SplineFormer: An Explainable Transformer Network for Autonomous Endovascular Navigation

Tudor Jianu[1], Shayan Doust[1], Mengyun Li[1], Baoru Huang[1*], Tuong Do[1], Hoan Nguyen[2], Karl Bates[3], Tung D. Ta[4], Sebastiano Fichera[5], Pierre Berthet-Rayne[6,7], Anh Nguyen[1]

***Abstract*— Robot-assisted endovascular navigation provides significant advantages, including reduced radiation exposure for surgeons and improved patient safety. However, a major challenge is to control curvilinear instruments like guidewires precisely for smooth and accurate navigation while adapting to anatomical variations and external forces. Traditional segmentation-based approaches struggle with real-time prediction of the guidewire's evolving shape, limiting their effectiveness in navigation tasks. In this paper, we propose SplineFormer, an explainable transformer network that predicts the continuous, structured representation of the guidewire as a B-spline. This formulation enables a compact, smooth, and explainable state representation that facilitates downstream navigation. By leveraging SplineFormer's predictions within an imitation learning framework, our system successfully performs autonomous endovascular navigation. Experimental results show that SplineFormer achieves a 50% success rate when fully autonomously cannulating the Brachiocephalic Artery in a real robotic setup, demonstrating its potential for improved autonomous navigation in endovascular interventions.**

I. Introduction

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for over a million deaths annually, with coronary heart disease and cerebrovascular disease being the primary contributors [1]. Endovascular interventions, such as Percutaneous Coronary Intervention (PCI), Pulmonary Vein Isolation (PVI), and Mechanical Thrombectomy (MT), have become well-established procedures for treating these conditions [2]–[4]. These minimally invasive techniques rely on the precise *navigation* of a *guidewire* and *catheter* through the vasculature to the target site, guided by intraoperative fluoroscopy. Once the target is reached, procedures such as thrombus removal, stent deployment, or tissue ablation are performed [5]. However, achieving *safe and efficient navigation* remains a critical challenge, as misalignment or excessive force can lead to vessel injury. Furthermore, the time-sensitive nature of these interventions, particularly in acute cases such as stroke, necessitates rapid and accurate navigation, where delays beyond 7.3 hours significantly reduce the benefits of MT [6]. Yet, only a fraction of eligible patients receive timely treatment, underscoring the need for advancements in autonomous navigation systems to improve procedural efficiency and accessibility [7].

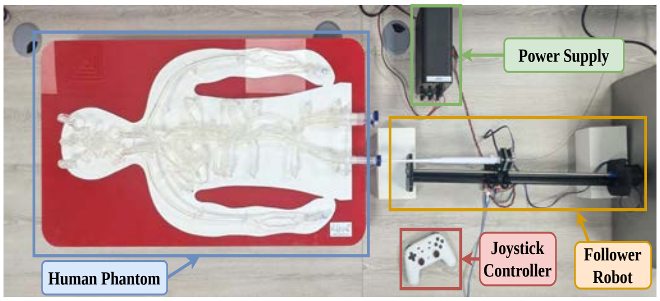


Fig. 1: **System Overview:** The experimental setup includes three main components: *i)* an anatomically accurate half-body vascular phantom model from Elastrat Sarl Ltd.; *ii)* a joystick controller (robotic leader); and *iii)* a robotic follower. Data is collected through teleoperation, where the robotic follower is controlled by the joystick controller.

To achieve precise endovascular navigation, operators primarily rely on fluoroscopic imaging to visualize the vasculature and manually guide instruments. However, prolonged fluoroscopy use poses risks such as radiation exposure to both patients and surgeons, along with potential nephrotoxicity from contrast agents used during angiography [8]. Moreover, manual operation demands significant skill and dexterity, increasing the risk of complications such as vessel perforation, misalignment, and distal embolization [9]. While robotic and semi-autonomous systems have been introduced to assist operators, they still require human oversight for fine adjustments, limiting scalability and increasing cognitive load. In contrast, fully autonomous approaches can reduce operator dependency and improve procedural efficiency in high-risk interventions.

Recent advancements in machine learning offer promising solutions to improve autonomous endovascular navigation [10]. Among various learning-based approaches, Reinforcement Learning (RL) has been extensively explored, showing the potential to optimize guidewire control strategies [11]–[15]. While RL methods have demonstrated early success, their deployment in real-world endovascular navigation remains limited due to factors such as high cognitive workload for operators, the need for precise state representations, and the difficulty of translating learned policies to real robots [16], [17]. Furthermore, ensuring explainability and safety in autonomous navigation is crucial as these technologies advance towards clinical applications [18].

[1] Department of Computer Science, University of Liverpool, UK
[2] University of Information Technology, VNUHCM, Vietnam
[3] Faculty of Health and Life Sciences, University of Liverpool, UK
[4] Department of Creative Informatics, The University of Tokyo, Japan
[5] Department of Mechanical, Materials and Aerospace Engineering, University of Liverpool, UK
[6] Honorary Fellow, University of Liverpool, UK
[7] 3IA Cote d'Azur, Sophia Antipolis, France
[*] Corresponding author

In this paper, we introduce *SplineFormer*, a transformer-based model designed to predict the guidewire's geometry in a structured and explainable manner. Unlike conventional segmentation-based methods, which generate dense pixel-wise representations, our approach predicts a continuous B-spline representation that captures the smooth and natural curvature of the guidewire. This formulation provides an efficient, compact, and structured representation, which facilitates downstream navigation by offering a state-space representation that can be leveraged for action selection. We validate our method in a high-fidelity experimental setup featuring a robotic system and an anatomically accurate human vascular phantom (Fig. 1). Utilizing SplineFormer's predictions, the robot successfully performs autonomous endovascular navigation, achieving a 50% success rate in cannulating the Brachiocephalic Artery (BCA). Our contributions are summarized as follows:

1) We introduce SplineFormer, an explainable transformer network capable of inferring the guidewire's geometry in a constrained parametric space.
2) We train our network to *i)* retrieve meaningful and concise representations, and *ii)* use this latent space to derive the appropriate actions to successfully navigate the anatomy.

