Data-driven Intra-operative Estimation of Anatomical Attachments for Autonomous Tissue Dissection*

Eleonora Tagliabue[1], Diego Dall'Alba[1], Micha Pfeiffer[2], Marco Piccinelli[1], Riccardo Marin[1], Umberto Castellani[1], Stefanie Speidel[2] and Paolo Fiorini[1]

***Abstract*— The execution of surgical tasks by an Autonomous Robotic System (ARS) requires an up-to-date model of the current surgical environment, which has to be deduced from measurements collected during task execution. In this work, we propose to automate tissue dissection tasks by introducing a convolutional neural network, called BA-Net, to predict the location of attachment points between adjacent tissues. BA-Net identifies the attachment areas from a single partial view of the deformed surface, without any a-priori knowledge about their location. The proposed method guarantees a very fast prediction time, which makes it ideal for intra-operative applications. Experimental validation is carried out on both simulated and real world phantom data of soft tissue manipulation performed with the da Vinci Research Kit (dVRK). The obtained results demonstrate that BA-Net provides robust predictions at varying geometric configurations, material properties, distributions of attachment points and grasping point locations. The estimation of attachment points provided by BA-Net improves the simulation of the anatomical environment where the system is acting, leading to a median simulation error below 5mm in all the tested conditions. BA-Net can thus further support an ARS by providing a more robust test bench for the robotic actions intra-operatively, in particular when replanning is needed. The method and collected dataset are available at `https://gitlab.com/altairLab/banet`.**

***Index Terms*— AI-based methods; Surgical Robotics: Laparoscopy;**

I. INTRODUCTION

Surgical robotic systems have rapidly advanced in recent years, as confirmed by their wider adoption in the clinical field. The next frontier in surgical robotics is the introduction of increasing levels of autonomy [1]. Ideally, an autonomous surgical robot is provided with the sequence of actions to perform, which is decided based on patient's pre-operative information and a-priori clinical knowledge. However, such information is often not sufficient to thoroughly characterize the uncertain anatomical environment and define all the intervention steps, thus requiring the plan to be corrected while surgery is already taking place. In this work, we consider the process of automating soft tissue dissection, a very common surgical step which consists in separating two anatomical layers to access the region of interest. During this task, it is essential to carefully identify the resection points, to limit tissue damage which is inevitably introduced each time soft tissues are separated by the surgical instrument (e.g. monopolar scissors). For example, tissue dissection is performed in robotic partial nephrectomy to separate the perinephric fat tissue from the kidney in order to expose the tumor to excise. To accomplish the task, surgeons identify the attachment points between the adipose tissue and the kidney, where dissection has to take place, via manipulation of the tissue itself. In the same manner, an autonomous agent has to find candidate regions for dissection during task execution, since such attachments do not have a standard location and cannot be identified from pre-operative data. However, intra-operative identification of attachment regions involves some challenges related to the fact that such areas are often hidden from the partial view of the scene provided by intra-operative sensors.

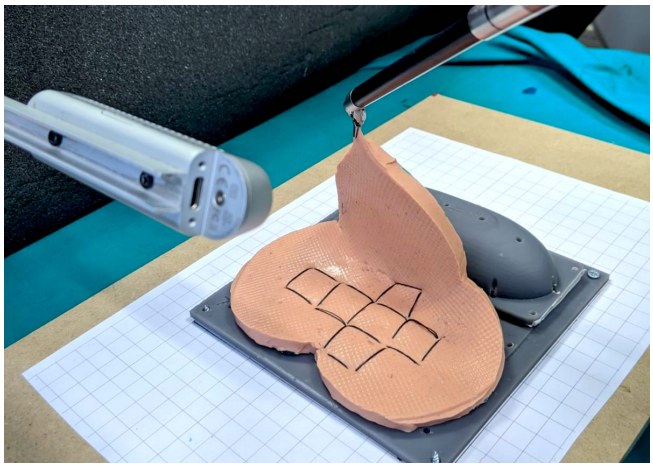

Fig. 1. A single dVRK PSM interacts with deformable phantoms stitched to the calibration base in correspondence of the attachment points. The PSM lifts the phantoms from their rest configurations. A snap-fit capsule is placed on the experiment board to induce a pre-deformation. The RGBD camera is used to acquire the point cloud of the deformed configurations.

An additional aspect to consider during the execution of a surgical task is that unexpected situations can occur because of the unpredictable behavior of the anatomical environment, thus requiring actions replanning. Before an autonomous agent executes any new action on the real system, it is strongly advisable to test the updated motion in simulation to guarantee that it can be safely performed. As a consequence, accurate simulation of the environment where the system is acting becomes of paramount importance. Precise knowledge of the location of attachment points would

[1]Authors are with Department of Computer Science, University of Verona, Verona, Italy `eleonora.tagliabue@univr.it`

[2]Authors are with National Center for Tumor Diseases (NCT), Dresden, Germany

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742671 "ARS").

DOI 10.1109/LRA.2021.3060655

not only help to identify the dissection region, but it would also improve the simulation of the anatomical environment. In fact, attachment points play the role of Dirichlet boundary conditions for soft tissue simulations, since they act as constraints to the motion of specific points (i.e. attachments act as fixed points). Previous works have demonstrated that correct definition of boundary conditions can highly impact simulation accuracy [2], especially in tasks where the driving input is a displacement, as is the case in many surgical systems, which lack of contact force-torque measurement.

