Deformable Model-to-Image Registration Toward Augmented Reality-Guided Endovascular Interventions

Li, Zhen; Contini, Letizia; Ippoliti, Alessandro Maria; Bastianelli, Elena; De Marco, Federico; Dankelman, Jenny; De Momi, Elena

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Deformable Model-to-Image Registration Toward Augmented Reality-Guided Endovascular Interventions

Zhen Li, *Graduate Student Member, IEEE*, Letizia Contini, Alessandro Maria Ippoliti,
Elena Bastianelli, Federico De Marco, Jenny Dankelman, *Member, IEEE*,
and Elena De Momi, *Senior Member, IEEE*

***Abstract*—Endovascular interventions are minimally invasive procedures that utilize the vascular system to access anatomical regions deep within the body. Image-guided assistance provides valuable real-time information about the dynamic state of the vascular environment. However, the reliance on intraoperative 2-D fluoroscopy images limits depth perception, prompting the demand for intraoperative 3-D imaging. Existing image registration methods face difficulties in accurately incorporating tissue deformations compared to the preoperative 3-D model, particularly in a weakly supervised manner. Additionally, reconstructing deformations from 2-D to 3-D space and presenting this intraoperative model visually to clinicians poses further complexities. To address these challenges, this study introduces a novel deformable model-to-image registration framework using deep learning. Furthermore, this research proposes a visualization method through augmented reality to guide endovascular interventions. This study utilized image data collected from nine patients who underwent transcatheter aortic valve implantation (TAVI) procedures. The registration results in 2-D indicate that the proposed deformable model-to-image registration framework achieves a modified dice similarity coefficient (MDSC) value of 0.89 ± 0.02 and a penalization of deformations in spare space (PDSS) value of 0.04 ± 0.01, offering an improvement of 3.5%–98.6% over the state-of-the-art image registration approach. Additionally, the accuracy of registration in 3-D was evaluated using a dataset obtained from an intervention simulator, resulting in a mean absolute error (MAE) of 1.51 ± 1.02 mm within the region of interest. Overall, the study validates the feasibility and accuracy of the proposed weakly supervised deformable model-to-image registration framework, demonstrating its potential to provide intraoperative 3-D imaging as intervention assistance in dynamic vascular environments.**

***Index Terms*— Augmented reality, deep learning, deformation, image registration, image-guided interventions.**
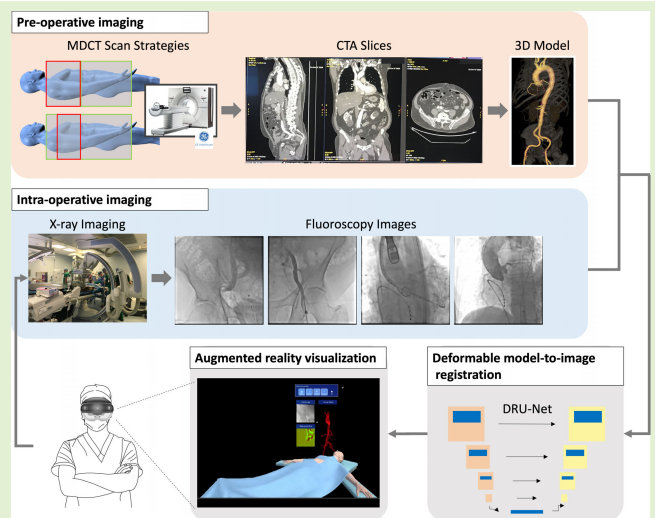
I. Introduction

THE rising demand for minimally invasive procedures has expedited the acceptance and implementation of endovascular interventions [1]. Endovascular interventions

utilize the vascular system to access anatomical regions deep within the body. For example, transcatheter aortic valve implantation (TAVI) procedure, a percutaneous cardiological intervention, facilitates the implantation of a miniaturized biological valve prosthesis into the aortic root. This minimally invasive approach is designed to address aortic valve pathologies, specifically aortic stenosis and steno-insufficiency [2].

During endovascular interventions, image-based guidance plays a crucial role in providing clinicians and robotics systems with valuable insights into the dynamic vascular environment [3], [4]. However, conventional 2-D images commonly used for intervention guidance, such as 2-D X-ray fluoroscopy and digital subtraction angiography (DSA), are often deemed insufficient due to limited information and the absence of depth perception. Consequently, there is a growing demand for intraoperative 3-D imaging [5]. By fusing 3-D preoperative data with 2-D intraoperative images, complex clinical

procedures can benefit from enhanced visualization of concealed structures and a more comprehensive anatomical model [6], [7], [8]. The necessity for introducing a deformable model-to-image registration approach stems from various physiological factors such as heartbeat, respiration, patient movement, and instrument insertion, all of which can induce vascular deformations and adversely impact registration accuracy [9].

Existing image registration methods to reconstruct vascular deformation can be broadly categorized into optimization-based approaches, which rely on iterative optimization processes, and learning-based approaches that leverage neural networks. Zhang et al. [11] proposed a method to reconstruct a deformed intraoperative 3-D aortic model using a preoperative 3-D model and intraoperative fluoroscopy images. They formulated the deformation estimation process as a nonlinear optimization problem based on the deformation graph approach, utilizing the comparison between preoperative model projection contours and intraoperative segmented aortic shape contours. However, optimization-based methods often suffer from high computational complexity [12], [13]. To illustrate, the iterative closest point (ICP) method, for instance, necessitates more than 10 min for execution [12]. Haskins et al. [13] conducted a survey on learning-based methods and highlighted a significant limitation: most studies in the literature rely on landmarks or manually annotated features, making them severely constrained by the laborious task of generating datasets. In [14], a multichannel convolutional neural network (CNN) was employed to achieve favorable registration results, demonstrating an average error of approximately 0.3 mm. This approach, however, necessitates the definition of a mathematical model for periodic deformation, which is only feasible when a complete dataset representing all phases of the periodic movement is available. This requirement implies the need for a long exposure time and significant amounts of contrast media. Hu et al. [15] proposed a weakly supervised CNN, which, however, does not include 3-D model deformation reconstruction. Overall, existing methods encounter challenges in effectively handling tissue deformations in a weakly supervised manner and accurately reconstructing deformations from 2-D to 3-D space.

Augmented reality (AR) visualization has been widely acknowledged in numerous studies for its ability to offer crucial advantages during clinical procedures. These benefits include providing valuable insights into the physiology of deformable organs and enabling clinicians to integrate information from multiple sources seamlessly, all while maintaining a clear line of sight with the patient in the operating room [16], [17], [18]. These advantages are particularly significant in the context of minimally invasive interventions, where direct visual observation is inherently limited [19]. Moreover, several studies have highlighted the positive outcomes achieved through the integration of AR with robotic-assisted procedures, as evidenced by user-centric and ergonomic evaluation criteria [20], [21], [22].

To overcome the aforementioned challenges, this article presents a novel deformable model-to-image registration framework using deep learning, specifically tailored for augmented reality-guided endovascular catheterization. Building upon our previous work [10], which proposed an affine model-to-image registration approach to align segmented fluoroscopy images with a preoperative 3-D model reconstructed from computed tomography angiography (CTA) scans, this study extends the registration pipeline. We introduce a phase for deformation prediction and reconstruction and incorporate immersive AR visualization using a head-mounted display (HMD) device. The main contributions of this research can be summarized as follows.

1) Proposal of an accurate deep-learning-based deformable model-to-image registration framework for predicting and reconstructing deformations from 2-D images onto the preoperative 3-D model, using a deep residual U-Net (DRU-Net) model with a customized loss function to adequately capture the registration accuracy.

2) Development of an immersive visualization interface for intraoperative 3-D models using the AR HMD.

3) Validation of the registration accuracy both in 2-D using a dataset comprising nine patients and in 3-D using a dataset obtained from an intervention simulator.

The article is organized as follows: Section II provides an overview of the proposed deformable model-to-image registration framework. Section III explains the experimental design and the performance metrics employed to evaluate the accuracy of the results. Section IV showcases the results, accompanied by a relevant discussion. Finally, Section V concludes the article and outlines future directions for research.

II. MATERIALS AND METHODS

The proposed deformable model-to-image registration framework is illustrated in Fig. 1, consisting of six modules.

1) *3-D Model Reconstruction:* Semi-automatic 3-D reconstruction of the patient's model from preoperative CTA images. These images are obtained using two typical multidetector computed tomography (MDCT) scan strategies [23]. The first strategy involves electrocardiogram (ECG)-synchronized CTA of the aortic root and heart, followed by non-ECG-synchronized helical CTA of the thorax, abdomen, and pelvis. The second strategy comprises ECG-synchronized CTA of the thorax, followed by non-ECG-synchronized helical CTA of the abdomen and pelvis.

