Pose Estimation and Non-rigid Registration for Augmented Reality during Neurosurgery

Nazim Haouchine, Parikshit Juvekar, Michael Nercessian, William Wells III,
Alexandra Golby and Sarah Frisken

***Abstract*— Objective: A craniotomy is the removal of a part of the skull to allow surgeons to have access to the brain and treat tumors. When accessing the brain, a tissue deformation occurs and can negatively influence the surgical procedure outcome. In this work, we present a novel Augmented Reality neurosurgical system to superimpose pre-operative 3D meshes derived from MRI onto a view of the brain surface acquired during surgery. Methods: Our method uses cortical vessels as main features to drive a rigid then non-rigid 3D/2D registration. We first use a feature extractor network to produce probability maps that are fed to a pose estimator network to infer the 6-DoF rigid pose. Then, to account for brain deformation, we add a non-rigid refinement step formulated as a Shape-from-Template problem using physics-based constraints that helps propagate the deformation to sub-cortical level and update tumor location. Results: We tested our method retrospectively on 6 clinical datasets and obtained low pose error, and showed using synthetic dataset that considerable brain shift compensation and low TRE can be achieved at cortical and sub-cortical levels. Conclusion: The results show that our solution achieved accuracy below the actual clinical errors demonstrating the feasibility of practical use of our system. Significance: This work shows that we can provide coherent Augmented Reality visualization of 3D cortical vessels observed through the craniotomy using a single camera view and that cortical vessels provide strong features for performing both rigid and non-rigid registration.**

***Index Terms*— Augmented Reality, Image-guided Intervention, Pose Estimation, Registration, Neurosurgery**

I. Introduction

NEARLY all brain tumors are initially treated with surgical resection. The amount-of-resection is highly correlated with patients' chances of survival [1]. Indeed, brain tissues surrounding the tumor should be carefully considered, and complete resection must be balanced against the risk of causing new neurological deficits. Image-guided neurosurgical systems help register preoperative images to intraoperative patient coordinates so surgeons can view the locations of their surgical instruments relative to their surgical plan [2] [3]. They also optimize the extent of resection while avoiding damage to surrounding tissues structures. However, commercial systems rely on registration techniques based on patients' skin surface which leads to initialisation errors [4]. Indeed, using the skin surface [5] instead of the brain surface makes it difficult to find strong and reliable features to drive the registration given the shape of the head. In addition the skin surface may deform (particularly in older and obese patients) and registration may thus be innacurate. On the other hand, the brain surface has more structures, such as cortical vessels and sulcal folds that can be used to drive both rigid and non-rigid registration more effectively. In this paper, we present a novel Augmented Reality neurosurgical system to superimpose pre-operative 3D meshes onto a view of the brain surface acquired during surgery. Our method has two main advantages, first, it uses only a single-view image of the brain surface, thus does not need additional acquisition devices and 2) it automatically aligns the 3D models of the cortical vessels generated from preoperative imaging on the intra-operative view of the craniotomy in a non-rigid manner, taking into account initial brain deformation that occurs as a result of the craniotomy as illustrated in Figure 1.

Nazim Haouchine and Sarah Frisken are with Harvard Medical School and the Department of Radiology at Brigham and Women's Hospital, Boston, MA, USA (e-mail: nhaouchine@bwh.harvard.edu).

Parikshit Juvekar and Alexandra Golby are with Harvard Medical School and the Department of Neurosurgery at Brigham and Women's Hospital, Boston, MA, USA.

Michael Nercessian is with Cornell University, Ithaca, NY, USA, and the Department of Neurosurgery at Brigham and Women's Hospital, Boston, MA, USA

William Wells III is with the Massachusetts Institute of Technology, Harvard Medical School and the Department of Radiology at Brigham and Women's Hospital, Boston, MA, USA

II. Related Work

During neurosurgery the brain moves and deforms [6] [7], this deformation, also called brain shift, may considerably impact the surgical outcome since neurosurgical procedures are often based on pre-operative planning where brain shift is not considered. Many methods have been proposed to compensate for brain shift, either by using additional intra-operative imaging data [8] [9] [10] [11] or advanced brain models to predict intra-operative outcome pre-operatively [12] [13] [14]. In order to improve trajectory planning during deep brain stimulation, Essert *et al.* [14] proposed to learn from a database of brain shift cases to build a geometric template of possible pathways. In the same context, Bilger *et al.* [12] proposed a method to estimate a brain shift risk map using a simulation that gathers statistics on the displacement of anatomical landmarks from different clinical studies. A color-coded scheme based on vessel proximity is used to visualize this risk map pre-operatively. This method was later improved

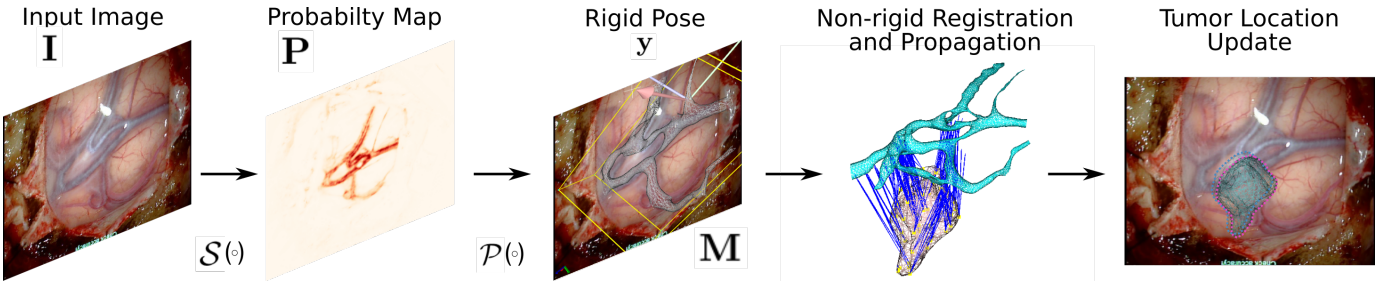


Fig. 1: Our pose estimation model is composed of a feature extractor $\mathcal{S}$ and a pose estimator $\mathcal{P}$. The feature extractor generates from the input image $\mathbf{I}$ a per-pixel probability map $\mathbf{P}$ to highlight the cortical vessels. The pose estimator uses the probability map to predict the 3D pose $\mathbf{y}$. The resulting pose $\mathbf{y}$ is used to overlay the pre-operative model $\mathbf{M}$ on the surgeon's view $\mathbf{I}$. A non-rigid refinement follows to estimate brain shift and updates tumor's position.

by Hamze *et al.* [13] using more advanced biomechanical models. Pre-operative simulation approaches have the benefit of being hardware-independent and do not involve any additional clinical step in the operating room. Nevertheless, they are built using case databases, or generic models that do not depict the actual events that occur during the procedure. Intra-operative data acquisition may be necessary to obtain a precise brain shift estimation.