II. Related Works

Guidewire and Catheter Representation. Accurate modeling of guidewires and catheters is a critical challenge in endovascular interventions, as it directly impacts the precision of robotic navigation. Traditional approaches have predominantly relied on segmentation-based methods, wherein the guidewire is identified from fluoroscopic images using classical image processing techniques such as pixel intensity analysis, texture feature extraction, and histogram-based methods [19]–[23]. More advanced methodologies, including the Hough transform and curvature-based techniques, have been employed to enhance detection accuracy [24]–[26]. However, these approaches often exhibit limitations in the presence of low contrast, occlusions, and imaging artifacts, thereby reducing their reliability for real-time navigation in dynamic clinical environments.

Deep learning methods have introduced more robust solutions for guidewire detection and segmentation. Convolutional Neural Networks (CNN) have been widely adopted for surgical tool localization, demonstrating significant improvements in accuracy [27]–[31]. In particular, U-Net-based architectures [32] have achieved state-of-the-art performance in segmenting guidewires from X-ray images, with refinements such as Recurrent Neural Networks (RNN) and adaptive binarization further improving accuracy [33], [34]. However, these methods primarily operate on pixel-wise segmentations, which fail to provide a structured, continuous representation of the guidewire's geometry. One common failure mode is fragmentation, where discontinuities appear in the predicted segmentation mask (Fig. 2), making it unsuitable for stable robotic navigation [35].

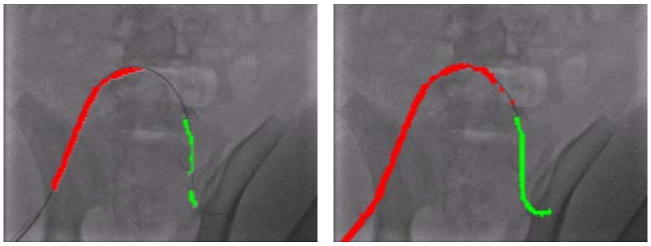


Fig. 2: **Segmentation Failure Cases.** Due to challenges in capturing thin and elongated guidewires, segmentation models often produce discontinuous segmented maps, which are unsuitable for stable robotic navigation.

Unlike segmentation-based methods, B-spline representations offer a structured and interpretable formulation of the guidewire's geometry. By parameterizing the guidewire as a set of control points, B-splines inherently enforce smoothness and geometric consistency, thereby reducing sensitivity to noise and enhancing navigation stability. This parametric representation is particularly advantageous for robotic applications, where compact, continuous state encodings are preferable to high-dimensional, pixel-wise outputs. Building upon this principle, our approach leverages SplineFormer to directly predict B-spline parameters, enabling seamless integration into downstream robotic navigation policies.

Autonomous Navigation. Autonomous navigation of guidewires within the vasculature is a highly complex task that has garnered significant attention due to its potential benefits, including reduced operative times, minimized radiation exposure, and improved procedural success rates. Recent advancements have increasingly focused on learning-based approaches to enhance the autonomous control of surgical tools, leveraging data-driven models to improve precision, adaptability, and robustness in dynamic clinical environments. Reinforcement Learning (RL) has been extensively explored in endovascular navigation, demonstrating the capability to learn complex control policies [12]–[15], [36]–[38]. Various RL-based approaches have been proposed to train agents for autonomous guidewire navigation. However, their performance is highly contingent on the quality of the state representation. Many existing methods rely on raw image-based observations, which are inherently high-dimensional and lack an explicit geometric structure, posing challenges for interpretability and efficient decision-making.

To mitigate these challenges, several studies have integrated Learning from Demonstration (LfD) techniques, such as Generative Adversarial Imitation Learning (GAIL) [37] and Behavioural Cloning (BC) [36], where expert trajectories serve as supervisory signals to guide learning. These approaches have demonstrated potential in reducing the sample complexity of RL algorithms by incorporating expert knowledge. However, existing methods often suffer from poor generalization due to the absence of structured state representations. In contrast, our proposed approach employs B-spline representations to construct a compact and structured state space, facilitating more stable and interpretable endovascular navigation.

