



HAL
open science

Physics-based Deep Neural Network for Augmented Reality during Liver Surgery

Jean-Nicolas Brunet, Andrea Mendizabal, Antoine Petit, Nicolas Golse, Stéphane Cotin

► **To cite this version:**

Jean-Nicolas Brunet, Andrea Mendizabal, Antoine Petit, Nicolas Golse, Stéphane Cotin. Physics-based Deep Neural Network for Augmented Reality during Liver Surgery. Medical image computing and computer-assisted intervention: MICCAI .. International Conference on Medical Image Computing and Computer-Assisted Intervention, In press, 1-8. hal-02158862v1

HAL Id: hal-02158862

<https://inria.hal.science/hal-02158862v1>

Submitted on 18 Jun 2019 (v1), last revised 19 Jun 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Physics-based Deep Neural Network for Augmented Reality during Liver Surgery

Jean-Nicolas Brunet^{1,*}, Andrea Mendizabal^{1,*}, Antoine Petit¹, Nicolas Golse²,
and Stéphane Cotin¹

¹ INRIA. Strasbourg, France
`stephane.cotin@inria.fr`

² Hopital Paul-Brousse. Paris, France

* J-N. Brunet and A. Mendizabal contributed equally to the paper.

Abstract. In this paper we present an approach combining a finite element method and a deep neural network to learn complex elastic deformations with the objective of providing augmented reality during hepatic surgery. Derived from the U-Net architecture, our network is built entirely from physically-based simulations of a preoperative segmentation of the organ. These simulations are performed using an immersed-boundary method, which offers several numerical and practical benefits, such as not requiring boundary-conforming volume elements. We perform a quantitative assessment of the method using synthetic and *ex vivo* patient data. Results show that the network is capable of solving the deformed state of the organ using only a sparse partial surface displacement data and achieve similar accuracy as a FEM solution, while being about 100x faster. When applied to an *ex vivo* liver example, we achieve the registration in only 3 *ms* with a mean target registration error (TRE) of 2.9 *mm*.

Keywords: Deep learning · real-time simulation · augmented reality

1 Introduction

In computer-assisted interventions, correctly aligning preoperative data to real-time acquired intraoperative images remains a challenging topic, especially when large deformations are involved and only sparse input data is available. This is typically the case when trying to provide an augmented view of an organ during surgery. In this context, just about 30% of the surface of the organ is visible due to the limited field of view of the laparoscopic camera or size of the incision [2]. Several works have demonstrated the benefits of physics-based models, particularly patient-specific biomechanical models, for accurate registration between different preoperative 3D anatomical model and intraoperative data [3,4,5,6]. These models are usually derived from continuum mechanics theory, where finite element (FE) methods are preferred due to their ability to numerically solve the complex partial differential equations associated with the constitutive models.

When considering augmented reality or, more generally, real-time tracking of the organ's deformation to provide up-to-date guidance, the computational

efficiency of the FE method is essential. To this end, various solutions have been proposed, with different trade-offs regarding the ratio between computation time and model accuracy [8,5,7]. The choice of the model, and its parameterization, are obviously key to an accurate registration. It is usually acknowledged that a registration of internal structures below 5mm is needed for best clinical impact, such as targeting relatively small tumors.

Without lack of generality, let us consider the scenario where an augmented view of the liver has to be provided during hepatic surgery. The intraoperative data is provided as a point cloud, reconstructed from the laparoscopic images [14,13] or direct camera view [9] of the operative field. Images are usually acquired at 20Hz or more, and latency between data acquisition and display of the augmented image needs to be minimized. This leads to update times of less than 50ms, during which image acquisition, image processing and model update need to take place. As a result, FE computation times should require less than 30ms. This is usually achieved using efficient implementation of specific elastic models.

In surgery where only small deformations take place, achieving such computation times is feasible, as illustrated in [10]. However, when considering large, non-linear deformations, involving typical soft tissue hyper-elastic behavior, computation times become incompatible with augmented reality constraints. A solution towards this objective has been to use a co-rotational technique, allowing to cope with large displacements (but small strain) behavior [4,9]. Yet, if more advanced laws from biomechanics are to be used, these optimizations (and simplifications) no longer hold. Alternative solutions have been proposed to avoid these restrictions. For instance, Marchesseau *et al.* [11] introduced the Multiplicative Jacobian Energy Decomposition (MJED) which is an alternative to the Galerkin FEM formulation. This method for discretizing non-linear hyperelastic materials on linear tetrahedral meshes can lead to fast and realistic liver deformations including hyperelasticity, porosity and viscosity. Also, Miller *et al.* [12] introduced TLED, an efficient numerical algorithm for computing deformations of very soft tissues. By using a total Lagrangian formulation, most spatial derivatives can be pre-computed, leading to fast simulations of non-linear material.

