nitrex (straight) guidewire.

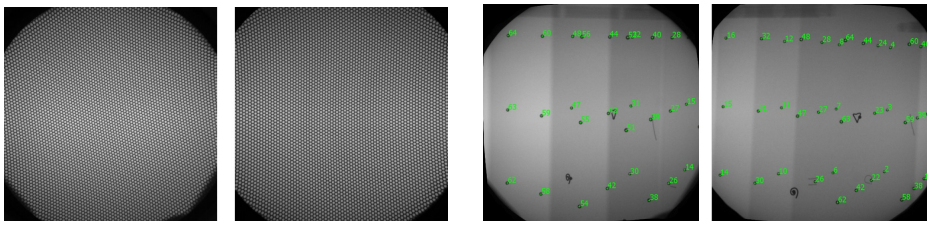
Sample Type	Radifocus™ Guide Wire (Angled)	Nitrex Guidewire (Straight)	Total
w fluid	3,664	484	4,148
w/o fluid	2,472	2,126	4,598
Total	6,136	2,610	8,746

Table 2: Dataset Composition Overview.

nol, has a gold-tungsten straight tip for enhanced radiopacity within fluoroscopic visualization. It measures 0.89mm in diameter and 400cm in length with a 15cm tip. Unlike the Radifocus guidewire, the Nitrex guidewire is generally used for accessing or maintaining position while exchanging catheters. Both guidewires were selected for their relevance to the operation in real-world settings and to maximize the diversity of the data in our dataset.

3.2 Data Acquisition, Labeling, and Statistics

Utilizing the materials described in Subsection 3.1, we compiled a dataset of 8,746 high-resolution samples ($1,024 \times 1,024$ pixels). This dataset includes 4,373 paired instances with and without a simulated blood flow medium. Specifically, it comprises 6,136 samples from the Radifocus Guidewire and 2,610 from the Nitrex guidewire, establishing a robust foundation for automated guidewire tracking in bi-planar scanner images. Manual annotation was performed using the Computer Vision Annotation Tool (CVAT) [11], where polylines were created to accurately track the dynamic path of the guidewire. The decision to represent the guidewire as a polyline was due to the inherent structure of a guidewire, where



(a) undistortion **(b)** calibration

Fig. 3: Fluoroscopic Calibration: *a)* Undistortion grid application, and *b)* Point identification on calibration frame.

certain parts would inevitably overlap, making a segmentation mask an unsuitable representation. In contrast, a polyline can effectively represent a looping guidewire, providing better accuracy. As detailed in Table 2, the dataset includes 3,664 instances of angled guidewires with fluid and 484 without, while straight guidewires are represented by 2,472 instances with fluid and 2,126 without. This distribution illustrates a variety of procedural contexts. We note that all 8,746 images in our dataset are equipped with *manual segmentation ground truth*, hence supporting the development of algorithms that need segmentation maps as the reference.

3.3 Calibration

We extract the camera parameters by following the traditional undistortion and calibration method. Initially, undistortion is achieved using a local weighted mean (LWM) algorithm, applying a perforated steel sheet with a hexagonal pattern (from McMaster-Carr Ltd., part number 9255T645) as framing reference, and employing a blob detection algorithm for precise identification of distortion points [6]. This setup establishes correspondences between distorted and undistorted positions, enabling accurate distortion correction as per Verdonck *et al.* [34]. A subsequent calibration step is undertaken utilizing a semi-automatic approach for marker identification and a random sampling consensus (RANSAC) method to ensure robustness in computing the projection matrix and deriving intrinsic and extrinsic camera parameters. The calibration process is refined through direct linear transformation (DLT) and non-linear optimization, considering multiple poses of the calibration object to optimize the entire camera setup [38]. Figure 3 shows the calibration process.

3.4 Guidewire Reconstruction

Given polyline representations of a curve in both planes, we initiate the reconstruction process by parameterizing these curves using B-Spline interpolation. This is achieved by expressing each curve as a function of the cumulative distance along the curve. Specifically, let $\mathbf{C}_A(\mathbf{u}_A)$ and $\mathbf{C}_B(\mathbf{u}_B)$ be the parameterized B-Spline curves in their respective planes, where $\mathbf{u}_A$ and $\mathbf{u}_B$ are the normalized

arc-length parameters. Next we find the corresponding $\mathbf{u}_B$ for a given $\mathbf{u}_A$ using epipolar geometry. Once the corresponding points $\mathbf{C}_A(\mathbf{u}_A^i)$ and $\mathbf{C}_B(\mathbf{u}_B^i)$ are identified, the 3D coordinates $\mathbf{P}^i$ of these points are computed by triangulation. This results in a set of 3D points $\{\mathbf{P}^i\}_{i=1}^M$, where M is the total number of sampled points, effectively reconstructing the original curve in 3D space.