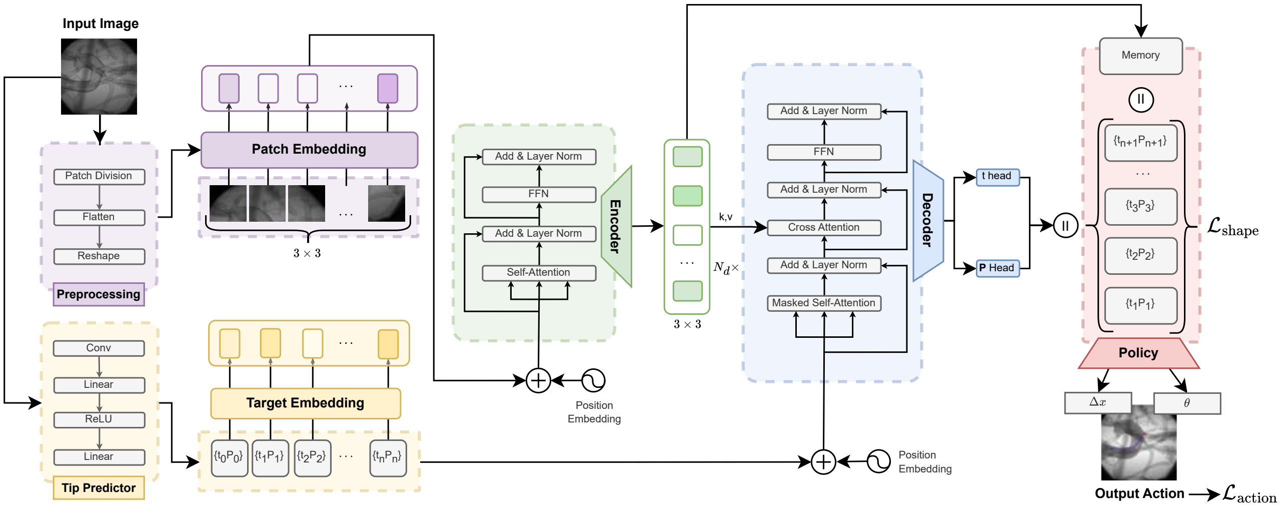


Fig. 3: **SplineFormer Network Architecture:** The input fluoroscopic image X is processed by a visual transformer encoder that divides the image into patches, embeds them, and generates visual feature representations X'. Simultaneously, a positional encoder processes embeddings of the target sequence. These encoded features are fed into a transformer decoder composed of N_D layers. Using masked self-attention and cross-attention mechanisms, the decoder sequentially generates the B-spline control points $\mathbf{P}_i$ and knots t_i that define the guidewire's geometry. The decoder predicts these parameters by projecting its outputs onto the B-spline dimensionality, yielding pairs $\{\mathbf{P}_0, t_0\}, \{\mathbf{P}_1, t_1\}, \ldots, \{\mathbf{P}_n, t_n\}$, which are subsequently passed to a multi-layer perceptron (MLP) to predict the navigation action $\mathbf{a}_t = (\Delta x, \theta)$. An independent tip predictor module initializes the generation by predicting the starting point $\{\mathbf{P}_0, t_0\}$.

III. B-Spline Representation for Autonomous Endovascular Navigation

Accurate geometric modeling of curvilinear guidewires and catheters is essential for precise navigation and control in interventional procedures. Conventional approaches primarily rely on dense image-based representations [39], which capture the tool's shape at the pixel level but lack an intrinsic parametric structure suitable for real-time navigation. To overcome these limitations, we develop SplineFormer (Fig. 3), which adopts a B-spline-based representation to encode the guidewire's shape in a compact, continuous, and structured form. By directly parameterizing the guidewire's geometry, our approach eliminates the need for post-processing steps required by segmentation-based methods, facilitating seamless integration into downstream robotic navigation policies.

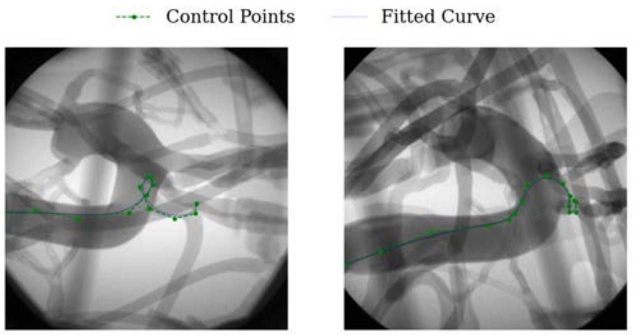


Fig. 4: **B-spline Representation of a Guidewire:** A guidewire can be represented as a continuous B-spline curve.

A. B-Spline Representation for Guidewire Modeling

A B-spline curve represents the guidewire's shape using a set of control points and a non-decreasing sequence of knots, providing a smooth and flexible parametric formulation (Fig. 4). This structured encoding inherently enforces continuity and geometric consistency, enabling stable and anatomically realistic trajectory adjustments during navigation. A B-spline curve $\mathbf{C}(t)$ of degree p is defined as:

$$\mathbf{C}(t) = \sum_{i=0}^{n} \mathbf{P}_i B_{i,p}(t), \tag{1}$$

where $\mathbf{P}_i$ denotes the control points, and $B_{i,p}(t)$ represents the B-spline basis functions of degree p, defined over the knot vector $t_0, t_1, \ldots, t_m$. The basis functions $B_{i,p}(t)$ are computed recursively as follows:

$$B_{i,0}(t) = \begin{cases} 1, & \text{if } t_i \leq t < t_{i+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

$$B_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} B_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} B_{i+1,p-1}(t). \tag{3}$$

Additionally, the constraint $\sum_{i=0}^{m-p-1} B_{i,p}(t) = 1$ ensures that the guidewire's shape remains consistent across varying imaging conditions, enabling real-time inference. To balance smoothness and computational efficiency, we set the B-spline degree to $p = 3$. The number of control points n is dynamically selected based on the guidewire's length, optimizing flexibility while maintaining model stability. The knot vector follows a uniform non-decreasing sequence, ensuring numerical robustness during optimization.

B. SplineFormer Network Architecture

Our SplineFormer (Fig. 3) is built upon a Vision Transformer (ViT) backbone [40] with $N_e = 12$ encoder layers and $N_D = 6$ decoder layers. Each encoder layer employs $H = 8$ attention heads with a hidden dimension of $d_k = 512$. Input fluoroscopic images are resized to 224×224 pixels and tokenized into 16×16 patches. The model is trained using the Adam optimizer on an NVIDIA A100 GPU with a batch size of 32, reaching convergence after 120K iterations.

Encoder. Two encoders are employed to extract structured features from fluoroscopic images, enabling precise guidewire representation for downstream navigation. The first encoder follows a ViT, where the input image $X \in \mathbb{R}^{H\times W}$ is divided into N_P patches, computed as:

$$N_P = \frac{H}{|P|} \times \frac{W}{|P|} \tag{4}$$

where $|P|$ denotes the patch size. These patches are flattened into a 1D tensor and embedded into a latent feature space via a linear projection layer. The second encoder is a sinusoidal positional encoder [41], applied to both the image patches P and the target sequence $Y_{\text{TGT}} \in \mathbb{R}^{S\times E}$, where S and E denote the sequence length and embedding dimension, respectively. This dual encoding mechanism ensures that the extracted features capture both local and global information, essential for accurately predicting the guidewire's shape.

The transformer encoder consists of N_e stacked identical layers, each comprising a Multi-Head Self-Attention (MHA) module followed by a positional Feed-Forward Network (FFN). The MHA module contains H parallel attention heads, each computing a scaled dot-product attention function. This architecture enables the network to attend to both local and global contextual information, which is critical for capturing the intricate bending and twisting of the guidewire during navigation. The outputs from all heads are aggregated via a learnable linear transformation W^O, formulated as:

$$\text{MHA}(Q, K, V) = \text{Concat}(h_1, h_2, \ldots, h_H)W^O \tag{5}$$

where the scaled dot-product attention for each head is computed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where $\{Q, K, V\} \in \mathbb{R}^{N_q\times d_k}$ represent the query, key, and value matrices, respectively.