2) *2-D Fluoroscopy Image Segmentation:* Acquisition of intraoperative fluoroscopy images depicting various field-of-view (FoV) along the insertion route, followed by automatic segmentation using a DRU-Net to generate binary images [10]. These fluoroscopic images predominantly focus on two main FoVs during interventions: the entry site, typically the femoral arteries, and the target site, generally the aortic root. These fluoroscopy images are typically captured at key stages of the intervention: first, following the insertion of the needle into the femoral arteries; second, before the inflation of the balloon catheter; and finally, after the placement of the stent at the aortic root.
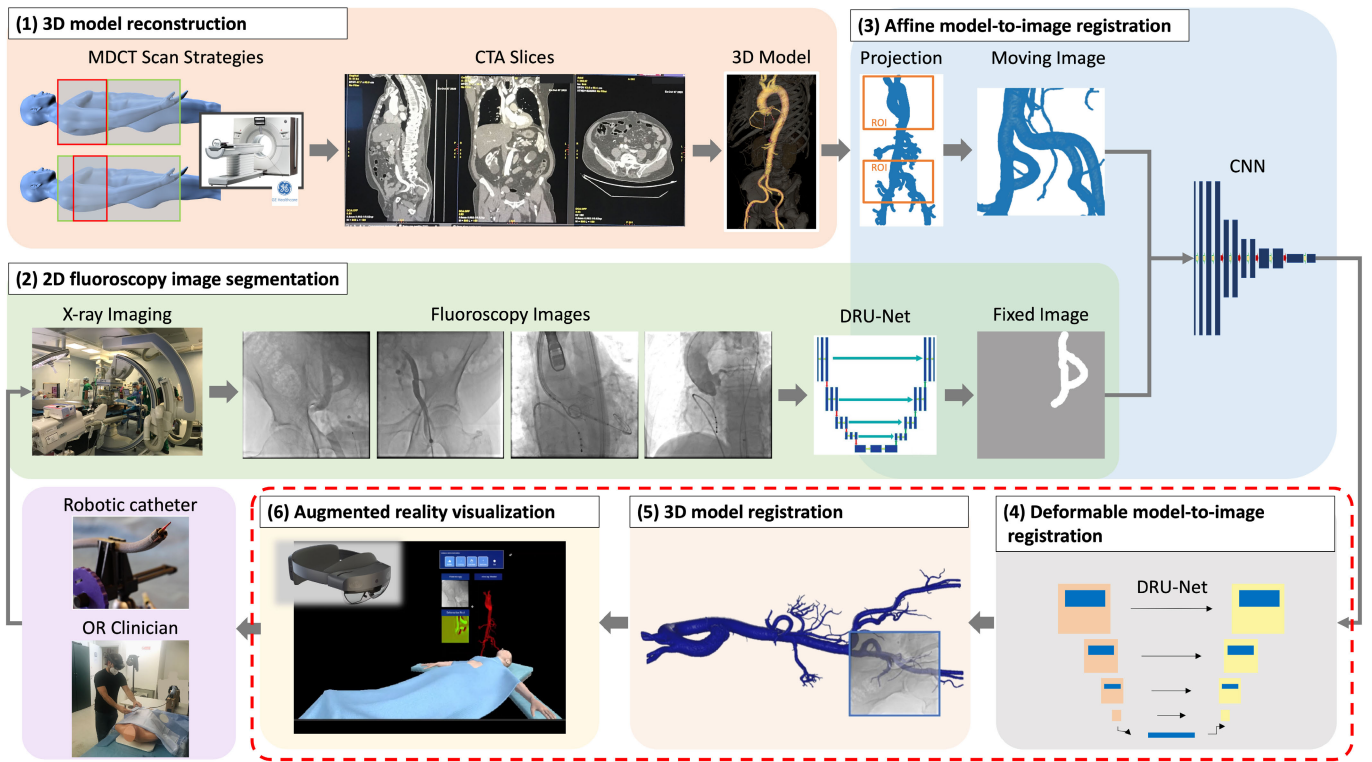
Fig. 1.   Architecture of the proposed framework consists of multiple modules. 1) Reconstruction of a preoperative 3-D aortic model from CTA slices. 2) Autonomous segmentation of vessels from intraoperative fluoroscopy images using a DRU-Net. 3) Generation of 2-D ROI projections and registration with segmented images using a hybrid approach incorporating a CNN. 4) The DRU-Net model to estimate the registration matrix and deformation field, respectively. 5) Application of these parameters to deform the preoperative 3-D model. 6) Visualization of the deformed 3-D model alongside the intraoperative fluoroscopy image using an AR HMD (Microsoft HoloLens 2[1]). The red dashed box highlights the extensions introduced in comparison to our previous work [10].

3) *Affine Model-to-Image Registration:* Conversion of the model-to-image registration problem into an image-to-image registration problem through the projection of a 2-D view of the Region of Interest (ROI) from the 3-D model, based on the fluoroscopy image. An CNN model is employed to estimate an affine registration matrix, which aligns the ROI projection with the binary image segmented from the fluoroscopy image [10]. The ROI of a patient-specific model is determined interactively by delineating a bounding box on a projection image of the 3-D model. This process involves the selection of two distinct ROIs for each patient: one encompassing a lower FoV, which includes the femoral arteries, and the other covering an upper FoV, providing visibility of the ascending aorta and aortic root.

Extensions in Comparison to Our Previous Work [10]:

4) *Deformable Model-to-Image Registration:* estimation of the deformation field describing the vessel deformation between the ROI projection and the segmented image using a DRU-Net model.

5) *3-D Model Registration:* application of the 2-D affine registration matrix and deformation field obtained from the previous modules to the preoperative 3-D model using a free form deformation (FFD) algorithm [24].

6) *Augmented Reality Visualization:* Visualization of the deformed 3-D model, along with the fluoroscopy image and deformation field, using the AR HMD device.

Our contribution to this extended framework introduced in comparison to the previous work [10] is depicted by the red dashed box in Fig. 1, including the model deformation prediction using a DRU-Net and the reconstruction of deformations onto the preoperative 3-D model. Furthermore, the visualization phase is enhanced by integration with AR. Detailed descriptions of these extensions will be provided in Sections II-B–II-D.

A. Image Dataset

For this study, a dataset was collected from nine patients who underwent TAVI procedures at the Centro Cardiologico Monzino (CCM) in Milan, Italy. The data collection process adhered to the ethical protocol approved by the CCM under the assigned code of *02_21 PA*.

The patient dataset consists of five males and four females, with an age of 81 ± 4. A collection of fluoroscopy images was obtained for each patient. The number of frames extracted varied according to the details presented in Table I. Before analysis, all images utilized in this study were resized to dimensions of 256×256. Subsequently, these images were segmented and aligned with the corresponding preoperative 3-D model via the previously proposed affine registration approach [10].

[1]Registered trademark.

TABLE I
NUMBER OF FLUOROSCOPY IMAGES EXTRACTED FROM EACH PATIENT

Patient No.	1	2	3	4	5	6	7	8	9
Age	84	85	82	85	85	75	75	79	80
Sex	M	F	F	M	F	M	M	F	M
Images	38	15	14	21	38	24	30	24	23

TABLE II
SUBDIVISION OF THE PATIENT IMAGE DATASET IN TRAINING,
VALIDATION, AND TESTING SETS

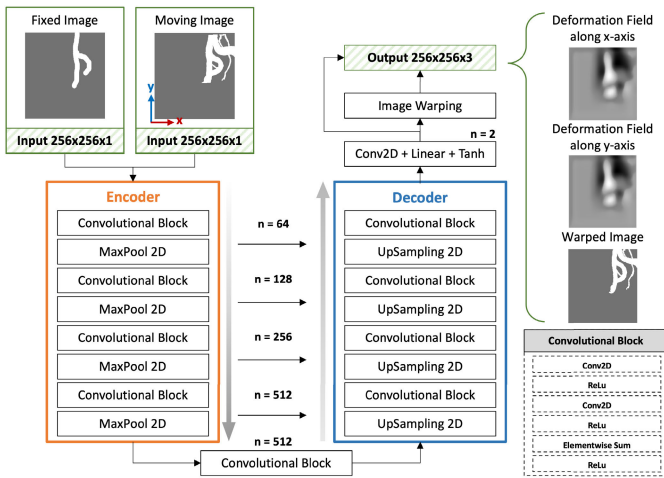| Phase	Dataset	Patient	Images
General Training ($W1$)	Training	1-7	180
	Validation	8	24
	Testing	9	15
Patient-specific Retraining ($W2$)	Training	9	6
	Validation	9	2
	Testing	9	15

Fig. 2. Sketch of the DRU-Net architecture for deformable model-to-image registration. Given a pair of input images, the DRU-Net model predicts deformation field components along the x- and y-axes, enabling the generation of a warped image by applying the deformation field to the moving image.