In Bayer *et al.* [15], a non-rigid registration appraoch is used based on vessel centerliens extracted from preoperative and intreoperative C-Arm cone beam CT images. An interpolation warping function is then used as dense deformation field to compensate for brain shift. In Machado *et al.* [16], intra-operative ultra-sound images and pre-operative MRI are registered using a feature-based method. The extracted features are used to find correspondences between both image modalities and estimate dense mappings throughout the image. An alternative to the features extractor is to rely on user-selected landmarks in order to improve robustness in a clinical routine [17]. More advanced deformation models have been proposed to compute deformations. Using intraoperative doppler-based ultrasound, Morin *et al.* [18] measured brain shift at the brain surface and sub-surface vessels thanks to the use of physics-based models to propagate the deformation to regions of the brain where arteries are not present.

A craniotomy exposes the cortical brain surface and provides an additional source of information to capture intra-operative brain shift. Marreiros *et al.* [19] used three Near-Infrared cameras to capture brain surface displacement to register the cortical surface to MRI scans using a point-to-point registration method by considering vessel centerlines as strong matching features. A new deformed MRI volume is generated using a geometric interpolations. The approach presented by Luo *et al.* [20] uses an optically tracked stylus to identify cortical surface vessel features; a model-based workflow is then used to estimate brain deformation correction from these features after a dural opening. Jiang *et al.* [21] use phase-shifted 3D measurement to capture 3D brain surface deformations. This method highlights the importance of using cortical vessels and sulci to obtain robust results. The presence of a microscope in the operating room makes it very convenient to deploy such methods clinically. Sun *et al.* [22] proposed to pre-compute brain deformations to build an atlas from a sparse set of image-points extracted from the cortical brain surface. These image-points are extracted using an optically tracked portable laser range scanner. Haouchine *et al.* [23] proposed to use a finite element model of the brain to propagate cortical deformation captured from the microscope to substructures. This method uses cortical vessels to drive a non-rigid point-to-point registration. However, a manual rigid alignment is needed to initialize the system and obtain the point-to-point correspondences. Mohammadi *et al.* [24] proposed a projection-based stereovision process to map brain surfaces with a pre-operative finite element model. A predefined pattern is used to recover the 3D brain surface. In order to build a dense brain surface and gather more precise positions, Ji *et al.* [25] proposed a stereo-based optical flow shape reconstruction. The 3D shapes are recovered at different surgical stages to obtain undeformed and deformed surfaces. These surfaces can be registered to determine the deformation of the exposed brain surface during surgery. Nevertheless, most of the surgical microscopes need tedious pre-calibration which limits the scope of use of 3D stereo reconstruction.

Contributions and Motivations

Previous approaches have shown that it is possible to exploit the brain surface instead of the patient's skin surface to perform preop-to-intraop registration. However, these methods need manual registration to initialize the system. This cumbersome initialization is an obstacle to the adoption and deployment of such approaches in the clinical routine. We propose a novel approach to tackle this problem using an efficient approach to infer the 3D pose of the pre-operative mesh from its projection on a single-view image. In comparison to previous methods, we rely solely on a monocular single-view image acquired from the surgical microscope (or possibly a ceiling mounted camera) , to avoid tedious calibration of the stereo camera, laser range finder or optical stylus. Instead of using the raw image of the brain surface, we rely on probability maps that extract features that outline the vessels from its background. These features are fed into a deep neural network to infer a *6-DoF* rigid pose to permit a coherent mesh-to-image superimposition. To train this deep neural network, we synthesize a set of projected probability

maps from patients' 3D preoperative meshes, using various rigid transformations. Following rigid registration, we perform a 3D/2D non-rigid registration by satisfying a combination of physical and projective constraints using a shape-from-template formulation. The results demonstrate that our solution is fast, easy-to-use and achieves low errors, thus demonstrating the potential for improving surgical guidance.

III. Method

A. Overview

As illustrated in Figure 1, given a 3D mesh of the cortical vessels $\mathbf{M}$ and an input image of the brain surface $\mathbf{I}$, our solution recovers the 3D pose $\mathbf{y}$ corresponding to the camera rotation and translation required to align the mesh $\mathbf{M}$ onto the image $\mathbf{I}$, and a displacement field $\delta\mathbf{u}$ that updates the vertices of $\mathbf{M}$ and describes the brain shift deformation. Our method is a *three-stage* solution composed of: 1) a feature extractor neural network $\mathcal{S}$ that estimates, from the input image $\mathbf{I}$, a probability map that extracts the cortical vessels from the parenchyma; 2) a pose estimator network $\mathcal{P}$, that given the previously extracted probability map, estimates the rigid 3D pose (rotation and translation) and 3) a non-rigid refinement stage, that given the camera pose, will deform $\mathbf{M}$ in a physically-realistic manner. In order to train the network $\mathcal{S}$, images of the brain surface are synthesized using a style-enforced neural image analogy, while a mapping network is used to synthetize probability maps from random rigid projections of $\mathbf{M}$. Composing the networks $\mathcal{S}$ and $\mathcal{P}$ will result in a predicted pose $\mathbf{y}$ that is used to align $\mathbf{M}$ on $\mathbf{I}$. Once aligned, a force-based Shape-from-Template problem is solved to estimate $\delta\mathbf{u}$ and deform the mesh at cortical and sub-cortical levels.

B. Segmentation of Cortical Vessels from Brain Surface

Let us define the training set $T^{\mathcal{S}} = \{(\mathbf{I}_i, \mathbf{L}_j)\}_i$ composed of images $\mathbf{I}_i$ of the brain surface and their corresponding binary labels $\mathbf{L}_i$, with $i \in |T^{\mathcal{S}}|$ We first start by training the feature extractor $\mathcal{S}(\mathbf{I}; \theta_{\mathcal{S}})$ with $\theta_{\mathcal{S}}$ being the learned parameters for network $\mathcal{S}$. We use the feature extractor to generate a per-pixel probability map $\mathbf{P}$ that will associate with each pixel of an image $\mathbf{I}$ the probability of that pixel being part of the vessels or the parenchyma.