Following the MHA module, the positional FFN is implemented using two fully connected layers with GELU activation and dropout regularization:

$$\text{FFN}(x) = \text{FC}_2\left(\text{Dropout}\left(\text{GELU}\left(\text{FC}_1(x)\right)\right)\right) \tag{7}$$

To enhance stability, each sublayer is equipped with a residual connection, followed by layer normalization:

$$x^{\text{out}} = \text{LayerNorm}\left(x^{\text{in}} + \text{Sublayer}(x^{\text{in}})\right) \tag{8}$$

Decoder. The decoder consists of N_D transformer layers that sequentially generate the B-spline control points and corresponding knots, defining the guidewire's smooth representation. Each layer comprises masked self-attention, multi-head cross-attention, and a FFN [41]. This hierarchical structure enables the model to iteratively refine the predicted guidewire shape while capturing its curvature and spatial relationships in a compact form.

The decoder's final output is projected onto the B-spline parameter space, yielding the structured representation:

$$\{\mathbf{P}_0, t_0\}, \{\mathbf{P}_1, t_1\}, \ldots, \{\mathbf{P}_n, t_n\} \tag{9}$$

where $\mathbf{P}_i$ are the control points and t_i are the associated knot values. This formulation defines the guidewire's geometry as a continuous parametric curve $\mathbf{C}(t)$, enabling accurate and interpretable navigation. By leveraging a structured and smooth state representation, this approach ensures both robust real-time inference and geometrically consistent trajectory estimation.

Tip Predictor. The tip predictor module establishes a geometrically consistent anchor for guidewire shape prediction by estimating the initial tip position in the image. It consists of a series of convolutional layers with ReLU activations, followed by linear transformations that regress the tip coordinates. Given an input image I, the module predicts the tip location:

$$\mathbf{P}_0 = (x_0, y_0), \tag{10}$$

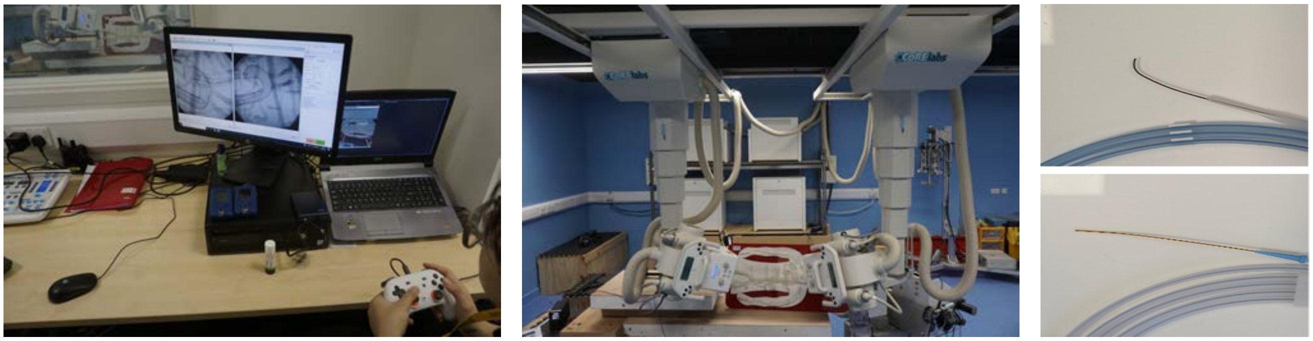
which serves as the initial token for autoregressive guidewire prediction, ensuring that the generated trajectory remains spatially coherent and well-anchored.

The encoder and tip predictor interact in a complementary manner: the encoder extracts hierarchical guidewire features from the fluoroscopic image, while the tip predictor refines the initial tip estimate using these extracted features. By leveraging encoder-derived representations, the tip predictor gains a richer contextual understanding of vessel topology, allowing it to produce a more precise and anatomically valid starting point. This is crucial, as an inaccurate tip estimation could introduce propagation errors in the subsequent trajectory prediction. The refined tip prediction is then passed to the decoder, which generates the full guidewire trajectory based on this initialization.

C. Loss Function

The shape loss function $\mathcal{L}_{\text{shape}}$ combines three weighted components, each scaled by a corresponding hyperparameter λ: *(i)* A Mean Squared Error (MSE) loss applied to the predicted and target sequences of control points and knots. *(ii)* A Binary Cross-Entropy (BCE) loss for the end-of-sequence (EOS) prediction. *(iii)* A curvature consistency loss, computed by sampling the predicted B-spline curve at n parameter values t_k, uniformly distributed over the valid parameter range $[t_p, t_{m-p}]$.

The sampled parameter values are defined as:



(a) (b) (c)

Fig. 5: **Robot Setup in X-ray Room:** The data acquisition and teleoperation process is highlighted, showcasing the user navigating the guidewire towards the designated target within the vascular phantom. The setup employs a Bi-planar X-ray. To enhance data variability, two different guidewires are used—the Radifocus™ Guide Wire M Stiff Type with an angled tip and the Nitrex Guidewire straight tip.

$$t_k = t_p + \frac{k}{n-1}(t_{m-p} - t_p), \quad k = 0, 1, \ldots, n-1. \tag{11}$$

The complete loss function is:

$$\begin{aligned}\mathcal{L}_{\text{shape}} = \frac{1}{N}\sum_{i=1}^{N}\Bigg(&\lambda_a\big(\|t_i - \hat{t}_i\|^2 + \|\mathbf{P}_i - \hat{\mathbf{P}}_i\|^2\big) + \\ &\lambda_b\big(-\mathbf{s}_i\log(\hat{\mathbf{s}}_i) - (1-\mathbf{s}_i)\log(1-\hat{\mathbf{s}}_i)\big) + \\ &\lambda_c\left\|\mathbf{C}(t_k) - \hat{\mathbf{C}}(t_k)\right\|^2\Bigg).\end{aligned} \tag{12}$$

The output of SplineFormer is a predicted sequence of control points and knots: This sequence serves not only as a shape descriptor but also as the input state for downstream action prediction. By coupling shape inference directly to action generation, the model forms a single cohesive pipeline from image to navigation decision.