Retrieving the Fundamental Matrix F. The relationship between the points in Image A ($\mathbf{I}_A$) and Image B ($\mathbf{I}_B$) is described by the fundamental matrix $\mathbf{F}$, which satisfies the condition $\mathbf{x}_B^T\mathbf{F}\mathbf{x}_A = 0$ for corresponding points $\mathbf{x}_A$ in $\mathbf{I}_A$ and $\mathbf{x}_B$ in $\mathbf{I}_B$. Given the calibration steps undertaken in Subsection 3.3, we now have the projection matrices $\mathbf{P}_A$ and $\mathbf{P}_B$. From these projection matrices, the fundamental matrix can be derived as follows:

$$\mathbf{F} = [\mathbf{e}_B]_\times \mathbf{P}_B \mathbf{P}_A^+ \tag{1}$$

where $\mathbf{e}_B$ is the epipole in Image B, defined as $\mathbf{e}_B = \mathbf{P}_B\mathbf{C}_A$, with $\mathbf{C}_A$ being the camera center of $\mathbf{P}_A$. The notation $[\mathbf{e}_B]_\times$ represents the skew-symmetric matrix of the epipole $\mathbf{e}_B$, which is given by:

$$[\mathbf{e}_B]_\times = \begin{bmatrix} 0 & -e_{B3} & e_{B2} \\ e_{B3} & 0 & -e_{B1} \\ -e_{B2} & e_{B1} & 0 \end{bmatrix} \tag{2}$$

Here, $\mathbf{e}_B = (e_{B1}, e_{B2}, e_{B3})^T$. The matrix $\mathbf{P}_A^+$ is the pseudoinverse of the projection matrix $\mathbf{P}_A$. The fundamental matrix $\mathbf{F}$ encapsulates the epipolar geometry between the two views, ensuring that corresponding points $\mathbf{x}_A$ and $\mathbf{x}_B$ lie on their respective epipolar lines.

Matching The matching phase begins with uniformly sampling points along the curve $\mathbf{C}_A(u_A)$ at intervals Δu_A. For each sampled point $x_A = \mathbf{C}_A(u_A)$, we project the epiline $l_B = Fx_A$ into Image B. We then determine the intersection of the epiline l_B with the curve $\mathbf{C}_B(u_B)$, thereby obtaining the corresponding parameter u_B for each u_A.

Due to errors in the projection matrices P_A and P_B, there are instances where the epiline l_B does not intersect with any part of the curve $\mathbf{C}_B$. To address this, similar to the approach in [1], we fit a monotonic function $f_A(u_A) \to u_B$ using a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP), thus interpolating the missing intersections. The matching can be visualized in Fig. 4.

Triangulation Given the projections from both the first and second cameras, we can construct a system of linear equations as:

$$\begin{aligned} \mathbf{x}_1 \times (\mathbf{P}_1\mathbf{X}) &= 0 \\ \mathbf{x}_2 \times (\mathbf{P}_2\mathbf{X}) &= 0 \end{aligned} \tag{3}$$

We therefore form a system of linear equations of the form $\mathbf{A}\mathbf{X} = 0$, wherein $\mathbf{A}$ is a 4×4 matrix composed of rows from the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ and the image coordinates $\mathbf{x}_A$ and $\mathbf{x}_B$. By employing the Singular Value Decomposition (SVD) [21] on $\mathbf{A}$, the desired $\mathbf{X}$ minimizing $||\mathbf{A}\mathbf{X}||$ subject to $||\mathbf{X}|| = 1$ is determined. The solution corresponds to the eigenvector associated with $\mathbf{A}$'s