To estimate deformations using a DRU-Net model, the dataset was divided into training, validation, and testing sets, as depicted in Table II. Notably, two distinct training phases were introduced: ($W1$) the initial phase involved standard training to obtain a well-trained model, while ($W2$) the subsequent phase focused on rapid retraining utilizing intraoperative images obtained during the procedure for the specific patient. This personalized fine-tuning phase is proposed to enhance registration accuracy in patient-specific scenarios. The acquisition of intraoperative data, such as fluoroscopy images, and its integration into neural networks during procedures have been demonstrated as feasible in the existing literature, as exemplified by references such as [25] and [26].

B. Deformable Model-to-Image Registration (Module 4)

The DRU-Net architecture of the "deformable model-to-image registration" module is illustrated in Fig. 2. In this module, the fixed image (i.e., the intraoperative segmentation) and the moving image (i.e., the preoperative ROI projection), which have been previously aligned through the "affine

model-to-image registration" module [10], are concatenated and fed into the DRU-Net.

The DRU-Net encoder consists of four residual blocks. Each residual block is followed by a 2×2 max-pooling layer to reduce the number of network parameters. Within each residual block, there are two convolutional layers with a kernel size of 3×3, followed by a ReLU activation layer. The number of filters used in each block is denoted by n and indicated in Fig. 2.

In the DRU-Net decoder, each block consists of a 2×2 up-sampling layer followed by a residual block. Subsequently, a convolutional layer with 2 filters having a kernel size of 1×1 and a linear activation function is used, followed by a fully connected layer with a hyperbolic activation function that produces output values within the range of $[-1, 1]$. To ensure a meaningful range of deformation while preventing overfitting, these values are scaled by a factor h, which represents the maximum possible amplitude of the deformation field value [27], [28]. As a result, the output values are confined within the range of $[-h, h]$. The output of this fully connected layer is a two-channel image containing the deformation field components along the x- and y-axes. This deformation field is applied to the moving image and concatenated with the warped result, resulting in a single $256 \times 256 \times 3$ output. As shown in Fig. 2, the DRU-Net generates a three-channel image, consisting of a warped binary image and deformation field components along the x- and y-axes.

The model loss between the warped and fixed image is then calculated and utilized to update the neural network parameters during the subsequent training iterations. This iterative process allows the DRU-Net to optimize its performance and improve the accuracy of the deformable registration.

The network training utilizes an Adam optimizer [29] to minimize a customized loss function, which is a linear combination of two components

$$\mathcal{L} = \alpha \mathcal{L}_A + \beta \mathcal{L}_B. \tag{1}$$

Here, $\mathcal{L}$ represents the combined loss proposed in this work, while α and β are the weights assigned to each component. The values of α and β can be determined based on their respective importance within the loss function. $\mathcal{L}_A$ corresponds to a customized similarity loss, and $\mathcal{L}_B$ refers to the customized penalization of deformations in spare space (PDSS). The following paragraphs present comprehensive explanations and definitions of these two loss components.

To enhance the performance of the standard dice similarity coefficient (DSC) [30], we introduce the concept of modified dice similarity coefficient (MDSC). This approach involves partitioning both images into $N \times N$ subregions, thereby enhancing the performance of the traditional DSC metric in image similarity analysis. The formulation is given by

$$\text{MDSC} = \frac{2\text{TP}}{\sum_{i=1}^{N} \frac{a_i}{a_i + s} b_i + \frac{b_i}{b_i + s} a_i} \tag{2}$$

where true positive (TP) represents the number of corresponding white pixels (vessels) in both images. a_i and b_i represent the number of white pixels in the ith subregion of the fixed
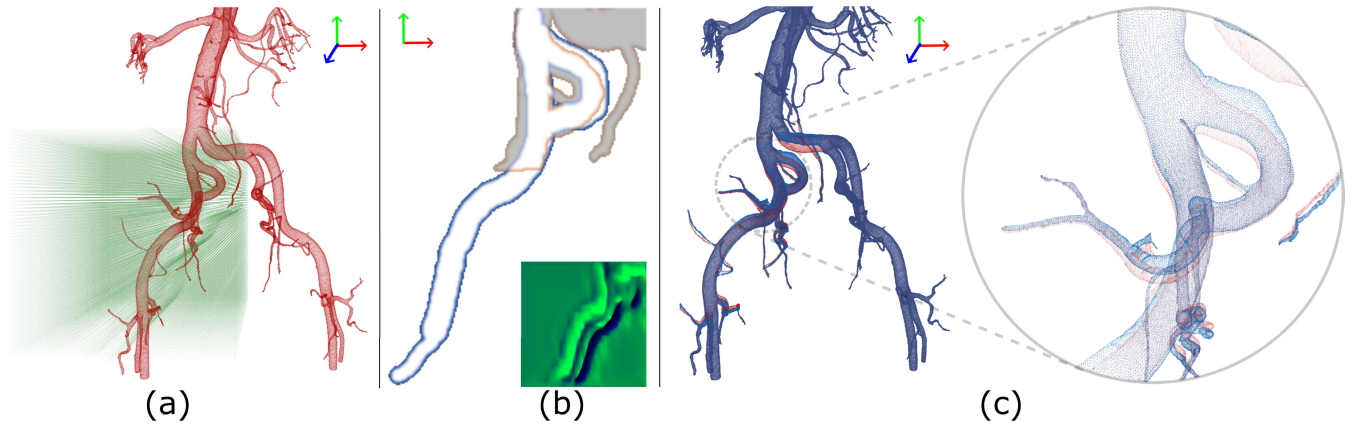
Fig. 3. Workflow illustrating the application of predicted deformation field on the preoperative 3-D model. (a) Evenly distributed control points (depicted in green) are generated within the ROI of the preoperative 3-D model (shown in red). (b) 2-D deformation field (the bottom image), predicted between the preoperative vessels (highlighted in red) and the intraoperative vessels (depicted in blue), is subsequently applied to the corresponding control points. (c) Resulting in a deformed 3-D mesh model (shown in blue).

and warped image, respectively, while s denotes a smoothing factor.

Compared to the traditional DSC, the improvement introduced in the denominator of MDSC reduces the dependence on vessel pixels that are absent in either the fixed and the warped images (where either a_i or b_i is equal to 0). This modification mitigates the adverse effects caused by incomplete vessel segmentation in fluoroscopy images, which has a substantial impact on the accuracy of the registration process. The loss component $\mathcal{L}_A$ in (1) is then defined as

$$\mathcal{L}_A = 1 - \text{MDSC}. \tag{3}$$

The loss component $\mathcal{L}_B$ is introduced to penalize deformations in sparse areas of the fixed image. To achieve this, both the fixed image and the deformation field in both x and y directions are divided into $N \times N$ subregions. $\mathcal{L}_B$ is then defined as follows:

$$\mathcal{L}_B = \text{PDSS} = \frac{1}{N^2} \sum_{i=1}^{N} \frac{\|1 - a_i\|}{ka_i + 1} \left(\frac{c_i + d_i}{2h}\right) \tag{4}$$

where c_i and d_i represent the maximum absolute values of the deformation field in the ith subregion along the x- and y-axes, respectively. The parameter k serves as an amplification factor.

The first part of the summation approaches 1 when the ith subregion of the fixed image is empty (i.e., $a_i = 0$) and approaches 0 otherwise. By doing so, it penalizes deformations in subregions of the moving image that do not correspond to any vessels in the fixed image.

C. 3-D Model Registration (Module 5)

The process of 3-D model registration involves reflecting the affine registration matrix and the 2-D deformation field onto the 3-D preoperative model. Specifically, the affine registration matrix obtained from Module 3, aligns the coordinate frames of the 3-D model and the fixed image. Subsequently, the estimated deformation field obtained from Module 4 is reconstructed onto the preoperative 3-D model using a FFD approach [24]. The details of the deformation reconstruction on the 3-D model are presented as follows.

To discretize the 3-D ROI in the preoperative model, an evenly distributed grid of size $m_x \times m_y \times m_z$ is employed. Each voxel corresponds to a control point [Fig. 3(a)] that applies a specific transformation in the three orthogonal axes based on its position. The number of control points m_x and m_y is set to be equal. The spacing between control points along the depth direction (z-axis) is set to be equal to the spacing in the other two directions.

Subsequently, the estimated deformation field obtained from Module 4 is applied to the control points [Fig. 3(b)]. The deformation between adjacent control points is determined using nonlinear cubic B-spline interpolation. We assume that the deformation field is consistent throughout the depth of the model, as the deformation information along the depth is not captured in fluoroscopy images. Thereafter, the resulting 3-D deformed model is illustrated in Fig. 3(c).

D. AR Visualization (Module 6)

As shown in Fig. 4(a), the 3-D visualization interface of the deformed 3-D model, along with the 2-D fluoroscopy image and deformation field, is developed using *Unity3D* under the support of Microsoft mixed reality toolkit (MRTK) [31]. Fig. 4(b) and (c) presents examples of the 3-D model visualization at two distinct time stamps. This interface application is deployed on the Microsoft HoloLens 2[1] AR device. Leveraging optical see-through display technology, digital elements are overlaid on real-world views with limited interaction.