D. Policy Training

The predicted control points and knots from Eq. 9 are directly passed to a Multi-Layered Perceptron (MLP) that predicts the next navigation action. The MLP processes these inputs—along with derived curvature features—through fully connected layers with ReLU activations, producing:

$$\mathbf{a}_t = (\Delta x, \theta), \tag{13}$$

where Δx denotes the translation step and θ denotes the rotation angle. This policy network is trained by minimizing the mean squared error between predicted and expert actions:

$$\mathcal{L}_{\text{action}} = \sum_{t}\|\mathbf{a}_t - \mathbf{a}_t^*\|^2. \tag{14}$$

This direct dependence between the predicted guidewire shape and the navigation action ensures that every control decision is grounded in the current estimate of the guidewire's geometry, providing both physical consistency and a structured link between perception and control.

IV. Experiments

To evaluate the effectiveness of our SplineFormer, we first setup the real endovascular robot and collect data to train the network. After training, the learned policy is evaluated on state-action pairs within our endovascular robotic system to assess how effectively the predicted B-spline representation supports autonomous guidewire navigation. All experiments were conducted on a computer equipped with an NVIDIA RTX 4080 GPU, 128 GB RAM, and an Intel Core i9-13900 processor, implemented using PyTorch.

A. Endovascular Robot Setup

Robot Setup. Endovascular robotic systems commonly utilize a leader-follower (master-slave) architecture, where the leader device—operated by a clinician—translates input commands to a follower robot that manipulates the catheter [42]. These systems typically offer up to six Degrees of Freedom (DoF), enabling precise translation and rotation control [43]. Human-Machine Interfaces (HMIs), such as multi-DoF joysticks or handheld controllers, convert operator movements into electromechanical actions, facilitating accurate catheter navigation [43], [44]. In our system, since only a guidewire is used, the robotic setup is streamlined to focus solely on translation and rotation movements. This simplification reduces mechanical complexity, making the design more accessible and easier to replicate compared to multi-DoF systems. The actuation mechanism comprises a NEMA 17 Bipolar stepper motor (59 N cm, 2 A) for linear motion, along with an additional motor for rotational control. System control is managed by an Arduino Uno Rev3 with a CNC shield and two A4988 drivers, powered by a 12 V DC power supply. Teleoperation input is provided via a Google Stadia joystick for intuitive manual control.

Data Collection. Our experiments utilize a Bi-planar X-ray system (Fig. 5) equipped with 60 kW Epsilon X-ray Generators (EMD Technologies Ltd.) and 16-inch Image Intensifier Tubes (Thales), incorporating dual focal spot Varian X-ray tubes for high-definition imaging. System calibration is achieved using acrylic mirrors and geometric alignment grids. To simulate human vascular anatomy, we employ

Method	Setup	Explainable?	BCA		LCCA	
			Success (%)	*Time (s)*	*Success (%)*	*Time (s)*
Expert Teleoperation	Fully Manual	-	100	32.1 ± 9.1	100	25.0 ± 6.7
GAIL-PPO [37]	Semi-Autonomous	No	69.4	52.1 ± 9.9	72.2	76.5 ± 24.1
Behavior Cloning [37]	Fully Autonomous	No	5.60	> 200	-	-
SplineFormer (ours)	Fully Autonomous	Yes	50.0	150 ± 45.6	-	-

TABLE I: Endovascular navigation results.

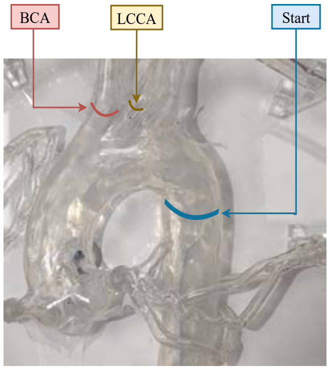


Fig. 6: Navigation setup.

a half-body vascular phantom model (Elastrat Sarl Ltd., Switzerland) enclosed in a transparent box and integrated into a closed water circuit to replicate blood flow. The model, constructed from soft silicone and featuring continuous flow pumps, was derived from detailed postmortem vascular casts, ensuring anatomical accuracy consistent with human vasculature [45], [46]. Finally, we utilized a Radifocus™ Guide Wire M Stiff Type (Terumo Ltd.), a $0.89\,\mathrm{mm}$ nitinol wire with a $3\,\mathrm{cm}$ angled tip and Nitrex Guidewire straight tip.

Dataset. We collect a dataset of 8,746 high-resolution samples ($1,024 \times 1,024$ pixels), consisting of $4,373$ paired instances with and without a simulated blood flow medium. Specifically, the dataset includes $6,136$ samples from the Radifocus Guidewire and $2,610$ from the Nitrex Guidewire, establishing a foundation for automated guidewire tracking in bi-planar X-ray images. Manual annotation was performed using CVAT tool [47], where polylines were meticulously created to capture the dynamic trajectory of the guidewire with high precision. To ensure balanced representation across different guidewire types and imaging conditions, the dataset was partitioned using a stratified sampling method.

B. Autonomous Navigation Results

Our SplineFormer was evaluated in a *fully autonomous* guidewire navigation task. The objective was to navigate from a predefined position in the descending aorta toward two distinct arterial targets: the Brachiocephalic Artery (BCA) and the Left Common Carotid Artery (LCCA), as illustrated in Fig. 6. For each target, the system performed 20 trials, recording trajectories to construct datasets of state-action pairs. The agent operated using fluoroscopic image observations, with an action space defined by translation within $\pm 2\mathrm{mm}$ and rotation within $\pm 15°$.