The network $\mathcal{S}$ is a deep CNN that follows a typical U-Net architecture [26] with a Softmax final layer. We optimize a *categorical cross-entropy* loss function using gradient descent over the parameters $\theta_{\mathcal{S}}$ following:

$$L(\theta_{\mathcal{S}}, T^{\mathcal{S}}) = \sum_{i \in |T^{\mathcal{S}}|} \left\| \mathcal{S}(\mathbf{I}_i; \theta_{\mathcal{S}}) - \mathbf{L}_i \right\|_2 \tag{1}$$

The model was trained over 94 epochs using mini-batches of size 8 with Adam optimizer and a learning rate of 0.001. In order to train $\mathcal{S}$ so that it generalizes well to new data, we needed to increase the amount of available training data $T^{\mathcal{S}}$ through data augmentation. We used neural image analogy [27] to generate an arbitrarily large and variable dataset of artificial images resembling real craniotomy images. This technique combines the style and texture of a source image with the high-level content representations of a target image. In our case we enforced the style on the source image over other features leading to generating many unique images that replicate the style of the craniotomy images. We randomly generated vessel-like binary images to mimic real craniotomy image labels using a procedural algorithm (see Figure 5). Then, using a limited set of real craniotomy images, we could generate a larger dataset composed of the pairs $\{(\mathbf{I}, \mathbf{L})\}$, which was then used to train $\mathcal{S}$.

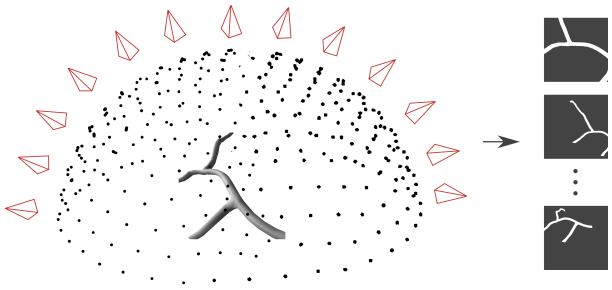


Fig. 2: Generating mesh projections. Using an input mesh $\mathbf{M}$, we generate binary images by randomly varying the 3D pose of a virtual camera (black dots) on the upper hemisphere of the 3D mesh.

C. 3D Pose Estimator

In order to align a 3D mesh model with its projection in a 2D image, we need to solve for its 3D pose (translation and rotation). In the general case, assuming a set of 2D coordinates $\mathbf{v} = \{v_i \in \mathbb{R}^2\}$ extracted from the image and their corresponding 3D coordinates $\mathbf{u} = \{u_i \in \mathbb{R}^3\}$, with $i \in |\mathbf{M}|$, solving for the 3D poses amounts at minimizing the reprojection error:

$$\sum_{i \in |\mathbf{M}|} \left\| \mathbf{A}[\mathbf{R}|\mathbf{t}]\mathbf{u} - \mathbf{v} \right\|_2^2 \tag{2}$$

where $\mathbf{R} \in \mathbb{R}^{3\times3}$ and $\mathbf{t} \in \mathbb{R}^3$ represent the 3D rotation and 3D translation, respectively, and $\mathbf{A}$ is the camera intrinsic matrix composed of the focal length and the principal points corresponding to the center of the image. Therefore, assuming a known focal length and using the quaternion representation for the rotation, there are altogether seven parameters to be estimated: rotations q_x, q_y, q_z, q_w and translations t_x, t_y, t_z.

In order to solve Eq. 2, we propose to train a pose estimator network $\mathcal{P}(\mathbf{P}; \theta_{\mathcal{P}})$ over the parameters $\theta_{\mathcal{P}}$. We use the synthetic training set $T^{\mathcal{P}} = \{(\mathbf{P}_i; \mathbf{y}_i)\}_i$, where $\mathbf{P}_i$ is the input probability map and $\mathbf{y}_k = [q_x, q_y, q_z, q_w, t_x, t_y, t_z]_i^{\mathsf{T}}$ is the output 3D pose; with $i \in |T^{\mathcal{P}}|$. This set is generated by creating binary projections of the 3D model $\mathbf{M}$ by randomly varying the 3D pose of a virtual camera (see Figure 2). Because the brain surface may deform due to brain shift [4], the preoperative model may differ from its expected image projection intraoperatively. We thus define a function $\mathcal{D} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that deforms the previously generated binary image. We can now apply, in addition to the rigid transformation, a set of non-rigid transformations to generate a set of binary

images representing the plausible projections of the model $\mathbf{M}$ during surgery.

As mentioned above, we use probability maps instead of binary images since in practice segmenting the cortical vessels will produce erroneous vessels/background classification that will lead to mis-alignments. In addition, binary images may discard meaningful information especially at the frontiers of classes that are present in the probability maps. We therefore train a mapper network $\mathcal{M}(\mathbf{L};\theta_{\mathcal{M}})$ with parameters $\theta_{\mathcal{T}}$ to translate binary images to probability maps. In order to train this network, we use the training set $T^{\mathcal{M}} = \{(\mathbf{L}_i;\mathbf{P}_i)\}_i$. This set makes use of the already available binary images used for the style-enforced analogy $\mathbf{L}_i$ and their corresponding inferences $\mathbf{P}_i$; with $i \in |T^{\mathcal{M}}|$. Putting all together, we can synthesize a new probability map following:

$$\mathbf{P}_i = \mathcal{D} \odot \mathcal{M}(\mathbf{\Pi}_{\text{proj}} \cdot \mathbf{M};\theta_{\mathcal{M}}) \tag{3}$$

where $\mathbf{\Pi}_{\text{proj}}$ is a virtual projection matrix that assembles the varying rotation and translation, in addition to pre-defined camera parameters. In practice, we only vary the rigid pose on the upper hemisphere of the 3D mesh to remain consistent with the potentiel camera positions w.r.t. patient's head during the craniotomy.