III. EXPERIMENT AND VALIDATION

A. Experimental Setup

The experimental setup for modules 1–3 remains consistent with the configuration described in [10].

1) Deformable Model-to-Image Registration: The DRU-Net was implemented in Python using the *Tensorflow* and *Keras* frameworks and trained on an NVIDIA GeForce RTX2080Ti GPU card.

The learning rate of the model was determined using *KerasTuner*, a scalable hyperparameter optimization framework for conducting hyperparameter search [32]. The search
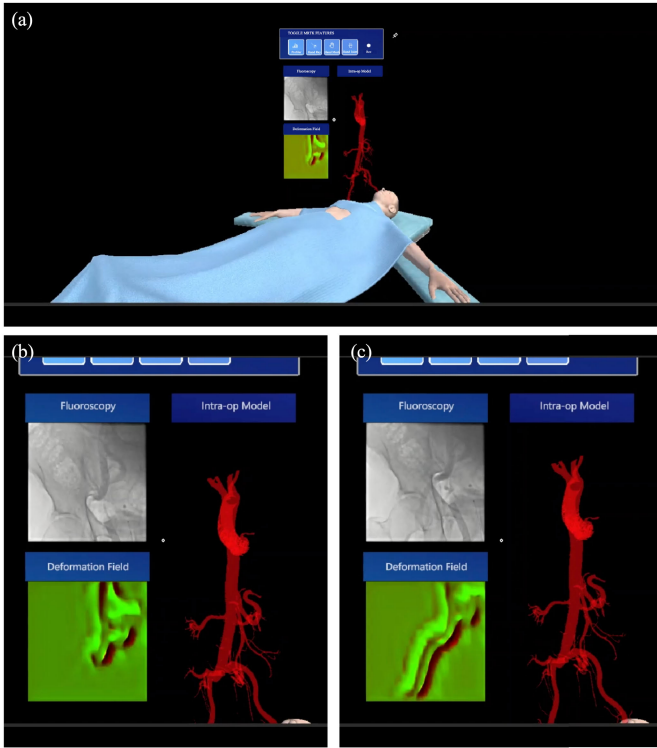
Fig. 4. Three-dimensional visualization interface. (a) Fluoroscopy image, the deformation field, and the deformed 3-D model. (b) and (c) Examples of 3-D model visualization at different time stamps.

process involved sampling values of the learning rate within the commonly used range $[1e-5, 1e-2]$ on a logarithmic scale. The optimal learning rate was selected based on achieving the lowest loss after a number of epochs (e.g., 20). The best learning rate was found to be within the range $[5e-5, 5e-4]$. A batch size of 4 was chosen to strike a balance between gradient estimation precision and iteration time. The training phases, $W1$ and $W2$, are each conducted for 30 and 20 epochs, respectively.

The weights in the loss function were set as $\alpha = 1$ and $\beta = 0.1$ to properly weigh the relative importance of image similarity and the penalization of spare space. The MDSC was computed using 16×16 subregions with $N = 16$ and a smoothing factor of $s = 1e-15$. The amplification factor k in the loss function was set to $1e6$, and $h = 15$ represented the maximum possible amplitude of the deformation field value.

2) 3-D Model Registration and Visualization: The 3-D model deformation is performed using the FFD approach, which is implemented utilizing the *PyGem* library [33].

The length of the ROI and the depth of the model in the corresponding volume were found to be similar across all patients. Consequently, when dealing with 256×256 images representing the estimated deformation field and $m_x = m_y = 256$, the number of control points surpassed the computational limitation of $1e6$, making precise deformations impractical. To address this constraint, the deformation field was downsampled to 64×64 using an averaging filter. This downsampling strategy reduced the number of control points to $m_x = m_y = 64$.

B. Registration Accuracy Validation

To validate the accuracy of registration in 3-D, a dataset was generated since ground-truth data for patients' intra-operative deformed mesh models was not available. This dataset was generated using a Position-based Dynamics (PBD) simulator [34]. External forces were applied at three different locations on the left and right femoral arteries in the simulator, resulting in three distinct groups of images and mesh models. The three mentioned locations represent typical regions that experience deformations due to catheter–vessel contact during procedures [27], [28].

These groups comprised 64, 46, and 33 pairs of images and mesh models, respectively. Subsequently, a split of 7:2:1 was employed to assign each group to the training, validation, and testing sets. The training set contained a total of 99 pairs of images and mesh models, while the validation and testing sets consisted of 27 and 17 pairs, respectively.

During the optimization process, the hyperparameter framework determined that the optimal learning rate for the DRU-Net was $2e-4$. Additionally, the maximum possible amplitude of the deformation field value was set to $h = 60$. The remaining parameters remained consistent with the description provided in Section III-A1.

C. Performance Metrics

1) Registration Accuracy in 2-D: The performance metrics employed to assess the quality of the deformable registration results included the DSC (5), Precision (6), Recall (7), MDSC (2), and PDSS (4)

$$\text{DSC} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{5}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

Here, TP represents the number of corresponding white pixels (vessels) in both images, false positive (FP) represents the number of pixels that are white in the first image (warped image) but black (background) in the second (fixed image), and false negative (FN) represents the number of pixels that are black in the first image but white in the second.

The first three metrics, namely DSC, Precision, and Recall, are commonly used [35]. The DSC provides a comprehensive assessment of the registration performance by considering both FP and FN, while Recall and Precision offer insights into the predominant type of matching error. On the other hand, the MDSC and PDSS metrics were specifically introduced in this study to address the challenges associated with deformable registration in partial regions.

2) Registration Accuracy in 3-D: To assess the accuracy of registration in 3-D, we employed the Mean Absolute Error (MAE) metric, which measures the discrepancy between the vertex positions of the deformed mesh model and the ground-truth positions of the dynamic mesh model obtained

from the simulator. The MAE values were computed for all vertices and specific axes, denoted as e for overall MAE and e_x, e_y, and e_z for MAE along the x-, y-, and z-axes, respectively,

$$e = \frac{\sum_{i=1}^{m} ||\boldsymbol{v}_i - \hat{\boldsymbol{v}}_i||}{m} \tag{8}$$

$$e_x = \frac{\sum_{i=1}^{m} |(\boldsymbol{v}_i - \hat{\boldsymbol{v}}_i) \cdot (1, 0, 0)|}{m}$$

$$e_y = \frac{\sum_{i=1}^{m} |(\boldsymbol{v}_i - \hat{\boldsymbol{v}}_i) \cdot (0, 1, 0)|}{m}$$

$$e_z = \frac{\sum_{i=1}^{m} |(\boldsymbol{v}_i - \hat{\boldsymbol{v}}_i) \cdot (0, 0, 1)|}{m}. \tag{9}$$

Here, m represents the total number of vertices in the mesh model. In the context of the deformed mesh model, $\boldsymbol{v}_i$ denotes the actual position of the ith vertex, representing the ground truth. Conversely, $\hat{\boldsymbol{v}}_i$ signifies the position of the ith vertex in the estimated mesh model, illustrating the approximation of the vertex's location. Moreover, we calculate the MAE specifically for the vertices within the ROI, which is referred to as e_{ROI}. Additionally, we calculate the MAE along specific axes within the ROI: $e_{x-\text{ROI}}$, $e_{y-\text{ROI}}$, and $e_{z-\text{ROI}}$, representing the MAE along the x-, y-, and z-axes, respectively.

3) Execution Time: Execution time, in the context of our study, pertains to the point at which the instructions within the computer programs or code are carried out.

To provide clarity on this matter, we distinguish between several key components: the execution time of the hyperparameter optimization phase, denoted as th, the execution time of the deformable model-to-image registration training phase, denoted as tr, and the execution time of the testing phase, denoted as te.

4) Statistical Significance: To assess the statistically significant differences between the method proposed in this study and the state-of-the-art approaches, the nonparametric Kruskal–Wallis test [36] was employed at a significance level of 0.05.

Two state-of-the-art approaches were chosen as comparative benchmarks: the traditional optimization-based Powell's method [37] and the state-of-the-art CNN approach [15]. It is worth noting that multiple optimization-based registration methods have been assessed in the work [37], with Powell's method, specifically Powell's conjugate direction method, consistently demonstrating reliability across various similarity measures. Consequently, it was chosen as a comparative benchmark.