In this work, we propose to use a convolutional neural network (CNN) to predict the location of attachment points. Our method identifies attachment regions directly on the pre-operative 3D anatomical model, purely based on positional information. The main contributions are the following:

1) we introduce a method to predict the attachment areas from a single partial view of the intra-operative surface, without any a-priori knowledge of their location;
2) we show that the proposed CNN guarantees inference times compatible with intra-operative applications;
3) we demonstrate the ability of the method to generalize to varying geometric configurations, material properties, distributions of attachment points and grasping point locations on both simulation and real world phantom experiments performed with the da Vinci Research Kit (dVRK) (Fig. 1) [3].

This is the first application of a deep network to estimate the attachment regions of a deformable organ intra-operatively, and would support an autonomous system by both refining the robot plan and updating the simulated model online.

II. Related works

The problem of automating soft tissue dissection has been addressed by some recent works. For example, [4] and [5] propose some methods to generate robot motion primitives exploiting information from stereo-camera images. However, these works do not entail autonomous identification of the dissection start point, assuming it to be manually specified by the surgeon. One of the first attempts to automate path planning during soft tissue retraction, the first step of dissection, is represented by [6]. Authors propose different optimization strategies based on the tissue state and/or the robot effort to generate the sequence of control actions. However, this approach relies on accurate modeling of the tissue mechanical behavior, which is not easy to obtain due to the uncertain and highly variable anatomical environment. Attanasio et al. [7] present a framework for both autonomous path planning and execution of tissue retraction. In their workflow, grasping points are identified based on the geometry and position of the tissue to grasp, which is segmented from endoscopic images using a deep network. Similarly to this work, we also rely on a deep architecture to extract information about the regions to grasp. However, the aim of our method is to directly identify attachment regions purely based on positional information from the pre-operative model and observed surface displacement, without relying on the video stream. The chosen approach further allows us to refer the attachment points to the original 3D anatomical model, thus enabling a more accurate definition of the simulation boundary conditions and, in turns, an improvement in simulation accuracy.

Some approaches have been also proposed to estimate boundary conditions for soft tissue simulations in the surgical field. In [8], attachments are estimated by matching two anatomical configurations extracted from CT scans acquired before and after patient repositioning. This approach assumes that the entire surface is available in both configurations, preventing from its application in real clinical settings, when only a partial surface view can be obtained intra-operatively. Plantefeve et al. [9] propose to initialize the position of attachment points based on statistical atlases. However, statistical atlases are not always able to adapt to patient-specific conditions due to the high inter-patient variability. Another line of research exploits Kalman Filters (KFs) to estimate boundary conditions in the context of liver surgery [10], [11]. The main constraints between the liver and the surrounding tissues is represented by the hepatic ligaments, whose location can be extracted from pre-operative data. As a consequence, these works have focused on the characterization of the ligaments elastic properties, based on the assumption that their location is known, which does not hold in our application. Methods based on KFs have the advantage to work with partial observations (i.e. partial visible surface) coming from intra-operative sensors. However, the convergence time of KFs is highly dependent on how close the initialization is to the optimal solution, thus not ensuring estimation time suitable for online model update. In this work, we present a method that can predict simulation boundary conditions from a partial view of the anatomical deformed state, without any prior knowledge about their location. By relying on a deep network, our approach further guarantees inference times compatible with intra-operative model update.

III. METHOD

We present BA-Net (Binary Attachments Network), a CNN which outputs a binary map of estimated attachment points starting from a pre-operative 3D model and the displacement field of the intra-operatively visible portion of surface. In order to exploit convolutional operators, our framework relies on a representation of the data on a regular grid of dimension 64^3 and $300\,mm$ side length. Despite relying on the same formalism proposed in [12], [13] for its capability to generalize to different geometric shapes, our method has a completely different goal. In fact, BA-Net does not aim at predicting a 3D displacement field that registers two anatomical configurations, but it predicts which points of the pre-operative 3D model act as attachment points. The proposed network is trained only with simulated samples, to cope with the lack of real world data where the attachment points are annotated on the corresponding pre-operative volumes.

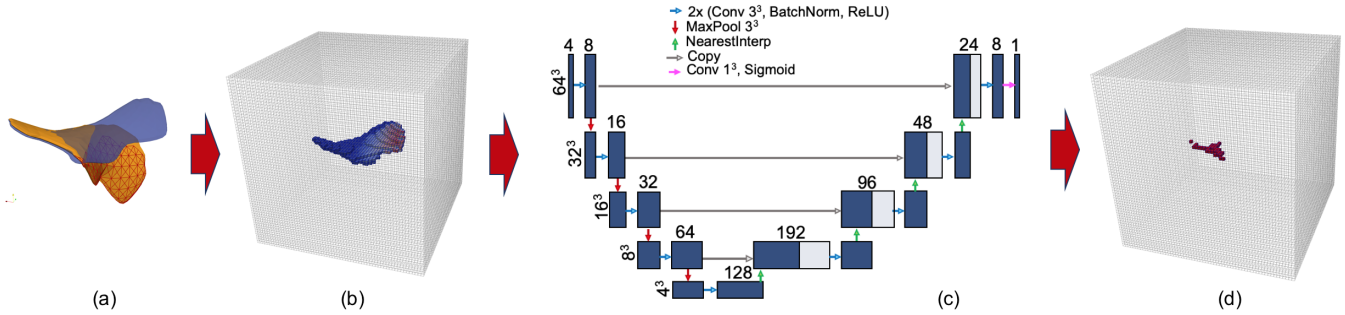

Fig. 2. Overview of our method. (a) Training data are generated from finite element simulations. The blue mesh represents one of the randomly generated surfaces, while the orange mesh represents the corresponding deformed configuration. A wireframe overlay is added in correspondence of the visible portion of the deformed mesh. (b) Input to the network is a structured grid where the initial undeformed surface is encoded through its signed distance field (sdf). Only grid cells belonging to the internal parts of the surface (sdf<voxel size) are displayed. Cell color is proportional to the magnitude of the associated displacement (higher displacement in red, zero displacement in blue). (c) The UNet architecture used. (d) Output of the network is a binary map of the attachment points. Such points are then converted from grid coordinates to original mesh coordinates.