Following training, SplineFormer was deployed on the robotic platform for fully autonomous navigation. The system achieved a 50% success rate in reaching the BCA, with a mean completion time of 2.5 ± 0.76min. This represents a substantial improvement over the baseline BCA method, which achieved only 5.6% success under identical autonomous conditions. The semi-autonomous GAIL-PPO approach, which incorporates human demonstrations, performed better with success rates of 69.4% for the BCA and 72.2% for the LCCA, but unlike SplineFormer, it still required human intervention.

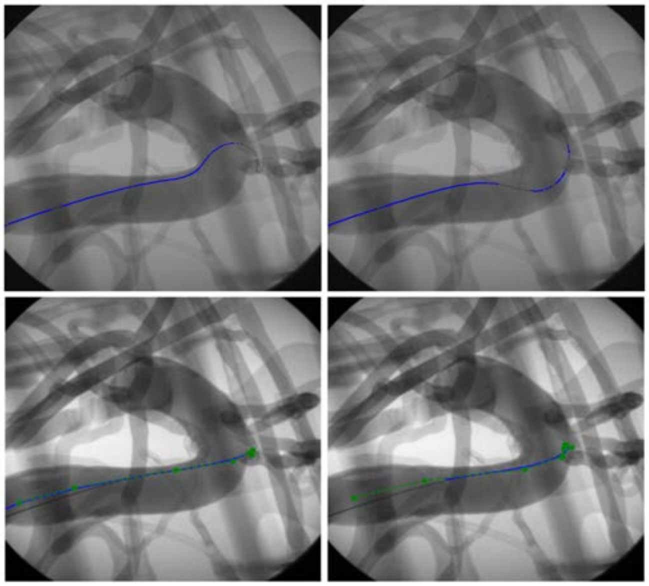


Fig. 7: **Qualitative comparison between SplineFormer and segmentation method.** U-Net (*top row*) produces segmentation masks, which often contain discontinuities and require additional post-processing before being used for robotic control. In contrast, our SplineFormer (*bottom row*) directly predicts the guidewire's geometry as a structured shape representation, making it more useful for navigation.

While our method did not surpass GAIL-PPO [37], a semi-autonomous method, in success rate, it offers the advantage of full autonomy. Moreover, our B-spline representation provides an explainable and structured state space, improving model interpretability. However, neither SplineFormer nor BC successfully cannulated the LCCA, highlighting challenges in navigating more complex vascular geometries. The failure in LCCA navigation can be attributed to its sharper curvature, narrower lumen diameter, and more abrupt bifurcation angle compared to the BCA. These factors increase resistance and the likelihood of collision, making it more difficult for the learned policy to generalize effectively.

C. Qualitative Results

SplineFormer was trained for 300 epochs on the annotated dataset from Section IV-A using the Adam optimizer with an initial learning rate of 1×10^{-5} and the loss function from Eq. 12. As illustrated in Fig. 7, the model effectively predicts the global guidewire shape within a compressed feature space, ensuring a compact and structured geometric representation. A key strength of SplineFormer is its ability to localize key guidewire points precisely. By leveraging a B-spline-based formulation, the model maintains smoothness and structural integrity, ensuring a continuous representation that aligns well with the guidewire's physical properties.

D. *Attention Visualization*

To better understand how the model processes fluoroscopic images, attention maps were generated from SplineFormer's transformer layers. Using maximal fusion across the final layer, with a discard factor to isolate key features, the resulting visualizations in Fig. 8 highlight the critical regions where the model concentrates its predictions. These attention maps reveal a strong focus on the guidewire tip and essential anatomical landmarks, including the central portion of the aortic arch and the BCA cannulation site.

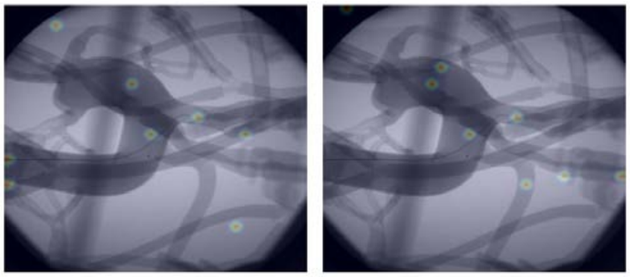


Fig. 8: **Attention visualization:** Attention maps highlight regions where SplineFormer focuses its predictions, including the guidewire tip and key vascular landmarks. This localized attention improves interpretability and facilitates precise autonomous navigation.

Unlike traditional segmentation methods, where attention is often dispersed across large areas of the image, SplineFormer exhibits highly localized attention, refining its focus on the most relevant regions for navigation. This precise attention mechanism enhances stability and accuracy in guidewire positioning, improving the reliability of autonomous navigation.

E. *Discussion*

We present SplineFormer, a new transformer network for predicting continuous guidewire geometries using a B-spline representation, enabling efficient encoding for real-time navigation. By parameterizing the guidewire with control points and knots, our approach ensures geometric consistency and provides a compact state representation for downstream robotic control. However, small variations in B-spline parameters can introduce spatial misalignment, affecting trajectory accuracy. In robotic experiments, our SplineFormer successfully cannulated the BCA but failed in the LCCA, consistent with prior findings [37], [48]. The LCCA's sharper turns, narrower lumen, and abrupt bifurcation angle increase resistance and the risk of collision, challenging model generalization. These findings highlight the need for improved adaptability in high-curvature vascular structures.

F. *Limitations*

While SplineFormer provides a structured and interpretable representation for autonomous navigation, its reliance on B-spline parameterization makes it sensitive to perturbations, which can affect trajectory stability. Additionally, its failure in LCCA navigation reinforces the well-documented challenges of traversing high-curvature vascular regions [37]. The LCCA's complex morphology makes navigation inherently more difficult than in the BCA, limiting policy generalization. Addressing these challenges will require adaptive trajectory optimization, uncertainty-aware planning, and biomechanical modeling to improve robustness. Despite these constraints, SplineFormer demonstrates the viability of shape-driven navigation strategies, motivating further research into learning-based autonomous interventions.