Furthermore, the state-of-the-art CNN approach [15] was selected for comparison, considering the constraints of the available image dataset. In their work, Hu et al. [15] introduced a weakly supervised CNN for multimodal image registration. The primary distinctions between this existing state-of-the-art CNN and the CNN proposed in this study lie in the unique design of the loss function and the form of deformation output. Specifically, the network proposed in this study generates a deformation field as its output, whereas the model presented by Hu et al. [15] employs a dense displacement field (DDF) as its output mechanism.
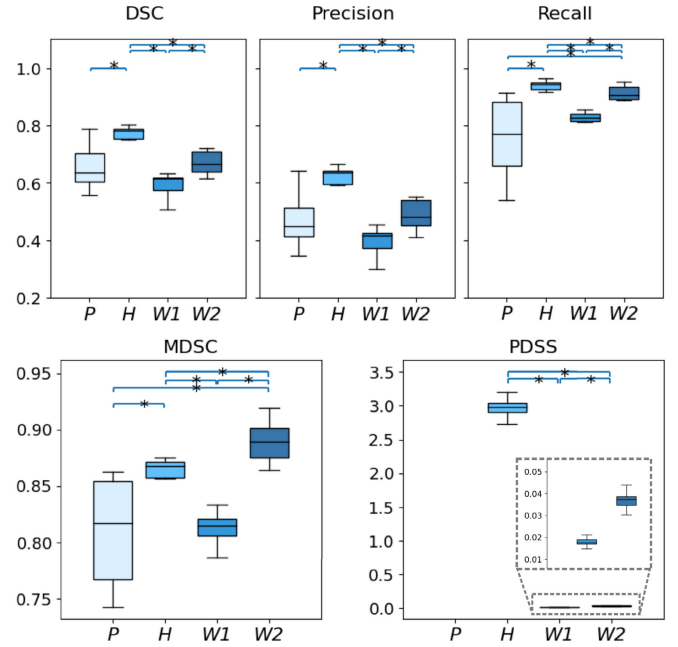

Fig. 5. Boxplots of the testing data depicting the distribution of performance metrics including DSC, Precision, Recall, MDSC, and PDSS. The label *P* represents the traditional optimization-based Powell's method [37]. The label *H* corresponds to the weakly supervised CNN proposed by Hu et al. [15]. The label *W*1 represents the proposed deformable registration after the general training phase, while the label *W*2 signifies the proposed deformable registration after the patient-specific retraining phase. Significance (*$p < 0.05$) was determined using the Kruskal–Wallis test. Note that a higher MDSC value indicates better performance, while a lower PDSS value indicates better performance.

IV. RESULTS AND DISCUSSION

A. Deformable Model-to-Image Registration

Fig. 5 presents the registration results of the patients' testing set, evaluated in terms of DSC, Precision, Recall, MDSC, and PDSS. We present results for four distinct approaches: the traditional optimization-based Powell's method [37] (abbreviated as *P*), the state-of-the-art CNN approach [15] (denoted as *H*), the DRU-Net model with general training only (*W*1), and the DRU-Net model enhanced through a patient-specific retraining phase (*W*2).

The testing results indicate that the retraining phase (*W*2) leads to a significant improvement across all performance metrics compared to the general training case (*W*1). Specifically, the mean and standard deviation values of DSC increased from 0.60 ± 0.06 to 0.68 ± 0.05, Precision improved from 0.40 ± 0.06 to 0.50 ± 0.07, Recall increased from 0.77 ± 0.14 to 0.85 ± 0.13, and MDSC improved from 0.81 ± 0.02 to 0.89 ± 0.02. However, PDSS slightly increased from 0.02 ± 0.00 to 0.04 ± 0.01. These findings suggest that incorporating a patient-specific retraining phase using intraoperative images can effectively enhance deformable model-to-image registration accuracy. It is noteworthy that the retraining process requires a relatively low number of epochs and minimal dataset extension, indicating the potential practicality of this technique in intraoperative applications.

In this specific image registration scenario, Recall emerges as the most reliable metric among the conventional metrics

(DSC, Precision, and Recall) for evaluating registration accuracy. This is attributed to Recall's dependency solely on the number of TP and FN in the predicted image, considering that the fixed image typically represents a partial segmentation of the vessels visible in the moving image. The relatively lower Precision values obtained align with the expected outcome, indicating that the majority of prediction errors correspond to FP value.

When compared with the traditional optimization-based Powell's method [37] (P), our proposed approach ($W2$) demonstrates superior performance across all performance metrics. Notably, we observe a significant enhancement, with a remarkable 18% improvement in the median value of Recall and a notable 9% improvement in the median of MDSC. It is essential to note that Powell's method is utilized to determine the optimal registration matrix, without an estimation of the deformation field. Consequently, computing the PDSS for Powell's method is less meaningful in this context.

In comparison to the state-of-the-art approach [15] (H), our proposed approach ($W2$) exhibits lower performance values in terms of DSC, Precision, and Recall. However, it is important to note that these metrics may not adequately capture the registration accuracy in the specific scenario characterized by vessels in the fixed image appearing in partial regions of the moving image only. Fig. 6 showcases registration examples using the traditional optimization-based Powell's method [37] (P), the state-of-the-art CNN approach [15] (H) and our proposed approach ($W1$, $W2$). Notably, our approach demonstrates improved performance in estimating deformations, particularly in regions where the vessels appear in the moving image but not in the fixed image. It is worth mentioning that due to the low contrast media dose, the fixed image derived from fluoroscopy usually contains only partial vessel branches present in the moving image. This incomplete segmentation introduces challenges to the stability of the deformable registration network, resulting in deformation artifacts in areas of projected vessels that do not exist in the fixed image.

The performance metrics MDSC and PDSS provide a more comprehensive representation of registration accuracy. As illustrated in Fig. 5, the DRU-Net model ($W2$) exhibits a significant improvement in MDSC accuracy compared to the results obtained through the literature [15] (H), with mean and standard deviation values of 0.89 ± 0.02 (compared to 0.86 ± 0.02). Moreover, the DRU-Net model achieves a reduced penalization loss in terms of PDSS, with mean and standard deviation values of 0.04 ± 0.01 (compared to 2.95 ± 0.15). Despite the incorporation of the component $\mathcal{L}_B$ in the loss function, residual artifacts persist even after the retraining phase. Notably, these artifacts become more apparent in regions where the fixed image contains smaller vessel sections, posing a concern as they can potentially provide incorrect guidance. It is therefore imperative to address this challenge by introducing a postprocessing step that targets the suppression of deformations outside the segmented vessel area and enhances deformation smoothness.

An additional limitation inherent to the application of our proposed method in medical contexts is its substantial
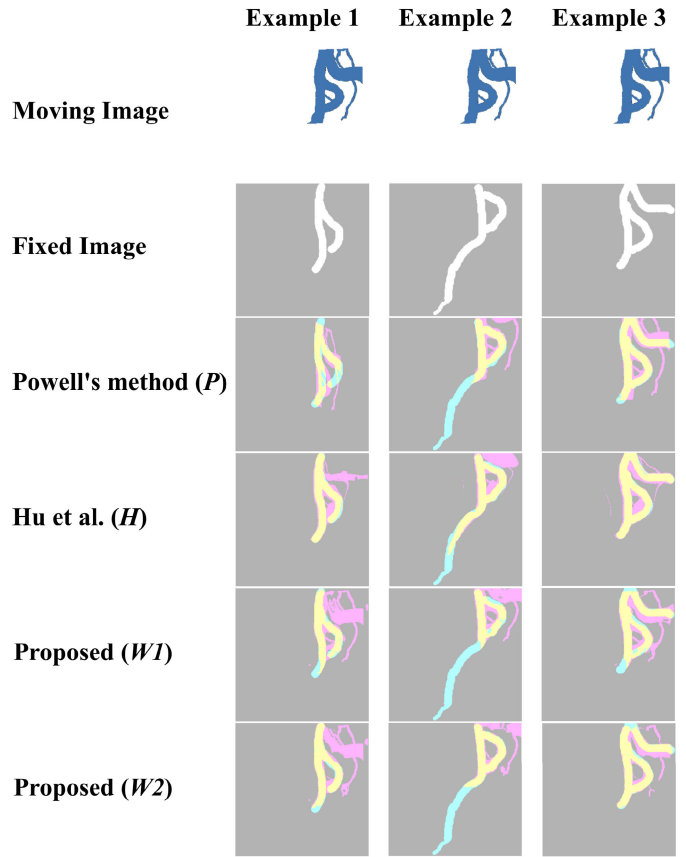

Fig. 6. Examples of model-to-image registration results obtained using the Powell's method [37] (P), the CNN method [15] (H), and the DRU-Net general training ($W1$) and retraining ($W2$). The last two columns display different elements of the confusion matrix between the fixed and warped images using distinct colors: yellow represents true positives (TP), gray represents true negatives (TN), cyan represents false negatives (FN), and magenta represents false positives (FP).

TABLE III
EXECUTION TIME ACROSS VARIOUS METHODS AND PHASES

Execution time	P	H	$W1$	$W2$
th [minutes]	-	32.2	60.7	-
tr [minutes]	-	7.4	17.5	2.4
te [seconds/image]	49.4	3.3	3.2	0.5

reliance on fluoroscopy images. The extraction of blood vessel deformations necessitates the use of intraoperative fluoroscopy images as primary data sources. Consequently, the accuracy of the deformation estimations is inherently contingent upon the quantity and frequency of injected contrast media. Variations in contrast media administration can introduce variability in image quality, potentially impacting the precision of deformation assessments. This dependence on contrast media may pose practical constraints in scenarios where such agents cannot be administered consistently or in desired quantities, thereby influencing the method's robustness in clinical applications.