A. Training Data Generation

A set of random surface meshes are generated by applying a series of morphological operations to an icosphere of random dimension. In order to mimic the pre-operative configuration of fat tissues as realistically as possible, the generated samples are clipped within two parallel planes to keep sample thickness below $20\,mm$, and twist and bend filters are applied to pre-deform the meshes. For this process, we make sure that the average edge length of the generated geometries is comparable to the grid voxel size (i.e., approximately $4.7\,mm$), to match the mesh and grid resolutions. After tetrahedrization of the resulting mesh, we extract a subset of the surface whose points will act as attachments. To define such subset with a realistic (thus irregular) profile, we associate to each point Q of the surface the value of the metric D_Q which acts as the likelyhood of that point being removed from the subset. After selecting a random mesh node P as center point, we define the distance metric D associated to each point Q as:

$$D_Q = w_d \cdot d_{PQ} + w_n \cdot n_{PQ} + w_p \cdot p_Q \qquad (1)$$

where d_{PQ} is the geodesic distance between P and Q, n_{PQ} is the angular distance between the normal of P and the normal of Q, and p is the value of perlin noise evaluated at position Q. These three contributions are weighted by w_d, w_n and w_p, whose values change for each new sample within specific bounds that can be adjusted to specify the relative importance of each term. For the extraction of attachment points, we sample w_n within a range of higher values with respect to the other two, to favor the extraction of regions belonging to the same side of the surface. Eventually, the extracted surface includes those points with the lowest values of D, until the desired percentage of surface points (between $(5, 50)\%$) is extracted.

We introduce a simulation environment to obtain tissues deformed state relying on the finite element (FE) method provided by SOFA framework [14]. Grasping action performed with a single dVRK arm is simulated by applying a force of random magnitude to a subset of surface nodes within a radius of $(4, 10)\,mm$, to simulate different amounts of grasped tissue. In order to make the network independent from specific mechanical properties, we consider varying Young's modulus $(3, 30)\,kPa$ and Poisson's ratio $(0.4, 0.45)$ such that they cover the range of values describing adipose tissues [15]. StVenant-Kirchhoff material is chosen to model tissue mechanics, since it represents the simplest generalization of elastic material to the non-linear regime while ensuring a good trade-off between simulation time and accuracy. A partial surface is then extracted from the deformed configuration relying the same method used for the extraction of attachment points, in a range between $(10, 100)\%$ of the entire surface, to simulate the partial view which is acquired by vision sensors intra-operatively. For each point of the undeformed mesh which belongs to the visible surface, we compute the displacement field which brings it to its deformed counterpart, while we associate all the remaining points with zero displacement.

The computed displacement field, together with the undeformed surface mesh, represent the input to our network. This input is converted into the grid-like structure required by the method following the same voxelization process described in [12], which encodes the undeformed mesh through its signed distance field and uses a Gaussian kernel to interpolate the displacement onto the grid points. By representing the data in this way, our method learns to interpret the geometry, allowing it to directly generalize to new geometries at inference time. The ground truth binary mask of attachments is defined by assigning a value of 1 to all grid cells which contain a fixed mesh node. An overview of the data conversion pipeline is provided in Fig. 2. Following this process, we generate a dataset composed of of 5000 samples, which are split into training and validation sets $(90 - 10\%)$.

B. BA-Net Architecture

The output of our method is a 3D binary map defined in the same domain as the input grid, where unitary values are assigned to each grid voxel containing attachment points. Our framework relies on the UNet architecture, which has been already successfully applied to learning deformation tasks

on 3D grids [12], [13], [16]. These works have confirmed UNet capability of learning both high-level representations of the data in the bottleneck layers and carrying high-level information using the skip-connections.

BA-Net architecture is illustrated in Fig. 2c. The network first contracts the input data into a 4^3 volume, to make sure that information coming from displacement fields taking place on one side of the input surface can influence the opposite side of the surface. The encoded information is then expanded back to original 64^3 space and converted into a 3D binary map via a final 1^3 convolution layer. Similarly to [13], interpolation operators are employed in the decoding path, to save computational time required by the standard up-convolutions. In addition, dropout layers at 50% are added after each max pool and interpolation operations, which allow to improve generalization capabilities of the network. The loss function L used for training is a linear combination of the Dice similarity coefficient (DSC) [17] and the binary cross entropy (BCE):

$$L = \frac{1}{N}\sum_{1}^{N}(1 - DSC + BCE) \tag{2}$$

where N is the batch size, which is 32 in our case. The network is trained with AdamW optimizer [18] and one cycle learning rate scheduler [19], on a workstation with Intel Xeon CPU and NVIDIA GeForce 2080 Ti GPU.