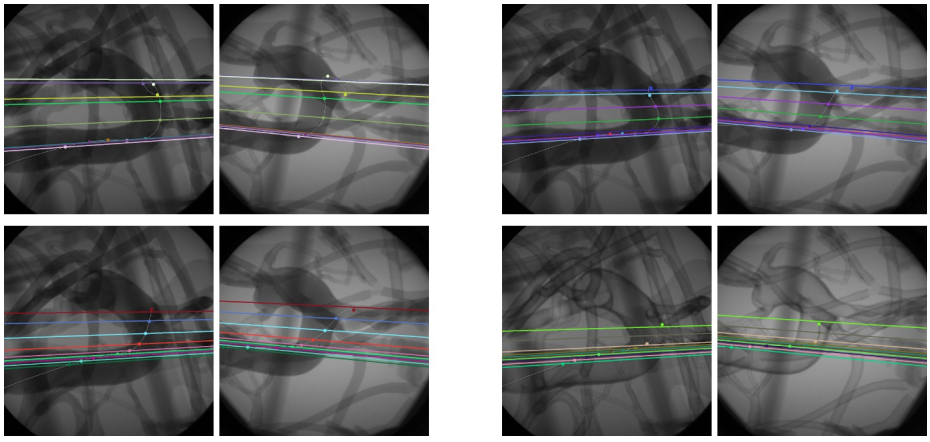


Fig. 4: Point Matching Process: Sampled points from image I_A ($C_A(u_A)$) and their corresponding epilines l_A on image I_B with their matches $C_B(u_B)$. The epilines for $C_B(u_B)$ are then computed and displayed on the image I_A.

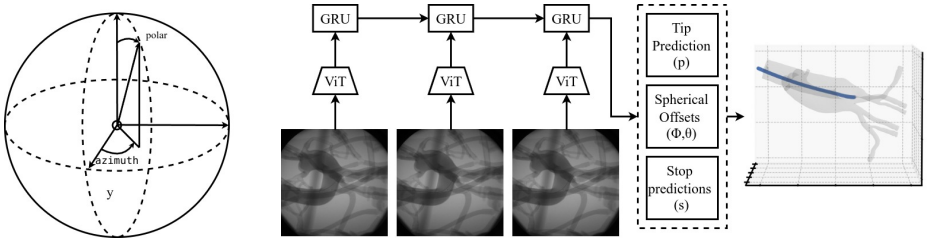
smallest singular value, thereby deriving the homogeneous coordinates for the triangulated 3D point. Finally, we perform a manual check of all reprojections to identify any degenerate curves. This step is crucial for addressing situations where partial occlusion of the guidewire or suboptimal acquisition angles result in either incorrect or missing corresponding points for the points derived. Samples with inaccurate reconstruction results are manually discarded.

3.5 Utility of Guide3D Dataset for the Research Community

Guide3D advances endovascular imaging by offering a bi-planar fluoroscopic dataset for *segmentation* and *3D reconstruction*, serving as an open-source benchmark. It supports precise algorithm comparisons for segmentation [2, 25, 32, 39] and facilitates method development in 3D reconstruction with its use of bi-planar imagery [3, 5, 7, 22]. With video data, Guide3D enables *video*-based methods, leveraging temporal dimensions for dynamic analysis, which enriches the segmentation and reconstruction capabilities whilst also adhering to the nature of the procedure. This versatility underscores Guide3D's crucial role in advancing endovascular imaging [29].

4 Guidewire Shape Prediction

Utilizing the Guide3D dataset, in this section, we build a benchmark for the shape prediction task, which is a popular task in endovascular intervention [25]. Accurate shape prediction of the guidewire is essential for successful navigation and intervention. In particular, we present a novel shape prediction network aimed at predicting the guidewire shape from a sequence of monoplanar images. The motivation behind this approach lies in the potential of deep learning to learn spatio-temporal correlations from a static camera with a dynamic



(a) Spherical Representation

(b) Network Architecture

Fig. 5: Network Key Components: The figure illustrates essential components of the proposed model. *a)* Spherical coordinates (r, θ, ϕ) used for predicting the guidewire shape. *b)* The model predicts the 3D shape of a guidewire from image sequences $\mathbf{I}_t$. A Vision Transformer (ViT) extracts spatial features $\mathbf{z}_t$, which a Gated Recurrent Unit (GRU) processes to capture temporal dependencies, producing hidden states $\mathbf{h}_t$. The final hidden state drives three prediction heads: the Tip Prediction Head for 3D tip position $\mathbf{p} \in \mathbb{R}^3$, the Spherical Offset Prediction Head for coordinate offsets $(\Delta\phi, \Delta\theta)$, and the Stop Prediction Head for terminal point probability $\mathbf{S}$.

scene [31]. Unlike the conventional reconstruction methods requiring biplanar images [8], our network leverages a sequence of images to provide temporal information, enabling the network to learn a mapping from a single image $\mathbf{I}_A$ to the 3D guidewire curve $\mathbf{C}(\mathbf{u})$. By adopting deep learning techniques, we aim to simplify the shape prediction process while maintaining high accuracy. This method holds promise for enhancing endovascular navigation by providing real-time, accurate predictions of the guidewire shape, ultimately improving procedural outcomes and reducing reliance on specialized equipment.