Table III provides a summary of the execution time for different methods and phases. Notably, the traditional optimization-based Powell's method [37] (P) has a significantly longer average execution time per testing image compared to the learning-based approaches. After performing
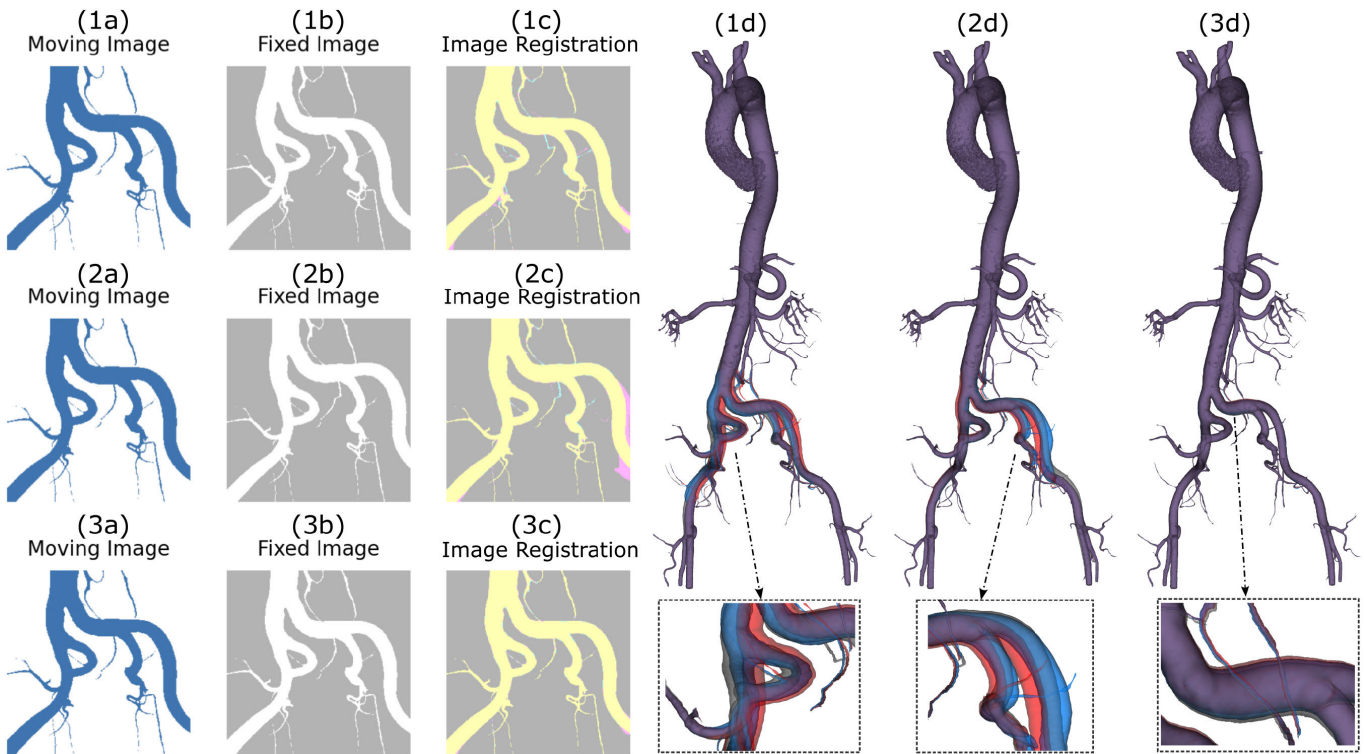
Fig. 7. Examples of model-to-image registration results using the dataset obtained from the intervention simulator. Starting from (a) moving image and (b) fixed image, (c) deformable model-to-image registration is performed, and (d) 3-D model registration and visualization are presented. In the third column (1c)–(3c), different elements of the confusion matrix between the fixed and warped image are represented by different colors: yellow indicates true positives (TP), gray indicates true negatives (TN), cyan indicates false negatives (FN), and magenta indicates false positives (FP). In (1d)–(3d), red represents the preoperative mesh model, blue represents the deformed mesh model predicted using the proposed approach, and gray represents the ground truth of the deformed mesh model.

the DRU-Net patient-specific retraining (which takes approximately 2.4 min), the average execution time of the testing phase is notably reduced to 0.5 s. This low computation time underscores the potential practicality and efficiency of our technique, particularly in intraoperative applications.

B. 3D Model Registration and Visualization

The deformed models provide improved accuracy and real-time representation of environmental changes during intraoperative procedures, benefiting both robotic systems and cardiologists. For a comprehensive demonstration of the 3-D visualization, we refer readers to the accompanying video[2] and Fig. 4. The visualization includes fluoroscopy images, estimated deformation fields obtained from the DRU-Net, and the corresponding deformed 3-D models. Consequently, the model-to-image registration facilitates 3-D visualization, enhancing visual guidance during procedures.

In our current methodology, we reconstruct 3-D models from intraoperative fluoroscopy images, which inherently capture the time-varying nature of the subject matter, specifically blood vessels. This temporal dimension, though implicit, effectively manifests as a form of 4D reconstruction and visualization [38], [39]. However, we acknowledge that our current neural network framework does not explicitly consider the temporal dimension. Future work could incorporate the

---

[2]https://youtu.be/3YbdejVkgzk

time factor into our model as [40] and [41]. This expansion would enable us to not only represent blood vessels in three spatial dimensions but also account for their dynamic changes over time. Such an enhancement could potentially contribute to predicting vessel deformation more accurately and ultimately lead to improved performance.

C. Registration Accuracy Validation

Table IV presents the results of the model-to-image registration performed on the dataset obtained from the intervention simulator [34]. The registration accuracy in 2-D exceeds 0.9, indicating excellent performance. The mean 3-D registration error for all vertices in the testing set is 0.39 mm, while for vertices within the ROI, it is 1.51 mm. The mean 3-D registration error observed within the ROI is notably higher compared to the average registration error across the entire dataset. This discrepancy can be attributed to the fact that the ROI is subject to a greater degree of deformations compared to regions that are more distant or less affected by such changes. The majority of the error arises from the x-y plane rather than the z-axis, indicating that our assumption of uniform deformation along the depth direction is reasonable. Fig. 7 showcases examples of model-to-image registration results. The preoperative mesh model (depicted in red) is transformed into an intraoperative mesh model (shown in blue) using the predicted deformation field obtained from our proposed DRU-Net approach. The similarity between the predicted

TABLE IV
PERFORMANCE METRICS ON THE DATASET OBTAINED
FROM THE INTERVENTION SIMULATOR

Performance Metrics		Testing
2D	DSC	0.97 ± 0.01
	Precision	0.95 ± 0.02
	Recall	0.98 ± 0.01
3D	e [mm]	0.39 ± 0.26
	e_x [mm]	0.26 ± 0.21
	e_y [mm]	0.19 ± 0.12
	e_z [mm]	0.09 ± 0.05
	e_{ROI} [mm]	1.51 ± 1.02
	$e_{x-\text{ROI}}$ [mm]	0.99 ± 0.84
	$e_{y-\text{ROI}}$ [mm]	0.79 ± 0.54
	$e_{z-\text{ROI}}$ [mm]	0.27 ± 0.14

deformed mesh model and the ground truth (represented in gray) confirms the feasibility and accuracy of the model-to-image registration. For a comprehensive recording of the 3-D visualization, refer to the accompanying video.[3]

In this study, the accuracy of the registration process is validated by applying a force near the femoral arteries in the simulator. This method effectively evaluates the precision of the registration under specific simulated conditions. However, this validation does not encompass the registration accuracy when considering deformations attributed to physiological factors such as heartbeat and pulsation. Future work could extend this validation process by incorporating these physiological factors into the simulator as described in [34]. This advancement would enable a more comprehensive assessment of the registration accuracy, particularly under varying conditions that mimic the realistic physiological environment of heartbeat and pulsation.

Moreover, it is important to note that the accuracy of the 3-D registration has not yet been validated using patient-specific image datasets due to the lack of ground truth, specifically intraoperative mesh models. This presents an opportunity for future research to explore and validate the accuracy of the proposed 3-D registration approach using such datasets, which would further strengthen the findings of this study.

D. Ablation Study

In our ablation study, we systematically evaluate the performance of the proposed loss function by comparing it with different loss configurations.

1) *T1:* The combined loss, as proposed in this work ($\alpha = 1$, $\beta = 0.1$).
2) *T2:* A loss function exclusively focused on PDSS, $\mathcal{L} = \beta \mathcal{L}_B$ (that is $\alpha = 0$, $\beta = 0.1$).
3) *T3:* A loss function exclusively focused on MDSC, $\mathcal{L} = \alpha \mathcal{L}_A$ (that is $\alpha = 1$, $\beta = 0$).
4) *T4:* The traditional dice loss ($\mathcal{L} = 1\text{-DSC}$).