The network architecture of $\mathcal{P}$ consists of 3 hierarchies composed of two convolutional and one ReLU layers. To decrease the spatial dimension, an average pooling layer with a stride of 2 follows each hierarchy except the last one. At the end of the last hierarchy we add one fully-connected layer with 128 neurons and ReLU activation followed by one fully-connected with 7 neurons with a linear activation. We finally optimize the following loss function over the parameters $\theta_{\mathcal{P}}$ of the network $\mathcal{P}$:

$$L(\theta_{\mathcal{P}}, T^{\mathcal{P}}) = \sum_{i \in |T^{\mathcal{P}}|} \left\| \mathcal{P}(\mathbf{P};\theta_{\mathcal{P}}) - \mathbf{y}_i \right\|_2 \tag{4}$$

The model was trained over 142 epochs using mini-batches of size 8 with Adam optimizer and a learning rate of 0.001. Finally, at run-time, given an image $\mathbf{I}$ we predict the corresponding pose $\mathbf{y}$ by composing the networks $\mathcal{S}$ and $\mathcal{P}$ together:

$$\mathbf{y} \leftarrow \mathcal{P}(\mathcal{S}(\mathbf{I};\widehat{\theta_{\mathcal{S}}});\widehat{\theta_{\mathcal{P}}}) \tag{5}$$

where $\widehat{\theta_{\mathcal{S}}}$ and $\widehat{\theta_{\mathcal{P}}}$ are the resulting parameters from the training.

D. 3D Vascular Network Modelling

In order to simulate brain deformation, we build from the pre-operative vessels geometry a mechanical model capable of handling 3D non-linear deformations in a consistent manner. We rely on a wire-like model to represent the behaviour of cortical vessels. This model has previously been used to simulate guide wires and catheters [28] and vessels [23]. In order to build this model, we first extract vessel centerlines from the 3D mesh $\mathbf{M}$, and build the model by connecting the 3D nodes $\mathbf{u}$ along the centerline in sequence. Each beam element is delimited by two nodes each having 3 spatial and angular degrees of freedom (see Figure 3-a). A stiffness matrix $\mathbf{K_e}$ is then defined to relate the node positions to the forces applied to them accounting for rotational degrees of freedom through a rotational matrix $\mathbf{Q_e}$. At each node i, the internal forces $\mathbf{f}_i$ generated by the deformation of the structures is formulated as:

$$\mathbf{f}_i = \sum_{e=i-1}^{i} \mathbf{Q}_e(\mathbf{u})\mathbf{K}_e(\mathbf{Q}_e(\mathbf{u})^T(\mathbf{u}-\mathbf{u}_e) - \mathbf{u}^{rest}) \tag{6}$$

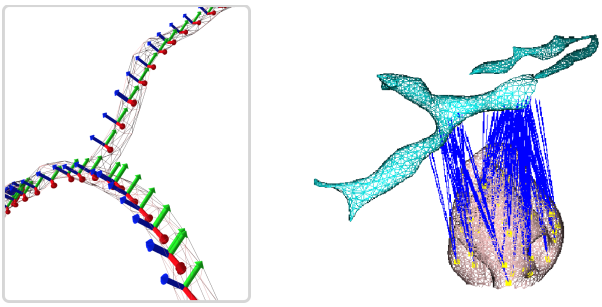
where e is the index of the two beams connected to this i^{th} node. $\mathbf{u}_{i-1}$, $\mathbf{u}_i$ and $\mathbf{u}_{i+1}$ are the degrees of freedom vectors of the three nodes (respectively $i-1$, i, $i+1$) and belong to the two beams in the global frame. $\mathbf{u}_{ej}$ denotes the middle frame of the j^{th} beam that is computed as an intermediate between the two nodes of the beam, and $\mathbf{u}^{rest}$ corresponds to the degrees of freedom at rest in the local frame.

Furthermore, in order to propagate the deformation from cortical to sub-cortical level, and more particularly to tumors, we rely on a linear geometrical barycentric function $\phi(\cdot)$ that can act as a mapping between the surface and the underlying tissue (see Figure 3-b;c).

Formally speaking, if we denote $\mathbf{r}$ the vector of vertices of the tumor model, we can express each vertex $\mathbf{r}_i$ using barycentric coordinates of facet vertices $\mathbf{u}$, such that

$$\mathbf{r}_i = \sum_{j=1}^{3} \phi_j(x_i, y_i, z_i)\mathbf{u}_j \tag{7}$$

where $\phi(x,y,z) = a + bx + cy$ with (a,b,c) being the barycentric coordinates of the triangle composed of nodal points $\mathbf{u}_j$, with $1 \leq j \leq 3$. Since craniotomy are performed to access structures beneath the opening region, this mapping is sufficient to propagate deformations to the immediate sub-cortical structures.



(a) (b)

Fig. 3: 3D vascular model and mechanical mapping. (a) Serially-linked beams on the mesh centerlines. Each beam element (represented by the frames) is delimited by two nodes that relate to the forces applied to them w.r.t a stiffness matrix. The wireframe mesh represent the vascular network. In (b), a visualization of the mapping mechanism with blue lines represents the attachments.

E. Force-based Shape-from-Template

Estimating a 3D deformed shape from a single projection knowing the 3D shape in rest configuration is in an ill-posed

problem also known as Shape-from-Template [29]. Its ill-posedness comes from its under constrained nature, where many 3D shapes may have identical 2D projections. In order to solve this problem, we can add constraints to reduce the space of plausible solutions. We chose to rely on physical constraints to force the 3D estimate shape to follow the laws of mechanics according to the beams model introduced in Eq. 6. In our case, the 3D shape at rest configuration consists of the pre-operative 3D centerlines $\mathbf{u}$ while its 2D projection consists of a set of points $\mathbf{v}$ extracted from the probability map along the vessel centerlines. The unknown 3D deformed shape $\delta\mathbf{u}$ is the displacement field induced by a potential brain shift. Using the previously estimates pose and assuming a known camera parameters $\mathbf{A}$ from pre-calibration, we can build the projection matrix $\mathbf{\Pi} = \mathbf{A}[\mathbf{R}|\mathbf{t}]$. Using $\mathbf{\Pi}$ we can establish a set of correspondences $\mathbf{c}$ so that if a 2D point $\mathbf{v}_i$ corresponds to a 3D point $\mathbf{u}_j$ then $\mathbf{c}_i = j$ for each point of the two sets. Deforming the 3D vessels so they fit their corresponding 2D projections amounts to minimize the re-projection error $\|\mathbf{\Pi}\mathbf{u}_{c_i} - \mathbf{v}_i\|$ for $i \in n_c$ while ensuring minimal internal forces $\mathbf{f} = \sum_i^{n_c} \mathbf{f}_i$ emanating from the vascular mechanical model. Combining these two constraints leads to the following force minimization expression:

$$\underset{\delta\mathbf{u}}{\operatorname{argmin}} \sum_{i}^{n_c} \left(\mathbf{f}_{c_i} - \kappa \|\mathbf{\Pi}\mathbf{u}_{c_i} - \mathbf{v}_i\| \right)^2 \tag{8}$$

where κ is the stiffness coefficient that permits translation of the image re-projection error to an image bending force. Note the subscript c_i that denotes the correspondence indices between the two point sets. This minimization can be seen as enforcing the 3D vessel centerlines to fit the sightlines that come from the camera position to the 2D vessels centerlines while maintaining a coherent 3D shape. It is computationally inexpensive to solve when formulated as linear complementary problem and solved using a Gauss-Seidel algorithm [23].

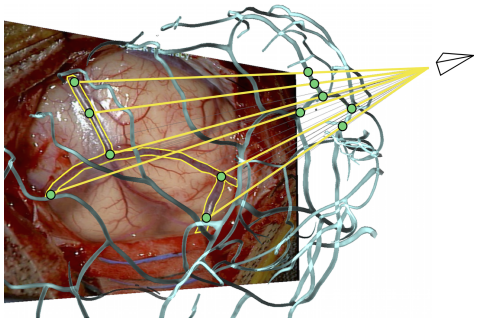
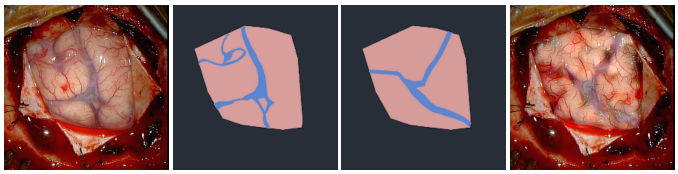


Fig. 4: Force-based Shape-from-Template, problem formulation: we aim at recovering the deformed 3D vessels shape $\delta\mathbf{u}$ from its known reprojection in the image $\mathbf{v}$, the known pre-operative 3D vessels at rest $\mathbf{u}$ and known rigid alignment, satisfying physical and reprojective constraints.

IV. Results

A. Dataset

To build the training set $T^{\mathcal{S}}$ we generated 1856 images using neural image analogy from 10 craniotomy style source images and 71 annotated inputs. These images were randomly split and 25% of the images were kept for validation. The model achieved an 0.71 IoU and 0.82 Dice accuracy on the validation set. Examples of probability maps generated using $\mathcal{S}$ are shown in Fig. 7.



(a) Source (b) Source labels (c) Target labels $\mathbf{L}$ (d) Target image $\mathbf{I}$

Fig. 5: Samples from dataset $T^{\mathcal{S}}$ used for training the segmentor network $\mathcal{S}$.

We tested our method retrospectively on six clinical datasets (see Fig. 7). These consisted of T1 contrast MRI scans and images of the brain surface after dura opening. The cortical vessels and the tumors were segmented and triangulated to generate 3D meshes using *3D Slicer* (www.slicer.org). Only a subpart of the cortical vessels is used. The microscopic images (left image of the stereoscopic camera) were acquired with a Carl Zeiss surgical microscope. The ground-truth pose is obtained by manually aligning the 3D meshes on their corresponding craniotomy images.

To build the training set $T^{\mathcal{P}}$ we generated around 1400 synthetic probability maps per case. We randomly sampled rotation and translation on the upper hemisphere of the 3D mesh. These images and poses were randomly split where 25% were kept for validation.

B. Pose Estimation

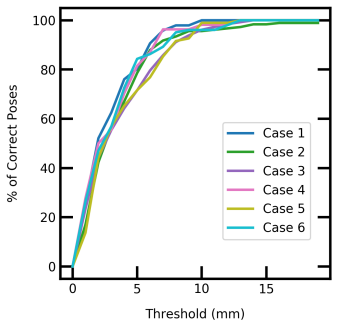


Fig. 6: Accuracy-threshold curves with average distance error on all validation sets.

We chose the average distance metric (ADM) as proposed in [30] for evaluation. Given a set of object's 3D vertices, the ground-truth and estimated rotations and translations, the ADM computes the mean of the pairwise distances between the 3D model points transformed. Using this metric, a pose is considered correct if the average distance is lower than a distance threshold (usually 10% of the object's diameter).

We first vary the distance threshold on the validation sets of the dataset in order to reveal how our method performs w.r.t that threshold. We plotted an accuracy-threshold curve showing the ADM variation with a threshold in a range of 0