4.1 Spherical Coordinates Representation

Predicting 3D points directly can be challenging due to the high degree of freedom. To mitigate this, we use spherical coordinates, which offer significant advantages over Cartesian coordinates for guidewire shape prediction. Spherical coordinates, as represented in Fig. 5a, defined by the radius r, polar angle θ, and azimuthal angle ϕ, provide a more natural representation for the position and orientation of points along the guidewire, which is typically elongated and curved. Mathematically, a point in spherical coordinates (r, θ, ϕ) can be converted to Cartesian coordinates (x, y, z) using the transformations $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, and $z = r\cos\theta$. This conversion simplifies the modeling of angular displacements and rotations, as spherical coordinates directly encode directional information.

Predicting angular displacements $(\Delta\theta, \Delta\phi)$ relative to a known radius r aligns with the physical constraints of the guidewire, facilitating more accurate and interpretable shape predictions. By predicting an initial point (tip position) and representing subsequent points as offsets in $\Delta\phi$ and $\Delta\theta$ while keeping r fixed, this method simplifies shape comparison and reduces the parameter space. This

approach enhances the model's ability to capture the guidewire's spatial configuration and improves overall prediction performance.

4.2 Network Architecture

The proposed model (shown in Fig. 5b) addresses the problem of predicting the 3D shape of a guidewire from a sequence of images. Each image sequence captures the guidewire from different time steps ($\mathbf{I}_{A,t}$), and the goal is to infer the continuous 3D shape $\mathbf{C}_t(\mathbf{u}_t)$. This many-to-one prediction task is akin to generating a variable-length sequence from variable-length input sequences, a technique commonly utilized in fields such as machine translation [33] and video analysis.

To achieve this, the input pipeline consists of a sequence of images depicting the guidewire. A Vision Transformer (ViT) [14], pre-trained on ImageNet, is employed to extract high-dimensional spatial feature representations from these images. The ViT generates feature maps $\mathbf{z}_t \in \mathbb{R}$. These feature maps are then fed into a Gated Recurrent Unit (GRU) to capture the temporal dependencies across the image sequence. The GRU processes the feature maps $\mathbf{z}_t$ from consecutive time steps, producing a sequence of hidden states $\mathbf{h}_t$. Formally, the GRU operation at time step t is defined as:

$$\mathbf{h}_t = \mathrm{GRU}(\mathbf{z}_t, \mathbf{h}_{t-1}), \tag{4}$$

The final hidden state $\mathbf{h}_t$ from the GRU is used by three distinct prediction heads, each tailored for a specific aspect of the guidewire shape prediction:

- **Tip Prediction Head:** responsible for predicting the 3D coordinates of the guidewire's tip. A fully connected layer maps the hidden state $\mathbf{h}_t$ to a Cartesian anchoring point $\mathbf{p} \in \mathbb{R}^3$.
- **Spherical Offset Prediction Head:** predicts the spherical coordinate offsets $(\Delta\phi, \Delta\theta)$ for points along the guidewire with a fixed radius r.
- **Stop Prediction Head:** outputs the probability distribution indicating the terminal point of the guidewire. It uses a softmax layer to produce a probability tensor $\mathbf{S}$, where each element $\mathbf{S}_j$ indicates the probability of the j-th point being the terminal point.

4.3 Loss Function

The custom loss function for training the model combines multiple components to handle the point-wise tip error, variable guidewire length (stop criteria), and tip position predictions. The overall loss function $\mathcal{L}_{\text{total}}$ is defined as:

$$\begin{aligned}\mathcal{L}_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N}\Big(&\lambda_{\text{tip}}\,\|\hat{\mathbf{p}}_i - \mathbf{p}_i\|^2 + \\ &\lambda_{\text{offset}}\big((\hat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i)^2 + (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)^2\big) + \\ &\lambda_{\text{stop}}\big(-\mathbf{s}_i \log(\hat{\mathbf{s}}_i) - (1-\mathbf{s}_i)\log(1-\hat{\mathbf{s}}_i)\big)\Big)\end{aligned} \tag{5}$$

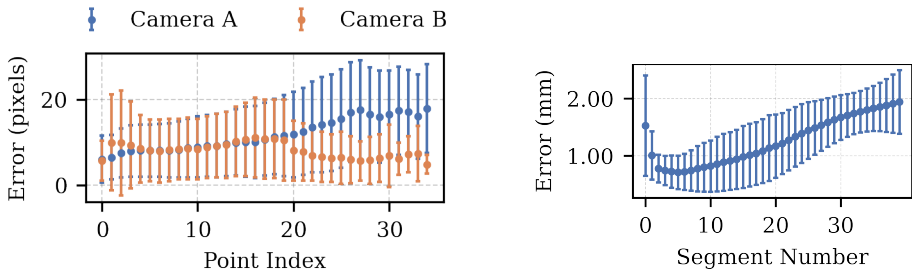


Fig. 6: Guidewire Reconstruction Error Analysis: (Left) Illustrates the distribution of reprojection errors, noting higher variability and peak errors in the mid-sections and reduced errors at the extremities. (Right) Presents the results of reconstruction validation.