IV. EXPERIMENTS

Performances of BA-Net have been assessed on both simulation and real world phantom data. We consider two metrics to evaluate the overall accuracy of attachment points prediction: the DSC coefficient and the true positive rate (TPR) [17]. The DSC coefficient is a measure of the overlap between the predicted and the ground truth areas. High values of the DSC coefficient are obtained when intersection between prediction and ground truth is maximized, and union is minimized. Despite being highly correlated with prediction accuracy, the DSC is strongly impacted by errors in the delineation of the contours of the region of interest. However, in our application it is more important to ensure that the predicted region includes all the true attachments even at the cost of introducing some errors in boundaries delimitation. Therefore, we also introduce TPR (or sensitivity) as evaluation metric, which measures the proportion of ground truth attachments that are also identified by the prediction, thus not influenced by errors at the boundaries.

Furthermore, we evaluate the simulation error introduced when relying on the predicted attachments. We compare the deformed configuration obtained with ground truth attachments as boundary conditions with the one obtained when fixing the regions predicted by BA-Net. Simulation error is computed as average volume error (AVE), where the volume error for one sample is defined as the mean square error (MSE) between the points of the deformed volume mesh with predicted attachments and the corresponding ones of the ground truth. In order to better represent the real surgical scenario, where there is no information about the tissue-robot interaction force, simulations based on predicted attachments are performed by considering the displacement of the surface nodes grasped by the simulated end-effector as driving input. This modelling choice allows also to obtain a deformation profile which is independent of possible inaccuracies in chosen material parameters [2]. Therefore, the defined AVE does not only contain the errors made by approximating the boundary conditions with BA-Net predictions, but also the error made by replacing the force input with a displacement input and the error made by discrepancies between simulation and reality. In order to isolate the error contribution due to imprecise estimation of attachment points, we provide a baseline value for the AVE by running forward simulations using ground truth attachment points as boundaries. We refer to this configuration with the word *Same* and to its corresponding AVE with AVE *Same*.

A. Simulated Data

A test dataset composed of 380 simulated samples (*Test*) is generated following the same pipeline described in Section III-A. The random simulations result in samples which have a median input displacement of $38.8\,mm$. In addition to reporting the metrics values obtained when ground truth attachments are used as boundary conditions (i.e. the *Same* configuration described above), we also detail the metrics when naive initializations of the boundaries are used. In particular, we fix (i) no points (*Zero*), (ii) all the points belonging to the lowest surface (*All*). Comparing the metrics on the test set with those obtained on these representative configurations makes it possible to assess how BA-Net predictions impact the simulation error. We also isolate values relative to the subset of test dataset whose samples have an associated visible surface below 50%, a condition which is closer to real cases (*TestV*).

B. Real World Phantom Data

BA-Net performances are tested on real world soft tissue manipulations performed with a single dVRK Patient Side Manipulator (PSM) arm. We fabricate four deformable phantoms using commercially available addition curing silicone rubber (Smooth-On Ecoflex materials) with different geometric shapes (circle, clover, rectangle and drop), thickness $(6, 6, 5, 12\,mm)$ and elastic properties (obtained by using silicone rubber with different shore hardness). For each phantom, we choose 3 configurations of fixed points, which are schematically summarized in Fig. 3. The position of fixed points is defined by stitching the phantoms on a 3D printed calibration base (dimensions $144 \times 144 \times 4\,mm$), called Reconfigurable Attachment Board (RAB), with regularly spaced holes in a 7×7 grid (distance between adjacent holes is $18\,mm$). Thanks to the RAB, each set of attachment points can be mapped to the corresponding simulated mesh to generate the virtual ground truth. For each attachments configuration, we select 3 or 4 grasping points distributed over the unconstrained portion of the phantom (see Fig. 3). Furthermore, for each configuration we consider two starting

conditions: the former where the phantom lies flat on the RAB, the latter where a snap-fit capsule structure is added to the RAB to introduce an initial pre-deformed state to the phantoms (Fig. 1), similarly to the real scenario where fat tissues lie on the kidney. In our experiment, we lift the tissue from each grasping point to the maximum feasible extent and we record the point cloud representing the current state of the surface at regular steps of $10\,mm$, while increasing the lifting. The point cloud is acquired by an Intel RealSense D435 RGB-D camera and is automatically registered to the virtual geometry thanks the initial system calibration, which is performed following the same process described in [20]. The displacement field relative to the visible phantom surface is retrieved by computing a dense point-to-point matching problem between the acquired point clouds and the undeformed phantom surface. We solve non-rigid correspondence in the functional space of the models using ZoomOut, a state-of-the-art refinement technique [21]. This method is particularly robust in presence of near-isometric deformations, which is our case. Eventually, voxelization is performed to convert the data into the input format required by the network.