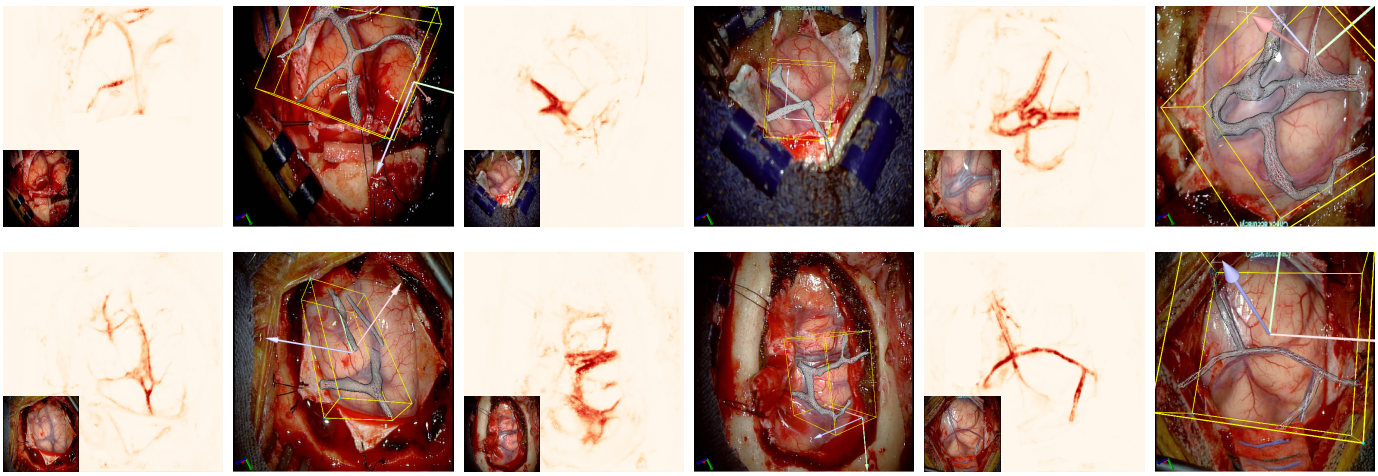


Fig. 7: Retrospective results on six cases. For each case the probability map (left) and the estimated pose are shown (right). The input images are displayed in the bottom left corner of the probability maps. Cases 1 to 6 are ordered from top to bottom, left to right.

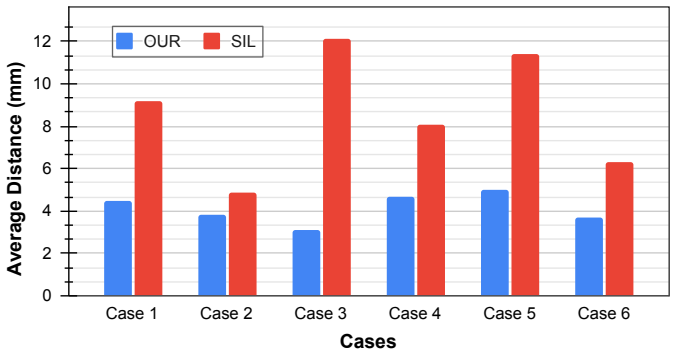


Fig. 8: Comparison between our method (**Our**) and the silhouette-based method (**SIL**) for each case.

mm to 20 mm. We can see in Figure 6 that for most of the sets, a 80% pose accuracy is reached with a threshold below 5 mm, except for case 2 and 3.

We then conducted experiments to analyze the effect of using probability maps. We compared our method (**Our**) with a baseline silhouette-based approach (**SIL**). The later approach estimates the pose using a binary image that corresponds to the silhouette of the 3D object. We report in Figure 8 the distance between the ground-truth and estimated pose where we can notice that our method outperforms the silhouette-based approach with average ADM errors of 4.13 ± 0.70 mm compared to 8.67 ± 2.84 mm. These errors should be compared to those of the study in [4], which reports a 5.26 ± 0.75 mm initialization error measured on 15 craniotomy cases when using commercially-available systems. Figure 7 shows our results as bounding boxes and overlaid meshes. Our methods perform well with case 3, 4, 5 and 6 with visually consistent alignements. Nevertheless, it should be noted that our method may produce misalignment errors as in case 1 and 2. This suggests that our model does not properly handle occlusions (case 1) and that the pose accuracy may decrease if the cortical vessels are only partially visible (case 2). However, in all cases, our predictions remain always close to the ground-truth.

From the computation time standpoint, it takes 2.07 seconds to extract the probability map and 1.23 seconds to predict the pose. We used the Tensorflow/Keras framework (www.tensorflow.org) on an NVidia GeForce GTX 1070 without any specific optimization.

C. Non-rigid Refinement

To evaluate the non-rigid refinement we first used simulated data of a 3D synthetic human brain. Using a mean curvature skeletonization technique we extract 3D vessel centerlines from the vessels mesh surface. The wire-like model is built upon the centerlines nodes with a subset nodes in range between 12 and 30 nodes. The mechanical parameters of the vessels are taken from a textbook [31] with a Young's modulus set to 0.6 MPa and Poisson's ratio to 0.45 to simulate a quasi-elastic and incomprehensible behaviour where vessels thickness is set to 0.5 mm These parameters are used to assemble the stiffness matrix of the vessels that will describe its elasticity. We used the Sofa framework (www.sofa-framework.org) to build the 3D beam-elements model.

To measure how the cortical deformation is propagated to sub-cortical levels, we measured the target registration error (TRE) on different locations. We also quantify the percentage of brain-shift compensation by measuring the difference between the initial and the corrected error (normalized by the initial error). This measure is important since the TRE depends on the amount of brain shift. We randomly placed 4 landmarks inside the simulated brain at different depths from 0 mm to 45 mm, 0 mm being the cortical level. The simulation of brain shift consisted of a protrusion deformation of 4 increasing amplitudes to mimic brain swelling after the craniotomy, Def1 $= 1.7$ mm, Def2 $= 4$ mm, Def3 $= 6$ mm and Def4 $= 10$ mm. We projected the deformed 3D centerlines w.r.t a known virtual camera to obtain the images and added