where N is the number of samples, and λ_{tip}, λ_{offset}, and λ_{stop} are weights that balance the contributions of each loss component. The tip prediction loss ($\mathcal{L}_{\text{tip}}$) uses mean squared error (MSE) to ensure accurate 3D tip coordinates. The spherical offset loss ($\mathcal{L}_{\text{offset}}$) also uses MSE to align predicted and ground truth angular offsets, capturing the guidewire's shape. The stop prediction loss ($\mathcal{L}_{\text{stop}}$) employs binary cross-entropy (BCE) to accurately predict the guidewire's endpoint.

4.4 Training Details

The model was trained end-to-end using the loss from Eq. 5. The NAdam [15] optimizer was used with an initial learning rate of 1×10^{-4}. Additionally, a learning rate scheduler was employed to adjust the learning rate dynamically based on the validation loss. Specifically, the ReduceLROnPlateau scheduler was configured to reduce the learning rate by a factor of 0.1 if the validation loss did not improve for 10 epochs. The model was trained for 400 epochs, with early stopping based on the validation loss to further prevent overfitting.

5 Experiments

We evaluate our proposed dataset, Guide3D, through a structured experimental analysis, as follows: *i)* initially, we assess the dataset's validity, focusing on reprojection errors and their distribution across the dataset to understand its accuracy (Section 5.1), *ii)* we then explore the applicability of Guide3D in a 3D reconstruction task (Section 5.2), and *iii)* finally, we benchmark several segmentation algorithms to assess their performance on Guide3D, providing insights into the dataset's utility (Section 5.3).

5.1 Dataset Validation

Our analysis revealed a non-uniform distribution of reprojection errors across the dataset, with the highest variability and errors concentrated at the proximal end of the guidewire reconstructions. Figure 6 shows the reprojection error patterns for both Camera A and Camera B. For Camera A, mean errors increase from approximately 6px to a peak of 20px, with standard deviations rising from 5px to 11px, indicating growing inaccuracies and variability over time. Significant fluctuations around indices 25 to 27 highlight periods of particularly high error. For Camera B, mean errors exhibit an initial peak of 9px at index 1, followed by fluctuations that decrease towards the end. The standard deviations for Camera B start high at 11px and decrease over time, reflecting a pattern of high initial variability that stabilizes later. These patterns are consistent with the inherent flexibility of the guidewire, which can form complex shapes such as loops.

	Cartesian	Spherical
MaxED ↓	10.00 ± 4.64	6.88 ± 5.23
METE ↓	6.93 ± 3.94	3.28 ± 2.59
MERS ↓	5.33 ± 2.73	4.54 ± 3.67

Table 3: Shape Comparison (mm).

	DiceM	mIoU	Jaccard
UNet [30]	92.25	36.60	86.57
TransUnet [10]	**95.06**	**41.20**	**91.10**
SwinUnet [9]	93.73	38.58	88.55

Table 4: Segmentation Results

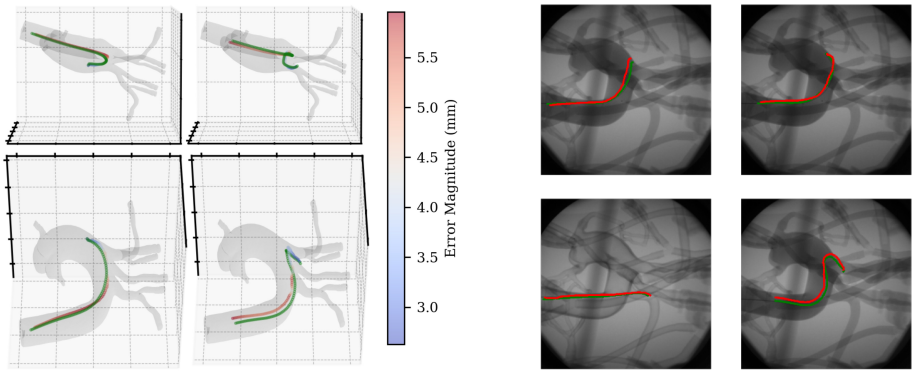
Furthermore, we conducted a validation procedure using CathSim [20], incorporating the aortic arch model described in Subsection 3.1 and a guidewire of similar diameter and properties. For sampling, we employed the soft actor-critic (SAC) algorithm with segmented guidewires and kinematic data, producing realistic validation samples. Evaluation metrics included maximum Euclidean distance (MaxED) at 2.880 ± 0.640 mm, mean error in tip tracking (METE) at 1.527 ± 0.877 mm, and mean error related to the robot's shape (MERS) at 0.001 ± 0.000. These results demonstrate the method's precision.