The results of this study, presented in Fig. 8, reveal notable insights. The combined loss, when compared to the PDSS-focused loss, exhibits significant enhancements across all performance metrics, except for PDSS itself. Additionally, when juxtaposed with the MDSC-focused loss, the combined
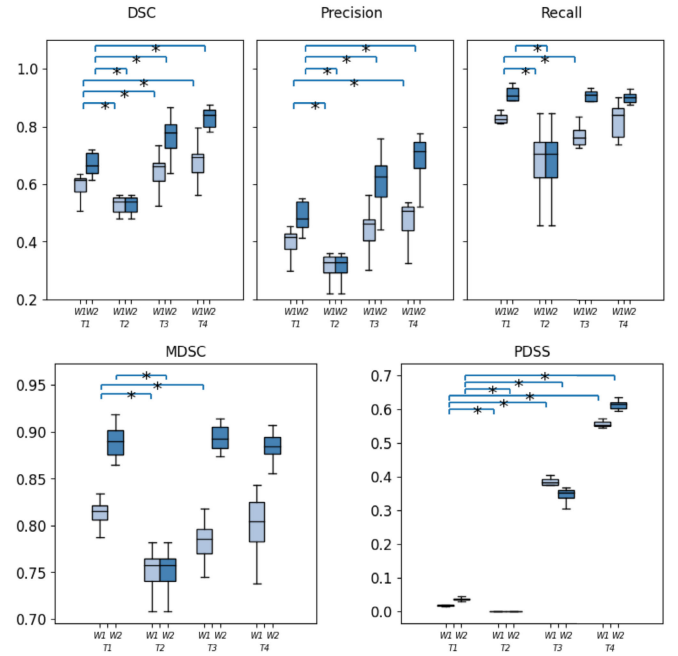
[3]https://youtu.be/2QFrwmTKIQs

Fig. 8. Boxplots illustrating the distribution of performance metrics in an ablation study conducted with varying losses: $T1$ ($\alpha = 1$, $\beta = 0.1$), $T2$ ($\alpha = 0$, $\beta = 0.1$), $T3$ ($\alpha = 1$, $\beta = 0$), and $T4$ ($\mathcal{L} = 1\text{-DSC}$). The comparison is made between the studies with altered parameters and the originally proposed study $T1$, assessing the significance of differences using the Kruskal–Wallis test (*$p < 0.05$). Note that a lower PDSS value indicates better performance.

loss showcases improvements in Recall, MDSC, and PDSS. Furthermore, in comparison to the traditional dice loss, the combined loss demonstrates improved performance in MDSC and PDSS. It is important to note that the combined loss effectively balances the tradeoff between maintaining vessel shape fidelity in the fixed image's spare space and ensuring accurate deformation field predictions in the rest space according to the fixed image.

V. CONCLUSION

In this study, we proposed a deformable model-to-image registration framework based on deep learning for augmented reality-guided endovascular interventions. The proposed framework encompasses several key components: 1) autonomous vessel segmentation of intraoperative fluoroscopy images through a DRU-Net; 2) affine model-to-image registration, achieved by employing a CNN to align the segmented images with the preoperative 3-D model reconstructed from CTA scans; 3) deformable model-to-image registration, accomplished by employing a DRU-Net model to predict and reconstruct deformations from 2-D images onto the preoperative 3-D model; and 4) an immersive visualization of intraoperative 3-D models using augmented reality. To provide a comprehensive evaluation of registration accuracy, we introduced a customized loss function and performance metrics, namely MDSC and PDSS.

This framework has the potential to assist clinicians during procedures by providing augmented reality visualization of patient-specific intraoperative vascular models. Our results

demonstrate improved accuracy and real-time representation of vascular changes compared to existing literature. The proposed DRU-Net approach achieved an increased MDSC value of 0.89 ± 0.02 (compared to 0.86 ± 0.02 in the literature) and a reduced PDSS value of 0.04 ± 0.01 (compared to 2.95 ± 0.15). It is important to note that the incorporation of a patient-specific retraining phase effectively enhanced deformable model-to-image registration accuracy.

To validate the registration accuracy in 3-D, we generated a dataset using an intervention simulator. The mean 3-D registration error for all vertices in the testing set was 0.39 mm, while for vertices within the ROI, it was 1.51 mm. The similarity between the predicted deformed mesh model and the ground truth further confirmed the feasibility and accuracy of the model-to-image registration approach.

Future work entails expanding the training set to enhance the robustness of the registration model, implementing post-processing techniques to address residual artifact suppression, and conducting end-user evaluations in the operating room. These endeavors will contribute to further advancements and practical application of the proposed framework in clinical settings.

Zhen Li is with the Department of Electronics, Information and Bioengineering, Politecnico di Milan, 20133 Milan, Italy, and also with the Department of Biomechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: zhen.li@polimi.it).

Letizia Contini and Elena De Momi are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: letizia.contini@polimi.it; elena.demomi@polimi.it).

Alessandro Maria Ippoliti is with Technische Universität Hamburg, 21073 Hamburg, Germany (e-mail: alessandro.ippoliti@tuhh.de).

Elena Bastianelli is with the University of Siena, 53100 Siena, Italy (e-mail: elena.bastianelli@mail.polimi.it).

Federico De Marco is with the Centro Cardiologico Monzino, IRCCS, 20138 Milan, Italy (e-mail: federico.demarco@gmail.com).

Jenny Dankelman is with the Department of Biomechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: j.dankelman@tudelft.nl).

REFERENCES

[1] A. Pore et al., "Autonomous navigation for robot-assisted intraluminal and endovascular procedures: A systematic review," *IEEE Trans. Robot.*, vol. 39, no. 4, pp. 2529–2548, Aug. 2023.

[2] L. Biasco et al., "Access sites for TAVI: Patient selection criteria, technical aspects, and outcomes," *Frontiers Cardiovascular Med.*, vol. 5, p. 88, Jul. 2018.

[3] S. R. Ravigopal, T. A. Brumfiel, and J. P. Desai, "Automated motion control of the COAST robotic guidewire under fluoroscopic guidance," in *Proc. Int. Symp. Med. Robot. (ISMR)*, Nov. 2021, pp. 1–7.

[4] M. H. D. Ansari et al., "Proof-of-concept medical robotic platform for endovascular catheterization," in *Proc. 11th Conf. New Technol. Comput. Robot Assist. Surgery (CRAS)*, 2022, pp. 66–67.

[5] M. Glöckler, J. Halbfaβ, A. Koch, S. Achenbach, and S. Dittrich, "Multimodality 3D-roadmap for cardiovascular interventions in congenital heart disease—A single-center, retrospective analysis of 78 cases," *Catheterization Cardiovascular Interventions*, vol. 82, no. 3, pp. 436–442, 2013.

[6] D. Toth et al., "3D/2D model-to-image registration by imitation learning for cardiac procedures," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 8, pp. 1141–1149, Aug. 2018.

[7] C. Heunis, J. Sikorski, and S. Misra, "Flexible instruments for endovascular interventions: Improved magnetic steering, actuation, and image-guided surgical instruments," *IEEE Robot. Autom. Mag.*, vol. 25, no. 3, pp. 71–82, Sep. 2018.

[8] C. Butter, H. Kaneko, G. Tambor, M. Hara, M. Neuss, and F. Hoelschermann, "Clinical utility of intraprocedural three-dimensional integrated image guided transcatheter aortic valve implantation using novel automated computed tomography software: A single-center preliminary experience," *Catheterization Cardiovascular Interventions*, vol. 93, no. 4, pp. 722–728, Mar. 2019.

[9] L. Zhao, S. Giannarou, S.-L. Lee, and G.-Z. Yang, "SCEM+: Real-time robust simultaneous catheter and environment modeling for endovascular navigation," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 961–968, Jul. 2016.

[10] Z. Li et al., "Model-to-image registration via deep learning towards image-guided endovascular interventions," in *Proc. Int. Symp. Med. Robot. (ISMR)*, Nov. 2021, pp. 1–6.

[11] Y. Zhang, L. Zhao, and S. Huang, "Aortic 3D deformation reconstruction using 2D X-ray fluoroscopy and 3D pre-operative data for endovascular interventions," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2393–2399.

[12] J.-Q. Zheng, X.-Y. Zhou, C. Riga, and G.-Z. Yang, "Towards 3D path planning from a single 2D fluoroscopic image for robot assisted fenestrated endovascular aortic repair," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, 2019, pp. 8747–8753.

[13] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Mach. Vis. Appl.*, vol. 31, no. 1, pp. 1–18, 2020.