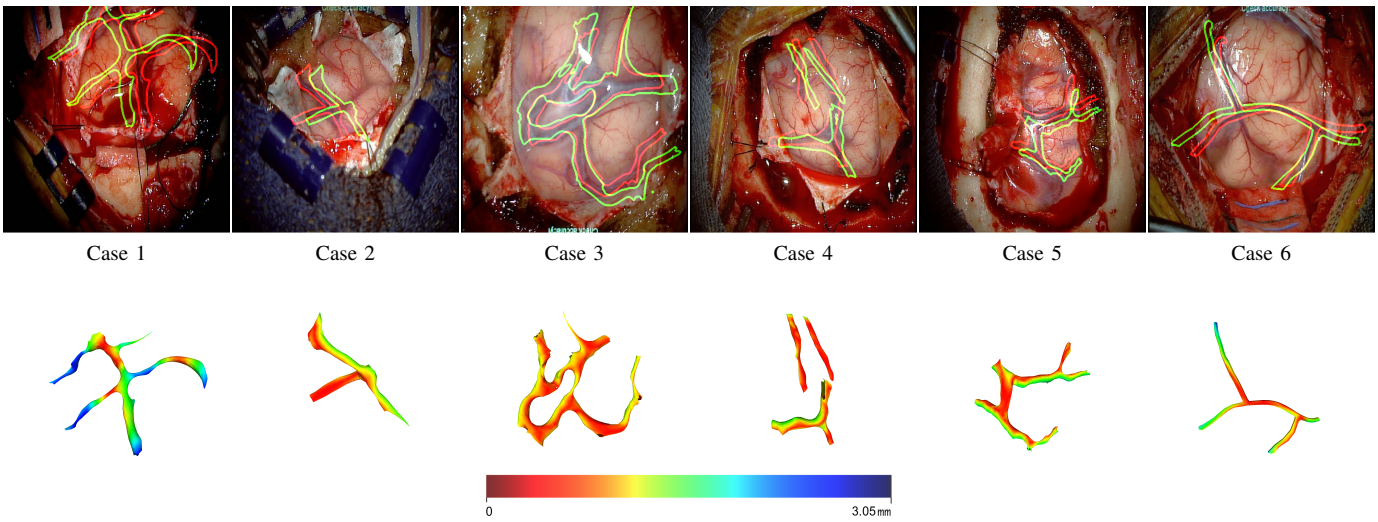


Fig. 9: The **first** row shows qualitative results obtained on our dataset retrospectively. The contours correspond to the pre-operative 3D meshes projected using the non-rigid refinement (red), and the estimated poses (green). Our predictions remain always close to the ground-truth. The **second** row shows a color-coded estimated stress on the vessels.

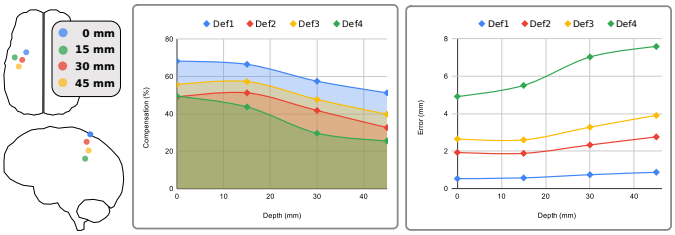


Fig. 10: Quantitative evaluation on simulated data: Charts (a) and (b) show the percentage of compensation and TRE respectively w.r.t to depth and brain shift amplitude. 0 mm in the "depth" axis represents the cortical level.

Gaussian noise with a standard deviation of 5 mm and a 5 clustering decimation on the set of 2D points extracted from the images.

We can notice in Figure 10 that using our method achieves small TRE ranging from 0.53 mm to 1.93 mm, on the cortical level and the immediate sub-cortical structures ($\leq$ 15 mm). Brain shift compensation of up to 68.2 % is achieved, and at least a 24.6 % compensation is reported for the worst configuration. We can also observe that errors and compensation at the cortical level is very close to the immediate sub-cortical location ($\leq$ 15 mm). It increases when the targets are locate deeper in the brain ($\geq$ 30 mm) or the amount of deformation increases ($\geq$ 6 mm). Finally, we show in Figure 9 the measured deformation amplitude on each vessel after non-rigid registration with color-coded stress visualization as well as the projection of the deformed vessels on the image.

V. Limitations

a) Robustness to Occlusions: As reported in the results, our method is limited when the 3D mesh is partially visible, as seen in case 1 and 2. Expanding the training set with occluded cases would permit handling of occlusions properly. Moreover, a minimal amount of visible vessels could naturally be considered as a hard condition for the success of our solution. Another possibility is to combine the vessels with the sulci which are more likely to be visible during a craniotomy and can also be segmented from the MRI scans.

b) Microscopic Camera Focal Changes: Although a camera calibration is a well-established technique to build the camera parameters $\mathbf{A}$, during a craniotomy, the focal length of the microscope may change and the pose estimation may be inaccurate. Including the camera focal length in the prediction can be done by adding a new element to the prediction vector $\mathbf{y}$ and training the model to take into account these focal length changes. This future work is aligned with our objective to make such a technique easy to use in the operating room.

c) Error Propagation: Although we reported errors below actual clinical practice, neurosurgery requires very accurate registration. The proposed solution does integrate potential brain deformations but solely as a refinement step. Handling deformations can create ambiguities when estimating the pose thus leading to registration errors. In addition, an inaccurate rigid registration will undoubtedly lead to an inaccurate non-rigid refinement. Our way to reduce error propagation consists of building a full model capable of predicting an n-DoF pose, accounting for rigid and non-rigid components simultaneously.

VI. Conclusion

We proposed a fully automated brain-shift aware Augmented Reality method for craniotomy guidance. Our method first performs a rigid alignment of the 3D pre-operative mesh onto a single-view image of the brain surface, then refines the alignment by deforming the 3D mesh to compensate for brain shift. The rigid registration removes the curbesome step of manually aligning a 3D mesh onto its projection

on an image in the context of craniotomy, while the non-rigid refinement step permits updating the underlying tumor's location. Restricting our method to a single-view image makes it more adequate to operating rooms but turned the registration process into an ill-posed problem that we solved using an elegant end-to-end solution.

The proposed method was evaluated both quantitatively and qualitatively on a clinical dataset where we demonstrated that our method is fast and can achieve ADM errors below current clinical margins. In addition, we showed that low TRE can be obtained at cortical and sub-cortical levels and large brain shift compensation can be achieved. We strongly believe that this method is adequate for the use in the operating room, thanks to its high degree of automation and low level of surgeon interaction.

ACKNOWLEDGEMENTS

This work was supported in part by the National Institute of Health under Grant R01 EB027134-01, R01 NS049251 and BWH Radiology Department Research Pilot Grant Award. Human datasets were used in this work. Approved, recognized procedures were followed for human subject inclusion.

REFERENCES

[1] J. M. Gonzalez-Darder, *State of the Art of the Craniotomy in the Early Twenty-First Century and Future Development*. Cham: Springer International Publishing, 2019, pp. 421–427.

[2] J. F. Fraser, T. H. Schwartz, and M. G. Kaplitt, *BrainLab Image Guided System*. Springer Berlin Heidelberg, 2009, pp. 567–581.

[3] R. Bucholz and L. McDurmont, *The History, Current Status, and Future of the StealthStation Treatment Guidance System*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 543–565.

[4] S. Frisken, M. Luo, P. Juvekar, A. Bunevicius, I. Machado, P. Unadkat, M. Bertotti, M. Toews, W. Wells, M. Miga, and A. Golby, "A comparison of thin-plate spline deformation and finite element modeling to compensate for brain shift during tumor resection," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, 08 2019.

[5] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. Sinclair, and D. Collins, "Augmented reality in neurovascular surgery: feasibility and first uses in the operating room," *Int J Comput Assist Radiol Surg*, vol. 10(11), pp. 1823–36, 02 2015.