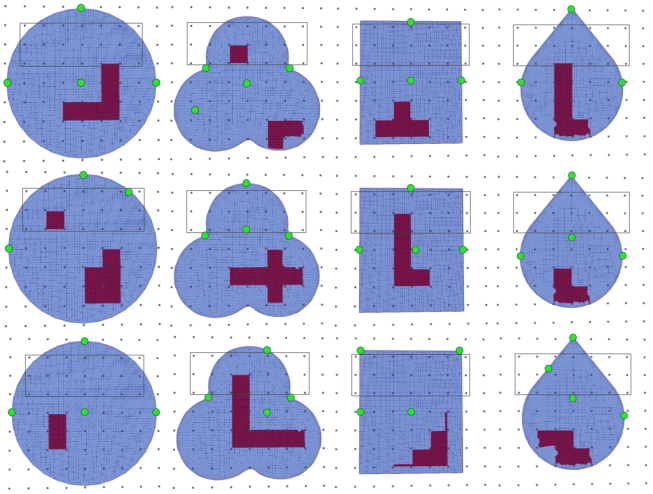

Fig. 3. Configurations of attachments considered in the real world phantom experiments. Red areas represent the defined attachment regions; green spots correspond to grasping points. Configurations are overlaid to the grid of RAB points, which allows to uniquely map the positions between the real and simulated environment. Gray rectangle defines the position of the snap-fit capsule that allows to obtain pre-deformed configurations.

V. RESULTS

A. Simulated Data

Values of evaluation metrics on test dataset (*Test*) are detailed in Table I. The maximum value of volume error (MVE) obtained in simulations with predicted attachments has a median of $7.0(4.0-12.7)\,mm$. Median MVE for the *Same* configuration is $5.2(2.24-10.4)\,mm$. In most of the samples, the maximum volume error is found close to the grasping point, which usually corresponds to the maximum deformation. The median time required by BA-Net to predict the attachment point position is $43.9\,ms$ (including time for data upload to GPU).

TABLE I

EVALUATION METRICS, AS MEDIAN (25TH-75TH PERCENTILE), FOR THE SIMULATION TEST SAMPLES.

	DSC	TPR [%]	AVE [mm]
Zero	0.00 (0.00-0.00)	0 (0-0)	27.0 (13.4-52.0)
All	0.42 (0.31-0.55)	93 (84-100)	6.7 (2.5-12.4)
Same	1.0 (1.0-1.0)	100 (100-100)	0.6 (0.2-1.7)
Test	**0.82 (0.70-0.88)**	**86 (75-91)**	**1.1 (0.5-2.1)**
TestV	0.76 (0.63-0.83)	83 (69-90)	1.2 (0.6-2.4)

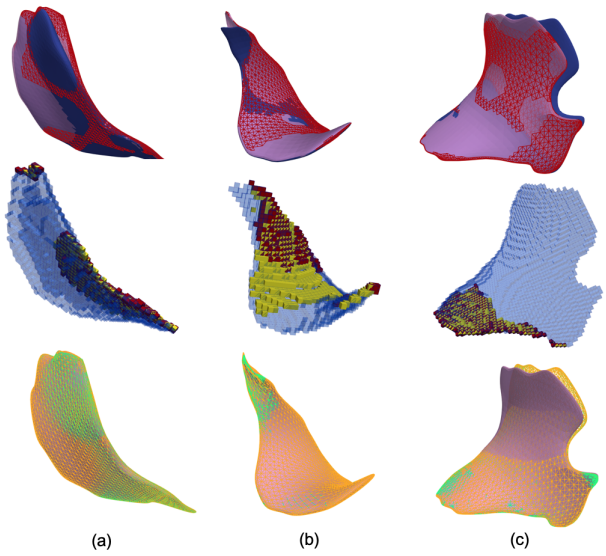

Fig. 4. BA-Net predictions for three simulated cases (one per column). First row: undeformed surface (blue), deformed surface (pink) and visible surface (red wireframe). Second row: voxelized initial surface (sdf $<$ grid voxel size) (blue), ground truth attachments (red) and predicted attachments (yellow). Last row: deformed surface (pink), deformed surface when using predicted attachments (orange), and deformed surface when using ground truth attachments (green). (a) Sample composed of two disjoint regions of attachments, correctly predicted by the network. (b) Sample associated to a low visible surface (36%). In this case, BA-Net overestimates the attached region. (c) Sample characterized by a good prediction accuracy but non-zero AVE. However, green and orange meshes are perfectly superimposed, thus high AVE is due to the different simulation method and not inaccurate prediction.

B. Real World Phantom Data

Table II summarizes the results obtained for the experiments conducted on the real scenario, grouped following different criteria. First, we consider the results relative to all the acquisitions associated to the same phantom (first four rows), to understand if the performances are influenced by the different geometries and properties. Secondly, we analyze metrics values at different levels of input displacement ($2, 4, 6\,cm$ from the rest configuration), to evaluate the influence of the lifting height on the predictions. Results are further grouped depending on the starting configuration (flat or pre-deformed). The AVE reported in Table II represents the error between each point in the acquired point cloud and

TABLE II

EVALUATION METRICS ON THE REAL EXPERIMENTS, EXPRESSED AS MEDIAN (25TH-75 PERCENTILE). LAST COLUMN REPORTS THE NUMBER OF ACQUISITIONS WHICH CONTRIBUTED TO THE STATISTICS OF THE CORRESPONDING ROW.

