Catheter PCI Intervention Sim2Real

Haoran Zhou[1], Weibo Gao[1], Katsuo Kurabayashi[1,*], Hao Su[1,*]

***Abstract*— Mitral valve closure remains an extremely challenging surgical procedure, primarily due to the difficulty of simultaneously navigating and controlling both catheter and guidewire from the femoral vein into the left atrium and ventricle. The venous pathways are often tortuous and narrow, further compounding the complexity of the intervention. To address these challenges, we propose a VAE-based deep reinforcement learning framework that integrates X-ray imaging, multi-instrument tracking, and virtual haptics to coordinate the maneuvering of the guidewire, microcatheter, and guide catheter from the femoral vein to the mitral valve. The policy is trained using heart and vascular phantoms in conjunction with a catheter–robot model developed within the SOFA simulation and Issac Lab environment, and is subsequently deployed on a physical robotic platform with vascular phantoms for validation. We further demonstrate zero-shot transfer from simulated pseudo X-ray images to real-time X-ray image segmentation and instrument tracking, as well as performance improvements achieved through the incorporation of virtual haptics.**

I. Introduction

II. Method

The VAE based multi-agent SAC policy was trained in a SOFA simulation environment and deployed on the robot and physical vascular phantom.

A. SOFA Simulation

We partially adopted the SOFA-based stEVE platform [1] as the foundational digital twin environment. On top of this, we further developed a delta-reward mechanism for multiple agents, incorporated a full vascular mesh from the femoral vein to the left atrium, implemented contact force modeling (normal and tangential), extended instrument dynamics and communication with the VAE-DRL environment, and enabled interactions with both the physical robot and the vascular phantom.

Vascular Digital Twin. The vascular digital twin was remeshed and decimated in Blender to enhance computational efficiency in SOFA. A new material property, approximating that of the physical vascular phantom, was assigned to the digital twin. In addition, reaction force modeling, including both tangential and normal components, was incorporated.

Instruments. Instrument morphologies were measured, and their simulation parameters were calibrated to match those of the real devices. Specifically, we modeled the XXXX guidewire, XXXX microcatheter, and XXXX guide catheter based on instruments from XXX company.

Simulation Camera and Pseudo X-Ray. We integrated XXX virtual cameras into the simulation environment. At each time step, the cameras captured images from XXX different viewpoints, which were further processed to approximate pseudo X-ray projections.

Virtual Haptics. By combining mesh properties, reaction forces, instrument characteristics, and motion velocities, we generated a virtual haptic feedback signal from the simulation.

Delta Reward.

Communication.

B. Foundation Model-Based Image Processing

The image stream captured by the SOFA simulation cameras was processed using the DINOv3 model [2], enabling vascular system segmentation as well as instrument segmentation and tracking across frames.

Vascular System Segmentation. For vascular segmentation, we applied few-shot training of the DINOv3 model using pseudo X-ray images with simple data augmentation strategies, achieving mAP XXX performance.

Instruments Tracking and Segmentation. We adopted a few-shot learning strategy to adapt the DINOv3 model for instrument tracking and segmentation. Pseudo X-ray images from multiple viewpoints were used for training, covering three types of instruments: the guidewire, microcatheter, and guide catheter. To extend beyond segmentation and tracking, we modified the final layer of DINOv3 and provided ground-truth 6-DoF poses of the instruments in the X-ray images. This enabled simultaneous segmentation, temporal tracking, and pose estimation, achieving XXX mAP for segmentation/tracking and XXX accuracy for pose estimation.

C. VAE based Multi Agent Soft Actor-Critic

Inputs. Image, 6DOF * 3, virtual haptics, reward.

Encoder. VAE

Multiagent SAC.

output

[1]Haoran Zhou is with the Mechanical and Aerospace Engineering Department, Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA.

[1]Weibo Gao is with the Biomedical Engineering Department, Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA.

[1]Katsuo Kurabayashi is with the Mechanical and Aerospace Engineering Department and Biomedical Engineering Department, Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA. katsuo.k@nyu.edu

[1]Hao Su is with the Biomedical Engineering Department, Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA. hao.su@nyu.edu