V. Conclusions

We introduce SplineFormer, a spline-based framework for autonomous endovascular navigation that encodes guidewire geometry in a compact and structured latent representation. Our method enhances trajectory planning and real-time decision-making, providing a foundation for precision-driven robotic control. While SplineFormer performs well in structured environments, addressing its limitations in high-curvature navigation is essential for real-world deployment. Future work will focus on adaptive learning strategies, biomechanical modeling, and preclinical validation to enhance generalization. By leveraging explainable curvilinear representations, SplineFormer shows the potential for fully autonomous endovascular interventions, contributing to safer and more efficient vascular navigation.

Acknowledgment

Hoan Nguyen was supported by the VNUHCM-University of Information Technology's Scientific Research Support Fund. Tung D. Ta and Anh Nguyen were supported by The Royal Society grant ICA\R1\231067.

References

[1] N. Townsend, L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, and M. Nichols, "Cardiovascular disease in europe: epidemiological update 2016," *Eur. Heart J.*, 2016.

[2] M. Goyal, B. K. Menon, W. H. Van Zwam, D. W. Dippel, P. J. Mitchell, A. M. Demchuk, A. Dá valos, C. B. Majoie, A. van Der Lugt, M. A. De Miquel, *et al.*, "Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials," *Lancet*, 2016.

[3] D. Giacoppo, R. Colleran, S. Cassese, A. H. Frangieh, J. Wiebe, M. Joner, H. Schunkert, A. Kastrati, and R. A. Byrne, "Percutaneous coronary intervention vs coronary artery bypass grafting in patients with left main coronary artery stenosis: a systematic review and meta-analysis," *JAMA Cardiol*, 2017.

[4] A. Lindgren, M. D. Vergouwen, I. van der Schaaf, A. Algra, M. Wermer, M. J. Clarke, and G. J. Rinkel, "Endovascular coiling versus neurosurgical clipping for people with aneurysmal subarachnoid haemorrhage," *Cochrane Database Syst. Rev.*, 2018.

[5] E. Brilakis, *Manual of percutaneous coronary interventions: a step-by-step approach.* Academic Press, 2020.

[6] J. L. Saver, M. Goyal, A. Van der Lugt, B. K. Menon, C. B. Majoie, D. W. Dippel, B. C. Campbell, R. G. Nogueira, A. M. Demchuk, A. Tomasello, *et al.*, "Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis," *JAMA*, 2016.

[7] P. McMeekin, P. White, M. A. James, C. I. Price, D. Flynn, and G. A. Ford, "Estimating the number of uk stroke patients eligible for endovascular thrombectomy," *Eur. Stroke J.*, 2017.

[8] M. R. Rudnick, S. Goldfarb, L. Wexler, P. A. Ludbrook, M. J. Murphy, E. F. Halpern, J. A. Hill, M. Winniford, M. B. Cohen, D. B. VanFossen, *et al.*, "Nephrotoxicity of ionic and nonionic contrast media in 1196 patients: a randomized trial," *Kidney Int.*, 1995.

[9] K. A. Hausegger, P. Schedlbauer, H. A. Deutschmann, and K. Tiesenhausen, "Complications in endoluminal repair of abdominal aortic aneurysms," *Eur. J. Radiol.*, 2001.

[10] P. Berthet-Rayne and G.-Z. Yang, "Navigation with minimal occupation volume for teleoperated snake-like surgical robots: Move," *Front. Robot. AI*, 2023.

[11] W. Chi, J. Liu, M. E. Abdelaziz, G. Dagnino, C. Riga, C. Bicknell, and G.-Z. Yang, "Trajectory optimization of robot-assisted endovascular catheterization with reinforcement learning," in *IROS*. Ieee, 2018.

[12] T. Behr, T. P. Pusch, M. Siegfarth, D. Hü sener, T. Mörschel, and L. Karstensen, "Deep Reinforcement Learning for the Navigation of Neurovascular Catheters," *Curr. Dir. Biomed. Eng.*, sep 2019.

[13] J. Kweon, K. Kim, C. Lee, H. Kwon, J. Park, K. Song, Y. I. Kim, J. Park, I. Back, J.-H. Roh, *et al.*, "Deep reinforcement learning for guidewire navigation in coronary artery phantom," *IEEE Access*, 2021.

[14] Y. Cho, J.-H. Park, J. Choi, and D. E. Chang, "Sim-to-real transfer of image-based autonomous guidewire navigation trained by deep deterministic policy gradient with behavior cloning for fast learning," in *IROS*. IEEE, 2022.

[15] L. Karstensen, J. Ritter, J. Hatzl, T. P ätz, J. Langejürgen, C. Uhl, and F. Mathis-Ullrich, "Learning-based autonomous vascular guidewire navigation without human demonstration in the venous system of a porcine liver," *Int. J. Comput. Assist. Radiol. Surg.*, 2022.

[16] M. Mofatteh, "Neurosurgery and artificial intelligence," *AIMS Neurosci.*, 2021.

[17] W. Crinnion, B. Jackson, A. Sood, J. Lynch, C. Bergeles, H. Liu, K. Rhode, V. M. Pereira, and T. C. Booth, "Robotics in neurointerventional surgery: a systematic review of the literature," *J. Neurointerv. Surg.*, 2022.

[18] E. Fosch-Villaronga and T. Mahler, "Cybersecurity, safety and robots: Strengthening the link between cybersecurity and safety in the context of care robots," *Comput. Law Secur. Rev.*, 2021.

[19] A. Brost, R. Liao, J. Hornegger, and N. Strobel, "3-d respiratory motion compensation during ep procedures by image-based 3-d lasso catheter model generation and tracking," in *MICCAI*, 2009.

[20] A. Brost, R. Liao, N. Strobel, and J. Hornegger, "Respiratory motion compensation by model-based catheter tracking during ep procedures," *Medical Image Analysis*, 2010.