5.2 Guidewire Prediction Results

We now demonstrate the capability of the network introduced in Sec. 4 and thus highlight the importance of the proposed dataset. We examine the network prediction in the following manner: *1)* we first conduct an analysis between the predicted and reconstructed curve by employing piecewise metrics, and *2)* we showcase the reprojection error.

Shape Prediction Errors Table 3 presents the comparison of different metrics for shape prediction accuracy. Similarly to Subsection 5.1 we quantify the shape differences using the following metrics: *1)* Maximum Euclidean Distance (MaxED), *2)* Mean Error in Tip Tracking (METE), and *3)* Mean Error in Robot Shape (MERS). For all the metrics, the shape of the guidewire, represented as a 3D curve $\mathbf{C}(u)$, is sampled at equidistant Δu intervals along the arclength parameter u. Therefore, the metrics represent the pointwise discrepancies between the two shapes along the curve's arclength.

The results indicate that the Spherical representation consistently outperforms the Cartesian representation across all metrics. Specifically, the Maximum



(a) Guidewire Shape Predictions **(b)** Reprojections

Fig. 7: The figure illustrates the reconstruction similarity of the guidewire when reprojected onto the images. It demonstrates the network's capability to accurately predict the guidewire shape, even in the presence of noticeable angles, highlighting the robustness of the prediction model.

Euclidean Distance (MaxED) shows a lower error in the Spherical representation (6.88 ± 5.23 mm) compared to the Cartesian representation (10.00 ± 4.64 mm). Similarly, the Mean Error in Tip Tracking (METE) is significantly lower in the Spherical representation (3.28 ± 2.59 mm) than in the Cartesian representation (6.93 ± 3.94 mm). For the Mean Error in Robot Shape (MERS), the Spherical representation also demonstrates a reduced error (4.54 ± 3.67 mm) compared to the Cartesian representation (5.33 ± 2.73 mm). Lastly, the Fréchet distance shows a smaller error for the Spherical representation (6.70 ± 5.16 mm) compared to the Cartesian representation (8.95 ± 4.37 mm). These results highlight the advantage of using the Spherical representation for more accurate shape prediction.

Shape Comparison Visualization. Figure 7a showcases two 3D plots from different angles, comparing the ground truth guidewire shape to the predicted shape by the network. The network demonstrates its capability to accurately predict the guidewire shape, even in the presence of a loop and self-obstruction in the image. The predicted shape aligns closely with the actual configuration of the guidewire. Notably, the proximal end manifests a more substantial error relative to the nominal error seen at the distal end. Discrepancies from the authentic guidewire shape span from a mere 2mm at the distal end to a noticeable 5mm at the proximal end. Impressively, the network evidences its capability to accurately predict the guidewire's shape using only consecutive singular plane images. Subsequently, the 3D points are reprojected onto the original images, as illustrated in Fig. 7b.

5.3 Segmentation Results

We demonstrate Guide3D's potential to advance guidewire segmentation research by evaluating the performance of three state-of-the-art network archi-

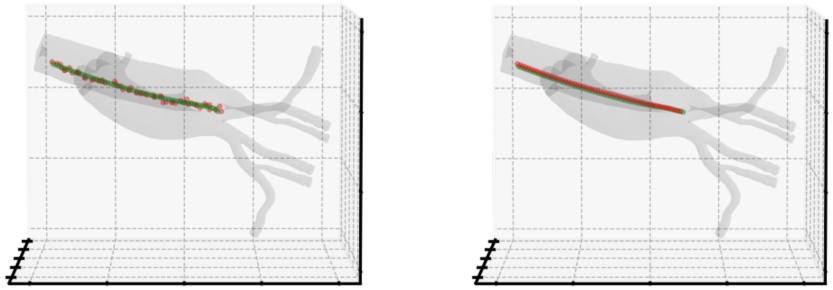


Fig. 8: Comparison between the predictions of the method not employing spherical coordinates (left) and the method employing spherical coordinates (right). It is evident that the method using spherical coordinates results in a more uniform and accurate shape reconstruction.

tectures: UNet [30] (learning rate: 1×10^{-5}, 135 epochs), TransUnet [10] (integrating ResNet50 and Vision Transformer [ViT-B-16], learning rate: 0.01, 199 epochs), and SwinUnet [9] (Swin Transformer architecture, learning rate: 0.01, 299 epochs). Performance metrics included the Dice coefficient (DiceM), mean Intersection over Union (mIoU), and Jaccard index, detailed in Table 4. The results indicate that UNet achieved a DiceM of 92.25, mIoU of 36.60, and Jaccard index of 86.57. TransUnet outperformed with a DiceM of 95.06, mIoU of 41.20, and Jaccard index of 91.10. SwinUnet recorded a DiceM of 93.73, mIoU of 38.58, and Jaccard index of 88.55. These findings benchmark the dataset's performance and suggest potential for future enhancements. Despite these promising results, the presence of loops and occlusions within the guidewire indicates that polyline prediction could significantly improve task utility.