	DSC	**TPR** [%]	**AVE** [mm]	**AVE** ***Same*** [mm]	**#samples**
Circle	0.28 (0.16-0.40)	68.7 (33.3-93.3)	3.2 (2.1-5.0)	3.5 (2.4-4.8)	101
Clover	0.40 (0.26-0.46)	62.8 (48.8-79.0)	3.1 (1.8-5.8)	3.0 (2.0-4.7)	79
Rectangle	0.43 (0.37-0.51)	86.7 (66.5-100.0)	3.7 (2.1-7.3)	4.2 (2.5-6.2)	85
Drop	0.54 (0.42-0.66)	75.6 (52.6-89.7)	3.4 (2.0-5.0)	3.6 (2.4-5.0)	80
Lift 2cm	0.41 (0.31-0.48)	72.6 (49.4-93.8)	3.9 (2.2-6.1)	3.8 (2.6-5.4)	85
Lift 4cm	0.44 (0.29-0.53)	70.6 (54.7-86.7)	5.0 (2.1-7.4)	4.8 (3.1-5.7)	44
Lift 6cm	0.44 (0.37-0.56)	76.3 (55.4-89.9)	3.9 (1.7-7.9)	4.5 (1.9-7.2)	15
Flat	0.41 (0.30-0.50)	82.1 (60.1-97.8)	2.8 (1.8-4.8)	3.1 (2.3-4.8)	198
Pre-deformed	0.40 (0.27-0.49)	65.1 (34.7-84.3)	4.1 (2.4-7.2)	3.8 (2.6-5.7)	147
Grasp	0.50 (0.23-0.56)	51.5 (35.4-60.0)	7.4 (4.5-9.8)	–	24

the corresponding one in the deformed configuration, relative to the AVE at rest (median value $4.8\,mm$). Error at rest includes the contributions of registration error, inaccuracies in the computed correspondences and sensor noise. Table II also details the values of AVE *Same*, which allows to assess the magnitude of the error we are making even when using the optimal boundaries, thus solely caused by discrepancy between simulation and reality. Visual examples of the network predictions are provided in Fig. 5. Finally, we want to assess the robustness of the predictions at different grasping points, for the same configuration of attachments. To this end, we compute metric values considering the intersection of the predictions at different grasping points, for the same lifting height (in our case $4\,cm$), i.e. fixing only those points which are predicted as fixed from all the grasping points (Table II, last row). Fig. 6 provides some visual examples of these predictions. It is worth noting that real world experiments are associated to a visible surface below 50%, which is the most challenging condition.

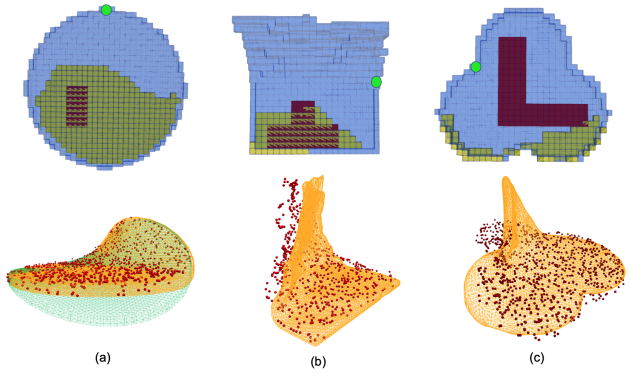

Fig. 5. BA-Net predictions for three real cases (one per column). Upper row: voxelized initial surface (sdf $<$ grid voxel size) (blue), ground truth attachments (red) and predicted attachments (yellow). The considered grasping point is indicated. Lower row: acquired point cloud (red) and deformed surface when using predicted attachments (orange). (a) Sample at a lifting level of $2\,cm$. The green mesh overlaid on the bottom configuration represents the deformed surface when using ground truth attachments. (b) Sample at a lifting level of $4\,cm$, starting from an initially deformed configuration. (c) Sample at a lifting level of $3\,cm$. In this configuration, the PSM occludes the upper part of acquired point cloud.

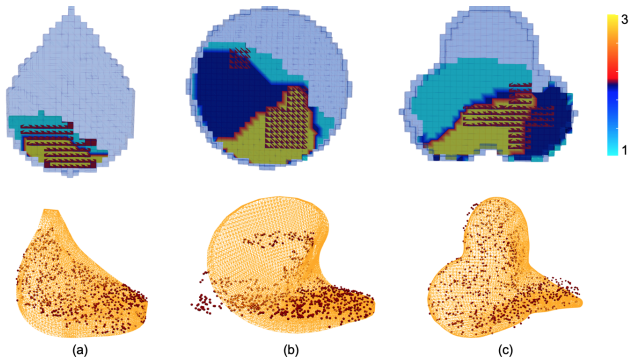

Fig. 6. Network performance when considering the intersection of the predictions at different grasping points, for the same configuration of attachments, in three different real cases. Upper row: voxelized initial surface (sdf $<$ grid voxel size) (blue) and ground truth attachments (red). Predicted attachments are rendered according to a colormap which maps a region with yellow if all the predictions considered it as attached. Lower row: acquired point cloud (red) and the deformed surface when fixing points predicted as attached by all configurations (orange).

VI. DISCUSSION

A. Simulated Data

BA-Net predictions maximize both DSC and TPR, showing that the network has learnt to generalize to new unseen geometries and configurations of fixed points (Fig. 4). The median AVE achieved by simulations with predicted attachments is $1.1\,mm$. Since this value is obtained as an average error over the entire considered geometries, which includes both internal and surface points, it represents an overall precise matching between the ground truth and the deformed state obtained with predicted attachments. This precise matching is further supported by considering the large applied input displacement, with a median value of $38\,mm$. Although the median MVE might seem large, its value differs from the one obtained when running forward simulations with ground truth attachments (*Same* configuration) by less than $2\,mm$. The fact that MVE is not zero with the *Same* configuration indicates that there is a baseline error introduced by applying an input displacement instead of a force (Fig. 4c). Overall, simulation accuracy has

significantly improved with respect to the cases where a naive initialization is given to the boundary conditions, i.e. when fixing either zero or all the points. Prediction accuracy is impacted by more challenging conditions, i.e. limited input information, as confirmed by metrics values when the visible surface is below 50% (*TestV*), that are slightly worse than the ones on the entire dataset.