[21] S. A. Baert, M. A. Viergever, and W. J. Niessen, "Guide-wire tracking during endovascular interventions," *TMI*, 2003.

[22] W. Wu, T. Chen, P. Wang, S. K. Zhou, D. Comaniciu, A. Barbu, and N. Strobel, "Learning-based hypothesis fusion for robust catheter tracking in 2d x-ray fluoroscopy," in *CVPR 2011*. IEEE, 2011.

[23] Y. Ma, N. Gogin, P. Cathier, R. J. Housden, G. Gijsbers, M. Cooklin, M. O'Neill, J. Gill, C. A. Rinaldi, R. Razavi, *et al.*, "Real-time x-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions," *Medical physics*, 2013.

[24] C. Sheng, L. Li, and W. Pei, "Automatic detection of supporting device positioning in intensive care unit radiography," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2009.

[25] E.-F. Kao, T.-S. Jaw, C.-W. Li, M.-C. Chou, and G.-C. Liu, "Automated detection of endotracheal tubes in paediatric chest radiographs," *Comput. Methods Programs Biomed.*, 2015.

[26] V. Bismuth, R. Vaillant, H. Talbot, and L. Najman, "Curvilinear structure enhancement with the polygonal path image-application to guide-wire segmentation in x-ray fluoroscopy," in *MICCAI*, 2012.

[27] P. Wang, T. Chen, Y. Zhu, W. Zhang, S. K. Zhou, and D. Comaniciu, "Robust guidewire tracking in fluoroscopy," in *CVPR*, 2009.

[28] L. Wang, X.-L. Xie, G.-B. Bian, Z.-G. Hou, X.-R. Cheng, and P. Prasong, "Guide-wire detection using region proposal network for x-ray image-guided navigation," in *IJCNN*, 2017.

[29] O. Pauly, H. Heibel, and N. Navab, "A machine learning approach for deformable guide-wire tracking in fluoroscopic sequences," in *MICCAI*, 2010.

[30] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, "Improving catheter segmentation & localization in 3d cardiac ultrasound using direction-fused fcn," in *ISBI*, 2019.

[31] P. Zaffino, G. Pernelle, A. Mastmeyer, A. Mehrtash, H. Zhang, R. Kikinis, T. Kapur, and M. F. Spadea, "Fully automatic catheter segmentation in mri with 3d convolutional neural networks: application to mri-guided gynecologic brachytherapy," *Phys. Med. Biol.*, 2019.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[33] X. Yi, S. Adams, P. Babyn, and A. Elnajmi, "Automatic catheter and tube detection in pediatric x-ray images using a scale-recurrent network and synthetic data," *J. Digit. Imaging*, 2019.

[34] P. Ambrosini, D. Ruijters, W. J. Niessen, A. Moelker, and T. van Walsum, "Fully automatic and real-time catheter segmentation in x-ray fluoroscopy," in *MICCAI*. Springer, 2017.

[35] B. Zhang, M. Bui, C. Wang, F. Bourier, H. Schunkert, and N. Navab, "Real-time guidewire tracking and segmentation in intraoperative x-ray," in *Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*. SPIE, 2022.

[36] W. Chi, J. Liu, H. Rafii-Tari, C. Riga, C. Bicknell, and G.-Z. Yang, "Learning-based endovascular navigation through the use of non-rigid registration for collaborative robotic catheterization," *Int. J. Comput. Assist. Radiol. Surg.*, 2018.

[37] W. Chi, G. Dagnino, T. M. Kwok, A. Nguyen, D. Kundrat, M. E. Abdelaziz, C. Riga, C. Bicknell, and G.-Z. Yang, "Collaborative robot-assisted endovascular catheterization with generative adversarial imitation learning," in *ICRA*. Ieee, 2020.

[38] H. You, E. Bae, Y. Moon, J. Kweon, and J. Choi, "Automatic control of cardiac ablation catheter with deep reinforcement learning method," *J. Mech. Sci. Technol.*, 2019.

[39] X. Wu, J. Housden, Y. Ma, B. Razavi, K. Rhode, and D. Rueckert, "Fast catheter segmentation from echocardiographic sequences based on segmentation from corresponding x-ray fluoroscopy for cardiac catheterization interventions," *TMI*, 2014.

[40] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[41] A. Vaswani, "Attention is all you need," *NeurIPS*, 2017.

[42] Y. Song, L. Li, Y. Tian, Z. Li, and X. Yin, "A novel Master-Slave interventional surgery robot with force feedback and collaborative operation," *Sensors*, mar 2023.

[43] W. Saliba, J. E. Cummings, S. Oh, Y. Zhang, T. N. Mazgalev, R. A. Schweikert, J. D. Burkhardt, and A. Natale, "Novel robotic catheter remote control system: feasibility and safety of transseptal puncture and endocardial catheter navigation," *Journal of cardiovascular electrophysiology*, 2006.

[44] E. M. Khan, W. Frumkin, G. A. Ng, S. Neelagaru, F. M. Abi-Samra, *et al.*, "First experience with a novel robotic remote catheter system: Amigo ™ mapping trial," *Journal of Interventional Cardiac Electrophysiology*, 2013.

[45] J.-B. Martin, Y. Sayegh, P. Gailloud, K. Sugiu, H. G. Khan, J. H. Fasel, and D. A. Rü fenacht, *In-vitro models of human carotid atheromatous disease*, 1998.

[46] P. Gailloud, M. Muster, M. Piotin, F. Mottu, K. J. Murphy, J. H. Fasel, and D. A. R üfenacht, "In vitro models of intracranial arteriovenous fistulas for the evaluation of new endovascular treatment materials," *AJNR Am. J. Neuroradiol.*, 1999.

[47] CVAT.ai Corporation, "Computer vision annotation tool (cvat)," nov 2023.

[48] T. Jianu, B. Huang, M. N. Vu, M. E. Abdelaziz, S. Fichera, C.-Y. Lee, P. Berthet-Rayne, F. R. y Baena, and A. Nguyen, "Cathsim: an open-source simulator for endovascular intervention," *T-MRB*, 2024.