[6] S. Bayer, A. Maier, M. Ostermeier, and R. Fahrig, "Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery," *International Journal of Biomedical Imaging*, vol. 2017, pp. 1–18, 06 2017.

[7] M. I. Miga, K. Sun, I. Chen, L. W. Clements, T. S. Pheiffer, A. L. Simpson, and R. C. Thompson, "Clinical evaluation of a model-updated image-guidance approach to brain shift compensation: experience in 16 cases," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 8, pp. 1467–1474, Aug 2016.

[8] D. Kuhnt, M. H. A. Bauer, and C. Nimsky, "Brain shift compensation and neurosurgical image fusion using intraoperative mri: Current status and future challenges," *Critical Reviews and trade in Biomedical Engineering*, vol. 40, no. 3, pp. 175–185, 2012.

[9] I. Reinertsen, F. Lindseth, C. Askeland, D. H. Iversen, and G. Unsgård, "Intra-operative correction of brain-shift," *Acta Neurochirurgica*, vol. 156, no. 7, pp. 1301–1310, Jul 2014.

[10] S. Ji, Z. Wu, A. Hartov, D. W. Roberts, and K. D. Paulsen, "Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery," *Medical Physics*, vol. 35, no. 10, pp. 4612–4624.

[11] H. Rivaz and D. L. Collins, "Deformable registration of preoperative mr, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 7, pp. 1017–1028, Jul 2015.

[12] A. Bilger, J. Dequidt, C. Duriez, and S. Cotin, "Biomechanical simulation of electrode migration for deep brain stimulation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 339–346.

[13] N. Hamzé, A. Bilger, C. Duriez, S. Cotin, and C. Essert, "Anticipation of brain shift in deep brain stimulation automatic planning," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 3635–3638.

[14] C. Essert, C. Haegelen, F. Lalys, A. Abadie, and P. Jannin, "Automatic computation of electrode trajectories for deep brain stimulation: A hybrid symbolic and numerical approach," *International journal of computer assisted radiology and surgery*, vol. 7, pp. 517–32, 08 2011.

[15] S. Bayer, N. Ravikumar, M. Strumia, X. Tong, Y. Gao, R. Fahrig, M. Ostermeier, and A. Maier, "Intraoperative brain shift compensation using a hybrid mixture model," 09 2018.

[16] I. Machado, M. Toews, J. Luo, P. Unadkat, W. Essayed, E. George, P. Teodoro, H. Carvalho, J. Martins, P. Golland, S. Pieper, S. Frisken, A. Golby, and W. Wells, "Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 10, pp. 1525–1538, Oct 2018.

[17] J. Luo, M. Toews, I. Machado, S. Frisken, M. Zhang, F. Preiswerk, A. Sedghi, H. Ding, S. Pieper, P. Golland, A. Golby, M. Sugiyama, and W. M. Wells III, "A feature-driven active framework for ultrasound-based brain shift compensation," in *MICCAI 2018*, 2018, pp. 30–38.

[18] F. Morin, H. Courtecuisse, I. Reinertsen, F. L. Lann, O. Palombi, Y. Payan, and M. Chabanas, "Brain-shift compensation using intraoperative ultrasound and constraint-based biomechanical simulation," *Medical Image Analysis*, vol. 40, pp. 133 – 153, 2017.

[19] F. M. M. Marreiros, S. Rossitti, C. Wang, and Ö. Smedby, "Non-rigid deformation pipeline for compensation of superficial brain shift," in *MICCAI 2013*. Springer Berlin Heidelberg, 2013, pp. 141–148.

[20] M. Luo, P. S. Larson, A. J. Martin, P. E. Konrad, and M. I. Miga, "An integrated multi-physics finite element modeling framework for deep brain stimulation: Preliminary study on impact of brain shift on neuronal pathways," in *Medical Image Computing and Computer Assisted Intervention – MICCAI*, 2019, pp. 682–690.

[21] J. Jiang, Y. Nakajima, Y. Sohma, T. Saito, T. Kin, H. Oyama, and N. Saito, "Marker-less tracking of brain surface deformations by non-rigid registration integrating surface and vessel/sulci features," *International journal of computer assisted radiology and surgery*, vol. 11, 03 2016.

[22] K. Sun, T. Pheiffer, A. Simpson, J. Weis, R. Thompson, and M. Miga, "Near real-time computer assisted surgery for brain shift correction using biomechanical models," *Translational Engineering in Health and Medicine, IEEE Journal of*, vol. 2, pp. 1–13, 04 2014.

[23] N. Haouchine, P. Juvekar, W. III, S. Cotin, A. Golby, and S. Frisken, "Deformation aware augmented reality for craniotomy using 3d/2d non-rigid registration of cortical vessels," *MICCAI 2020*, vol. 12264, pp. 735–744, 09 2020.

[24] A. Mohammadi, A. Ahmadian, A. D. Azar, A. D. Sheykh, F. Amiri, and J. Alirezaie, "Estimation of intraoperative brain shift by combination of stereovision and doppler ultrasound: phantom and animal model study," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 11, pp. 1753–1764, Nov 2015.

[25] S. Ji, X. Fan, D. W. Roberts, A. Hartov, and K. D. Paulsen, "Cortical surface shift estimation using stereovision and optical flow motion tracking via projection image registration," *Medical Image Analysis*, vol. 18, no. 7, pp. 1169 – 1183, 2014.

[26] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241.

[27] M. Nercessian, N. Haouchine, P. Juvekar, S. Frisken, and A. Golby, "Deep cortical vessel segmentation driven by data augmentation with neural image analogy," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 721–724.

[28] S. Cotin, C. Duriez, J. Lenoir, P. Neumann, and S. Dawson, "New approaches to catheter navigation for interventional radiology simulation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8, pp. 534–42, 02 2005.

[29] A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins, and D. Pizarro, "Shape-from-template," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2099–2118, Oct 2015.

[30] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes," in *Computer Vision – ACCV 2012*. Springer Berlin Heidelberg, 2013, pp. 548–562.

[31] A. Ebrahimi, "Mechanical properties of normal and diseased cerebrovascular system," *Journal of vascular and interventional neurology*, vol. 2, pp. 155–62, 04 2009.