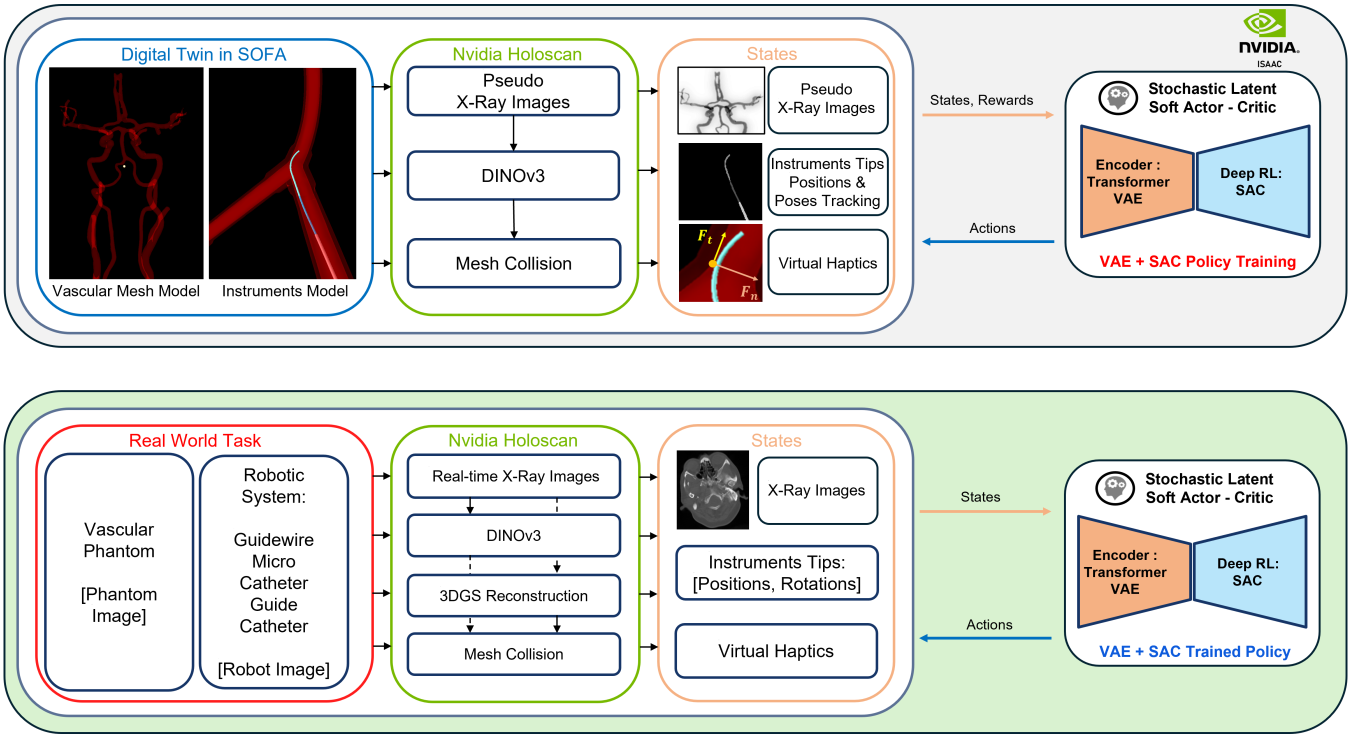

Fig. 1: **Catheter Intervention Learning in Simulation Flowchart: Policy Training(Up); Policy Deployment(Bottom)**

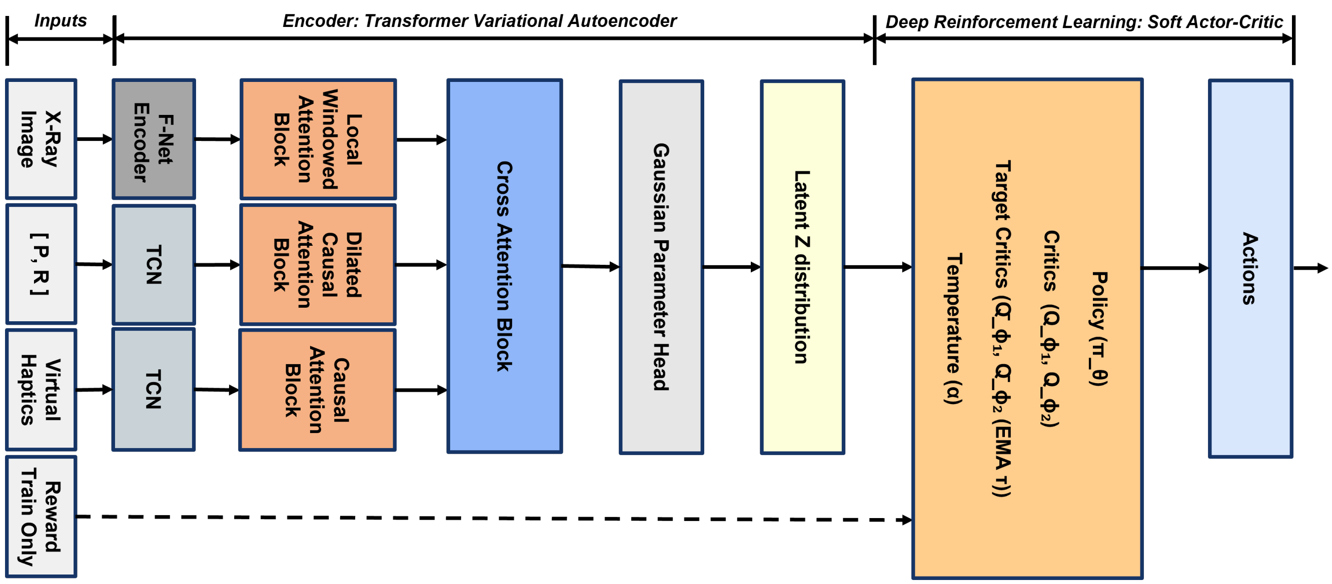

Fig. 2: **VAE Multi Agents Soft Actor-Critic Architecture**

III. Results

References

[1] L. Karstensen, H. Robertshaw, J. Hatzl, B. Jackson, J. Langejürgen, K. Breininger, C. Uhl, S. H. Sadati, T. Booth, C. Bergeles *et al.*, "Learning-based autonomous navigation, benchmark environments and simulation framework for endovascular interventions," *Computers in Biology and Medicine*, vol. 196, p. 110844, 2025.

[2] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa *et al.*, "Dinov3," *arXiv preprint arXiv:2508.10104*, 2025.