6 Discussion and Conclusion

In this study, we introduce Guide3D, a publicly available bi-planar endovascular navigation dataset designed for the segmentation and 3D reconstruction of flexible, curved endovascular tools, addressing a significant gap in medical imaging research. While extensive experiments demonstrate the dataset's value, we acknowledge limitations such as the absence of clinical real human data due to regulatory challenges and the focus on synthetic and experimental scenarios that may not fully capture the variability of real-world clinical environments. Nevertheless, our standardized platform accommodates both video and image-based approaches, providing a versatile resource to facilitate the translation of these technologies into clinical settings. By including complexities like the guidewire's flexibility and the presence of loops and occlusions, we aim to push the boundaries of current methodologies, although further validation with clinical data is necessary to ensure robustness and generalizability. Our objective is to bridge the disparity between research developments and clinical application by establishing a standardized framework for evaluating various methodologies.

References

1. Altingövde, O., Mishchuk, A., Ganeeva, G., Oveisi, E., Hebert, C., Fua, P.: 3d reconstruction of curvilinear structures with stereo matching deep convolutional neural networks. Ultramicroscopy **234**, 113460 (2022)
2. Ambrosini, P., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: Fully automatic and real-time catheter segmentation in x-ray fluoroscopy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2017)
3. Ambrosini, P., Smal, I., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: 3d catheter tip tracking in 2d x-ray image sequences using a hidden markov model and 3d rotational angiography. In: AE-CAI (2015)
4. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: CVPR (2007)
5. Baur, C., Milletari, F., Belagiannis, V., Navab, N., Fallavollita, P.: Automatic 3d reconstruction of electrophysiology catheters from two-view monoplane c-arm image sequences. Int J Comput Assist Radiol Surg (2016)
6. Brainerd, E.L., Baier, D.B., Gatesy, S.M., Hedrick, T.L., Metzger, K.A., Gilbert, S.L., Crisco, J.J.: X-ray reconstruction of moving morphology (xromm): precision, accuracy and applications in comparative biomechanics research. J Exp Zool A Ecol Genet Physiol (2010)
7. Brost, A., Wimmer, A., Liao, R., Hornegger, J., Strobel, N.: Catheter tracking: Filter-based vs. learning-based. In: DAGM (2010)
8. Burgner, J., Herrell, S.D., Webster III, R.J.: Toward fluoroscopic shape reconstruction for control of steerable medical devices. In: Dynamic Systems and Control Conference. vol. 54761, pp. 791–794 (2011)
9. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV (2022)
10. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv (2021)
11. CVAT.ai Corporation: Computer vision annotation tool (cvat) (Nov 2023). https://doi.org/10.5281/zenodo.4009388, https://cvat.ai/
12. Danilov, V.V., Kolpashchikov, D.Y., Gerget, O.M., Laptev, N.V., Proutski, A., Gómez, L.A.H., Alvarez, F., Ledesma-Carbayo, M.J.: Use of semi-synthetic data for catheter segmentation improvement. Comput Med Imaging Graph (2023)
13. Delmas, C., Berger, M.O., Kerrien, E., Riddell, C., Trousset, Y., Anxionnat, R., Bracard, S.: Three-dimensional curvilinear device reconstruction from two fluoroscopic views. In: Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling. vol. 9415, pp. 100–110. Spie (2015)
14. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
15. Dozat, T.: Incorporating nesterov momentum into adam. ICLR (Workshop) (2016)
16. Gailloud, P., Muster, M., Piotin, M., Mottu, F., Murphy, K.J., Fasel, J.H., Rüfenacht, D.A.: In vitro models of intracranial arteriovenous fistulas for the evaluation of new endovascular treatment materials. AJNR (1999)