B. Real World Phantom Data

Table II shows that values of the evaluation metrics are aligned for all the different experimental conditions. BA-Net is able to handle different geometries and material properties, with only slight differences. Worst results in terms of prediction accuracy are obtained for circle and clover. The reason for this is twofold. Firstly, they are the only samples for which we tested a configuration of attachments composed of two disjoint areas. Although samples with more than one attachment region are present in the training dataset (Fig. 4a), the network always predicts a single fixed region on the real data (Fig. 6b). The second reason for these suboptimal values is that the attachment region is often overestimated, as emerges by visual inspection of the results (Fig. 5a). However, this is not indicating bad prediction performance: BA-Net learns to model not only fixation points but also constraints imposed by the environment. In fact, the network predicts the whole area where the phantom is in contact with the RAB as attached, which constrains phantom motion but is not taken into account by the ground truth. This allows to achieve an overall good matching between simulation with predicted attachments and real deformed state, which actually outperforms the simulation result obtained when ground truth points are fixed (Fig. 5a). Although the best trade-off between the metrics assessing prediction accuracy is obtained for rectangle and drop, these shapes are associated to a higher AVE with respect to circle and clover. This is probably due to the fact that the considered constitutive law is not able to accurately describe the behavior of these two phantoms, fabricated with a different silicone rubber, which showed some time-dependent behavior (Fig. 5b). We expect that this error could be reduced by using a more accurate biomechanical model or injecting real samples in the dataset, but such fine tuning was out of the scope of this work.

Fig. 5c shows a failure case for BA-Net, which corresponds to the configuration associated with the leftmost grasping point. The heavy occlusion introduced by the PSM causes dramatic geometrical noise that perturbs the matching retrieved by [21], in particular due to ruined surface estimation. Even though BA-Net prediction is poor in this case, this is caused by limited surface visibility and not by failure of the method itself. The inaccurate matching could be resolved in the future by either injecting prior knowledge to the method (e.g. trusted landmarks or temporal constraints) or by introducing the second dVRK PSM, to limit occlusions.

BA-Net predictions are not influenced by the magnitude of the input displacement, as confirmed by the fact that higher input deformation introduces a limited gain in DSC and a slight reduction in TPR variability. The increase in the AVE values is due to the fact that bigger input displacements are more likely to introduce some instabilities in the simulations. We note that there is a slight difference in the performances depending on the starting state: even if the DSC is comparable between the two conditions, experiments starting from initially flat configurations are associated to higher percentage of correctly identified attachments. The difference in the AVE can be partially due to suboptimal 3D mesh when the configuration is initially deformed. While the geometry could be directly extracted from pre-operative data in real scenarios, for these experiments we had to warp the flat configuration based on the rest point cloud in the deformed state, which introduced some low quality regions that can have a negative impact on simulation accuracy.

BA-Net predictions are coherent when varying the grasping point (Fig. 6). However, when single states of deformation are viewed, the region of attachment points is often overestimated, likely because the network is unsure whether a region which does not move is fixed or not. Last row of Table II shows that the DSC tends to increase when using the intersection of the predictions from multiple views. This indicates a better matching of the actual attachment region, even though it sometimes comes at the expenses of some missed regions (lower TPR). A reason for this behavior is provided in Fig. 6b and c, which shows that if just one of the predictions misses a region, it will be excluded from the intersection (leading to worse AVE as well). In future works, we plan to test other strategies for combining the predictions obtained from different grasping points, for example weighting the different contributions based on the likelyhood of the predictions.

In general, median values for the DSC coefficient seem quite low if compared to the ones obtained on simulated samples. However, metrics values must be considered with respect to the application. As highlighted in Section I, the network could be exploited to support an autonomous surgical system in two ways: either by providing a guess of the attachment region to move towards or by improving simulation accuracy in case of replanning.

If the goal is to identify the location of attachment points, it is preferrable that most of the attached area is correctly identified, even at the cost of having added regions or inaccuracies in boundary delimitation. In this context, having high TPR values is more important than having a good DSC, because we want to minimize the amount of missed regions. Achieving a median TPR above 62% for all the real world phantom experiments tells us that most of the attached area is correctly identified. This is an interesting result if we consider that such accuracy is achieved by providing a single partial view of the deformed state of the tissue as input and without relying on any prior information and considering that the net is trained with only simulated samples. Even expert surgeons would find it challenging to precisely identify the attached area from a single manipulation, especially if the applied displacement is small. What experts would generally do is to move towards the expected area of attachments and perform further manipulations of the tissue, until they

are confident enough about the location of the attached points. We expect that BA-Net would benefit from a similar approach, and future works will focus on improving the prediction by providing sequential frames as input.