[14] S. Guan, C. Meng, Y. Xie, Q. Wang, K. Sun, and T. Wang, "Deformable cardiovascular image registration via multi-channel convolutional neural network," *IEEE Access*, vol. 7, pp. 17524–17534, 2019.

[15] Y. Hu et al., "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Image Anal.*, vol. 49, pp. 1–13, Oct. 2018.

[16] L. T. De Paolis and V. De Luca, "Augmented visualization with depth perception cues to improve the surgeon's performance in minimally invasive surgery," *Med. Biol. Eng. Comput.*, vol. 57, no. 5, pp. 995–1013, May 2019.

[17] R. Rahman, M. E. Wood, L. Qian, C. L. Price, A. A. Johnson, and G. M. Osgood, "Head-mounted display use in surgery: A systematic review," *Surgical Innov.*, vol. 27, no. 1, pp. 88–100, Feb. 2020.

[18] S. Chidambaram et al., "Applications of augmented reality in the neurosurgical operating room: A systematic review of the literature," *J. Clin. Neurosci.*, vol. 91, pp. 43–61, Sep. 2021.

[19] V. García-Vázquez et al., "Navigation and visualisation with HoloLens in endovascular aortic repair," *Innov. Surgical Sci.*, vol. 3, no. 3, pp. 167–177, Oct. 2018.

[20] H. Iqbal, F. Tatti, and F. Rodriguez y Baena, "Augmented reality in robotic assisted orthopaedic surgery: A pilot study," *J. Biomed. Informat.*, vol. 120, Aug. 2021, Art. no. 103841.

[21] L. Qian, A. Deguet, and P. Kazanzides, "ARssist: Augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technol. Lett.*, vol. 5, no. 5, pp. 194–200, Oct. 2018.

[22] L. Qian et al., "FlexiVision: Teleporting the surgeon's eyes via robotic flexible endoscope and head-mounted display," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 3281–3287.

[23] P. Blanke et al., "Computed tomography imaging in the context of transcatheter aortic valve implantation (TAVI)/transcatheter aortic valve replacement (tavr) an expert consensus document of the society of cardiovascular computed tomography," *JACC, Cardiovascular Imag.*, vol. 12, no. 1, pp. 1–24, 2019.

[24] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[25] Y. Sakata, R. Hirai, K. Kobuna, A. Tanizawa, and S. Mori, "A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy," *Phys. Med. Biol.*, vol. 65, no. 8, Apr. 2020, Art. no. 085014.

[26] P. Fischer et al., "An MR-based model for cardio-respiratory motion compensation of overlays in X-ray fluoroscopy," *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 47–60, Jan. 2018.

[27] G. Koutouzi, M. Pfister, K. Breininger, M. Hellström, H. Roos, and M. Falkenberg, "Iliac artery deformation during EVAR," *Vascular*, vol. 27, no. 5, pp. 511–517, Oct. 2019.

[28] L. Cercenelli et al., "Characterization of vessel deformations during EVAR: A preliminary retrospective analysis to improve fidelity of endovascular simulators," *J. Surgical Educ.*, vol. 75, no. 4, pp. 1096–1105, Jul. 2018.

[29] T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman, "Project adam: Building an efficient and scalable deep learning training system," in *Proc. 11th USENIX Symp. Operating Syst. Design Implement.*, 2014, pp. 571–582.

[30] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[31] Microsoft. (2021). *Mixed Reality Toolkit—Unity*. [Online]. Available: https://github.com/Microsoft/MixedRealityToolkit-Unity

[32] N. K. Manaswi, "Understanding and working with keras," in *Deep Learning with Applications Using Python*. Berlin, Germany: Springer, 2018, pp. 31–43.

[33] M. Tezzele, N. Demo, A. Mola, and G. Rozza, "PyGeM: Python geometrical morphing," *Softw. Impacts*, vol. 7, Feb. 2021, Art. no. 100047.

[34] Z. Li et al., "Position-based dynamics simulator of vessel deformations for path planning in robotic endovascular catheterization," *Med. Eng. Phys.*, vol. 110, Dec. 2022, Art. no. 103920.

[35] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-Net ensembled for nuclei segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 208–212.

[36] P. E. McKight and J. Najab, "Kruskal-wallis test," *Corsini Encyclopedia Psychol.*, p. 1, Jan. 2010, doi: 10.1002/9780470479216.corpsy0491.

[37] I. Van Der Bom, S. Klein, M. Staring, R. Homan, L. W. Bartels, and J. P. Pluim, "Evaluation of optimization methods for intensity-based 2D-3D registration in X-ray guided interventions," in *Proc. SPIE*, vol. 7962, pp. 657–671, Mar. 2011.

[38] P. Rodríguez-Gonzálvez et al., "4D reconstruction and visualization of cultural heritage: Analyzing our legacy through time," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 609–616, Feb. 2017.

[39] M. Agbalessi et al., "Tracking of blood vessels motion from 4D-flow MRI data," *Cardiovascular Eng. Technol.*, vol. 14, no. 4, pp. 577–604, Aug. 2023.

[40] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger, "Occupancy flow: 4D reconstruction by learning particle dynamics," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5379–5389.

[41] V. Vishnevskiy, J. Walheim, and S. Kozerke, "Deep variational network for rapid 4D flow MRI reconstruction," *Nature Mach. Intell.*, vol. 2, no. 4, pp. 228–235, Apr. 2020.

Zhen Li (Graduate Student Member, IEEE) received the B.Eng. degree in intelligent science and technology from Nankai University, Tianjin, China, in 2017, the M.Sc. European master's degree in advanced robotics (EMARO+) from Warsaw University of Technology, Warsaw, Poland, and Ecole Centrale de Nantes, Nantes, France, in 2019, and the dual Ph.D. degrees in bioengineering and biomechanical engineering from Politecnico di Milano, Milan, Italy, and Delft University of Technology, Delft, The Netherlands, in 2023.

She was an ESR with the Marie Skłodowska-Curie ATLAS Project under the European Union Innovative Training Network.

Letizia Contini received the bachelor's degree in engineering physics and the master's degree in biomedical engineering from the Politecnico di Milan, Milan, Italy, in 2020 and 2022, respectively, where she is currently pursuing the Ph.D. degree in physics.

Her work is focused on the use of photonics in biomedical applications.
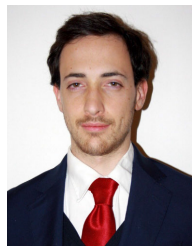
Alessandro Maria Ippoliti received the master's degree in biomedical engineering from the Politecnico di Milan, Milan, Italy, in 2022. He is currently pursuing the Ph.D. degree with Technische Universität Hamburg, Hamburg, Germany, under a Marie Skłodowska-Curie Actions scholarship for the Rewire project.

Elena Bastianelli received the master's degree in automation and control engineering from the Politecnico di Milan, Milan, Italy, in 2023.

She is currently a Research Fellow with the University of Siena, Siena, Italy, where she is going to start a Ph.D. program in Robotics and Intelligent Machines in November.

Federico De Marco received the Cardiology and Ph.D. degrees in cardiovascular physiopathology from the University of Milan, Milan, Italy, in 2005 and 2008, respectively.

He is the Director of the Interventional, Valvular and Structural Heart Cardiology, Centro Cardiologico Monzino, Milan. For over 16 years, he has been involved in endovascular interventions in the structural field, for which he is internationally recognized, with a particular interest in the treatment of heart valve disease. He has performed as a Lead or a Lead Operator in more than 5000 operations for structural heart disease.

Jenny Dankelman (Member, IEEE) received the master's degree in system and control engineering from the University of Groningen, Groningen, The Netherlands, in 1984, and the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 1989.

She is a Professor with the Minimally Invasive Surgery and Interventional Techniques (MISIT), Delft University of Technology. She became the Head of the MISIT Group. She cooperates with several hospitals, such as Leiden UMC, Leiden, The Netherlands; Erasmus MC Rotterdam, Rotterdam, The Netherlands; and the University Medical Center Amsterdam, Amsterdam, The Netherlands.

Dr. Dankelman became a Member of the Royal Netherlands Academy of Arts and Sciences in 2019. She was awarded the Antoni van Leeuwenhoek Chair in 2001. She has received the Royal Award of Knight in the Order of the Netherlands Lion.

Elena De Momi (Senior Member, IEEE) received the M.Sc. degree in biomedical engineering and the Ph.D. degree in bioengineering from the Politecnico di Milan, Milan, Italy, in 2002 and 2006, respectively.

She is currently an Assistant Professor with Politecnico di Milan, Milan, Italy.

Dr. De Momi has been an Associated Editor of IEEE ICRA, IROS, and BioRob, since 2016, and the Area Chair of MICCAI. She is currently the Publication Co-Chair of ICRA 2023. She is a PI for POLIMI of the EDEN2020 project and the ATLAS project, and a Co-PI of the ARTERY project. She has been an Evaluator and a Reviewer for the European Commission in FP6, FP7, and H2020.