17. Hoffmann, M., Brost, A., Jakob, C., Bourier, F., Koch, M., Kurzidim, K., Hornegger, J., Strobel, N.: Semi-automatic catheter reconstruction from two views. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part II 15. pp. 584–591. Springer (2012)
18. Hoffmann, M., Brost, A., Jakob, C., Koch, M., Bourier, F., Kurzidim, K., Hornegger, J., Strobel, N.: Reconstruction method for curvilinear structures from two views. In: Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling. vol. 8671, pp. 630–637. Spie (2013)
19. Hoffmann, M., Brost, A., Koch, M., Bourier, F., Maier, A., Kurzidim, K., Strobel, N., Hornegger, J.: Electrophysiology catheter detection and reconstruction from two views in fluoroscopic images. IEEE Trans. Med. Imaging **35**(2), 567–579 (2015)
20. Jianu, T., Huang, B., Vu, M.N., Abdelaziz, M.E., Fichera, S., Lee, C.Y., Berthet-Rayne, P., y Baena, F.R., Nguyen, A.: Cathsim: An open-source simulator for endovascular intervention. IEEE T-MRB (2024)
21. Klema, V., Laub, A.: The singular value decomposition: Its computation and some applications. IEEE Transactions on automatic control **25**(2), 164–176 (1980)
22. Ma, Y., King, A.P., Gogin, N., Rinaldi, C.A., Gill, J., Razavi, R., Rhode, K.S.: Real-time respiratory motion correction for cardiac electrophysiology procedures using image-based coronary sinus catheter tracking. In: MICCAI (2010)
23. Martin, J.B., Sayegh, Y., Gailloud, P., Sugiu, K., Khan, H.G., Fasel, J.H., Rüfenacht, D.A.: In-vitro models of human carotid atheromatous disease. International Course Book of Peripheral Vascular Intervention (1998)
24. Mastmeyer, A., Pernelle, G., Barber, L., Pieper, S., Fortmeier, D., Wells, S., Handels, H., Kapur, T.: Model-based catheter segmentation in mri-images. arXiv (2017)
25. Nguyen, A., Kundrat, D., Dagnino, G., Chi, W., Abdelaziz, M.E., Guo, Y., Ma, Y., Kwok, T.M., Riga, C., Yang, G.Z.: End-to-end real-time catheter segmentation with optical flow-guided warping during endovascular intervention. In: ICRA. pp. 9967–9973. IEEE (2020)
26. Petković, T., Homan, R., Lončarić, S.: Real-time 3d position reconstruction of guidewire for monoplane x-ray. Comput Med Imaging Graph (2014)
27. Püschel, A., Schafmayer, C., Groß, J.: Robot-assisted techniques in vascular and endovascular surgery. Langenbeck's Archives of Surgery **407**(5), 1789–1795 (2022)
28. Rafii-Tari, H., Payne, C.J., Yang, G.Z.: Current and emerging robot-assisted endovascular catheterization technologies: a review. Ann Biomed Eng (2014)
29. Ramadani, A., Bui, M., Wendler, T., Schunkert, H., Ewert, P., Navab, N.: A survey of catheter tracking concepts and methodologies. Med Image Anal (2022)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
31. Sitzmann, V., Thies, J., Heide, F., Nießner, M., Wetzstein, G., Zollhofer, M.: Deepvoxels: Learning persistent 3d feature embeddings. In: CVPR. pp. 2437–2446 (2019)
32. Subramanian, V., Wang, H., Wu, J.T., Wong, K.C., Sharma, A., Syeda-Mahmood, T.: Automated detection and type classification of central venous catheters in chest x-rays. In: MICCAI (2019)
33. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Advances in neural information processing systems **27** (2014)
34. Verdonck, B., Bourel, P., Coste, E., Gerritsen, F.A., Rousseau, J.: Variations in the geometrical distortion of x-ray image intensifiers. In: Physics of Medical Imaging (1999)

35. Wagner, M., Schafer, S., Strother, C., Mistretta, C.: 4d interventional device reconstruction from biplane fluoroscopy. Medical physics **43**(3), 1324–1334 (2016)
36. Wu, X., Housden, J., Ma, Y., Rhode, K., Rueckert, D.: A fast catheter segmentation and tracking from echocardiographic sequences based on corresponding x-ray fluoroscopic image segmentation and hierarchical graph modelling. In: ISBI (2014)
37. Yi, X., Adams, S., Babyn, P., Elnajmi, A.: Automatic catheter and tube detection in pediatric x-ray images using a scale-recurrent network and synthetic data. JDI (2020)
38. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. (2000)
39. Zhou, Y.J., Xie, X.L., Zhou, X.H., Liu, S.Q., Bian, G.B., Hou, Z.G.: A real-time multifunctional framework for guidewire morphological and positional analysis in interventional x-ray fluoroscopy. Trans. Cogn. Develop. Syst. (2020)