The other possible application of BA-Net in surgery deals with improving simulation accuracy. Simulations performed with the boundary conditions predicted by BA-Net lead to an AVE which is better than the one obtained when fixing ground truth points in most of the cases. This is due to the fact that simulations fixing ground truth points do not take into account the constraint provided by the RAB, while the network seems to learn to account for that. Simulations with ground truth attachments can be thought of as a real surgical scenario where we have some a priori knowledge about the attachment area, but no guess about the constraints provided by the surrounding anatomical environment. The reduction in AVE introduced by BA-Net tells us that the method has the potential to improve simulation accuracy with respect to the case when some a priori knowledge about the area of interest is available. Although obtained AVEs might seem large in absolute terms, we have to consider that it is obtained with sub-optimal simulations, characterized by quite coarse anatomical models and rough modelling assumptions. Using a higher resolution geometry and object-specific constitutive law would help to reduce such error. In the current work, our main focus was the assessment of the general prediction capabilities of the method and its ability to update biomechanical simulations, thus we relied on models that could guarantee a good trade-off between accuracy and computational performance.

VII. CONCLUSION

In this work, we have presented BA-Net, a framework for the intra-operative identification of attachment regions to automate robotic tissue dissection tasks. BA-Net has proven able to accurately estimate the location of the attachment points from a single partial view of the deformed surface with a very low inference time, thus making the method suitable for intra-operative model update. Despite being trained on a simulated dataset only, BA-Net generalizes to real world phantom data with variable properties and configurations. Future works will focus on improving the prediction performances by further reducing the gap between the synthetic training dataset and real world acquisitions and by showing the network multiple frames to account for time dynamics. Network performances will also be tested on adipose tissue manipulation in real anatomical environments.

REFERENCES

[1] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, *et al.*, "Medical roboticsregulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, p. 8638, 2017.

[2] K. Miller and J. Lu, "On the prospect of patient-specific biomechanics without patient-specific properties of tissues," *Journal of the mechanical behavior of biomedical materials*, vol. 27, pp. 154–166, 2013.

[3] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da vinci® surgical system," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 6434–6439.

[4] R. Elek, T. D. Nagy, D. Á. Nagy, T. Garamvölgyi, B. Takács, P. Galambos, J. K. Tar, I. J. Rudas, and T. Haidegger, "Towards surgical subtask automationblunt dissection," in *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*. IEEE, 2017, pp. 000 253–000 258.

[5] D. Á. Nagy, T. D. Nagy, R. Elek, I. J. Rudas, and T. Haidegger, "Ontology-based surgical subtask automation, automating blunt dissection," *Journal of Medical Robotics Research*, vol. 3, no. 03n04, p. 1841005, 2018.

[6] S. Patil and R. Alterovitz, "Toward automated tissue retraction in robot-assisted surgery," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 2088–2094.

[7] A. Attanasio, B. Scaglioni, M. Leonetti, A. F. Frangi, W. Cross, C. S. Biyani, and P. Valdastri, "Autonomous tissue retraction in robotic assisted minimally invasive surgery–a feasibility study," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6528–6535, 2020.

[8] I. Peterlik, H. Courtecuisse, C. Duriez, and S. Cotin, "Model-based identification of anatomical boundary conditions in living tissues," in *International Conference on Information Processing in Computer-Assisted Interventions*. Springer, 2014, pp. 196–205.

[9] R. Plantefève, I. Peterlik, N. Haouchine, and S. Cotin, "Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery," *Annals of biomedical engineering*, vol. 44, no. 1, pp. 139–153, 2016.

[10] I. Peterlik, N. Haouchine, L. Ručka, and S. Cotin, "Image-driven stochastic identification of boundary conditions for predictive simulation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 548–556.

[11] S. Nikolaev and S. Cotin, "Estimation of boundary conditions for patient-specific liver simulation during augmented surgery," *International Journal of Computer Assisted Radiology and Surgery*, 2020.

[12] M. Pfeiffer, C. Riediger, J. Weitz, and S. Speidel, "Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2019.

[13] M. Pfeiffer, C. Riediger, S. Leger, J.-P. Kühn, D. Seppelt, R.-T. Hoffmann, J. Weitz, and S. Speidel, "Non-rigid volume to surface registration using a data-driven biomechanical model," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Springer International Publishing, 2020.

[14] F. Faure, C. Duriez, H. Delingette, J. Allard, B. Gilles, S. Marchesseau, H. Talbot, H. Courtecuisse, G. Bousquet, I. Peterlik, *et al.*, "Sofa: A multi-model framework for interactive physical simulation," in *Soft tissue biomechanical modeling for computer assisted surgery*. Springer, 2012, pp. 283–321.

[15] N. Alkhouli, J. Mansfield, E. Green, J. Bell, B. Knight, N. Liversedge, J. C. Tham, R. Welbourn, A. C. Shore, K. Kos, *et al.*, "The mechanical properties of human adipose tissues and their relationships to the structure and composition of the extracellular matrix," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 305, no. 12, pp. E1427–E1435, 2013.

[16] A. Mendizabal, E. Tagliabue, T. Hoellinger, J.-N. Brunet, S. Nikolaev, and S. Cotin, "Data-driven simulation for augmented surgery," in *Developments and Novel Approaches in Biomechanics and Metamaterials*, July 2020, vol. 132, pp. 71–96.

[17] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, p. 29, 2015.

[18] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7

[19] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 464–472.

[20] E. Tagliabue, A. Pore, D. Dall'Alba, E. Magnabosco, M. Piccinelli, and P. Fiorini, "Soft tissue simulation environment to learn manipulation tasks in autonomous robotic surgery," in *2020 IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

[21] S. Melzi, J. Ren, E. Rodolà, A. Sharma, P. Wonka, and M. Ovsjanikov, "Zoomout: spectral upsampling for efficient shape correspondence," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, p. 155, 2019.