In this paper, we propose a method allowing for extremely fast and accurate simulations by using an artificial neural network that partially encodes the stress-strain relation in a low-dimensional space. Such a network can learn the desired constitutive model, and predict deformations under new, complex, conditions. As a result it can be used for augmented reality applications, where both complex tissue deformation and boundary conditions are involved. Our method consists in two main components: a deep network architecture, described in section 2.1, which is derived from model order reduction techniques, and an immersed boundary method, used for generating patient-specific simulations, using hexahedral finite elements 2.2. Results presented in section 3 show the efficiency of the method when applied to both synthetic and *ex vivo* scenarios.

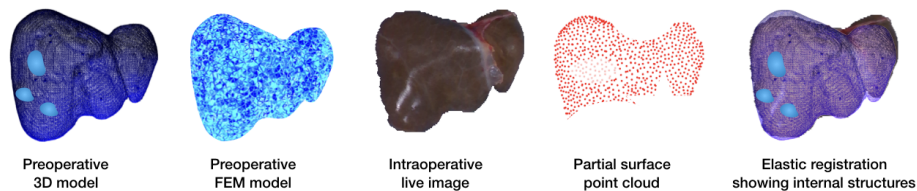


Fig. 1. Augmented reality pipeline: preoperative internal structures are mapped in real-time onto the live image of the organ using a FEM model.

2 Method

A typical pipeline for providing an augmented view of an organ during surgery is described by figure 1. The only assumption here is that we have a method for extracting a partial point cloud of the organ during surgery, either from a laparoscopic image or other system, such as RGB-D camera.

In such a pipeline we propose to accelerate the FE computation step by replacing it with a non-linear dimensionality reduction technique based on a U-Net architecture. Dimensionality reduction techniques have shown real benefits in speeding-up FE simulations. Among them, POD is a very popular one since it leads to very realistic real-time simulations. Although authors in [17] have shown good results using this method for large deformation, POD is better suited for linear or weakly non-linear processes. Our method uses a deep neural network to learn the relationship between sparse surface data and volumetric deformation of a given mesh.

2.1 U-Net architecture

The 3D point cloud provided by the RGB-D sensor is a sparse and partial view of the surface deformation of the organ. Our problem consists of finding the function f that produces the best estimation of the internal deformation of the organ from such point cloud. The function f is found by minimizing the expected error over a training set $\{(\mathbf{u}_s^n, \mathbf{u}_v^n)\}_{n=1}^N$ of N samples:

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^N \|f(\mathbf{u}_s^n) - \mathbf{u}_v^n\|_2^2. \quad (1)$$

where θ is the set of parameters of the network f , and \mathbf{u}_s and \mathbf{u}_v are the surface input and the volumetric output displacement fields of each sample. In order to characterize f , we propose to use the U-Net [1], a modified fully convolutional network initially built for precise medical image segmentations. As depicted in Fig. 2, the network is similar to an auto-encoder, with an encoding path to transform the input space into a low-dimensional representation, and a decoding path to expand it back to the original size. Additional skip connections transfer detailed information along matching levels from the encoding path to

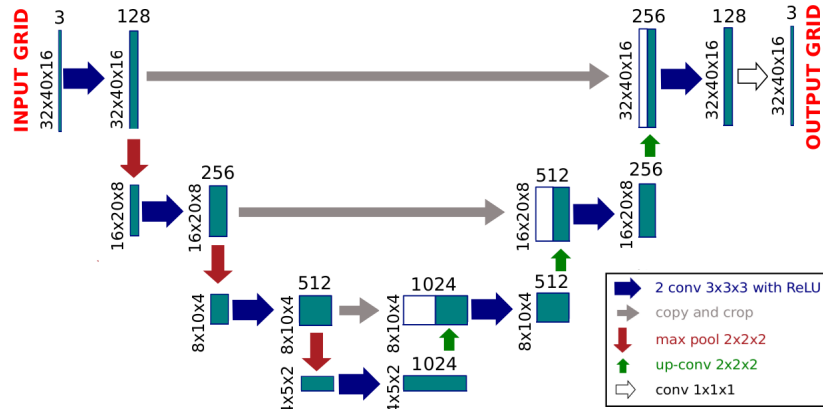


Fig. 2. U-Net architecture with 3 steps and 128 channels in the first layer for a padded input grid of size $32 \times 40 \times 16 \times 3$.

the decoding path. The encoding path consists of k sequences of two padded $3 \times 3 \times 3$ convolutions ($k = 4$ in [1]) and a $2 \times 2 \times 2$ max pooling operation (see Fig. 2). At each step, each feature map doubles the number of channels and halves the spatial dimensions. In the bottom part there are two extra $3 \times 3 \times 3$ convolutional layers leading to a 1024-dimensional array. In a symmetric manner, the decoding path consists of k sequences of an up-sampling $2 \times 2 \times 2$ transposed convolution followed by two padded $3 \times 3 \times 3$ convolutions. The features from the encoding path at the same stage are cropped and concatenated to the up-sampled feature maps. At each step of the decoding path, each feature map halves the number of channels and doubles the spatial dimensions. There is a final $1 \times 1 \times 1$ convolutional layer to transform the last feature map to the desired number of channels of the output (3 channels in our case). The up- and down-sampling process implies to use a grid-like structure for storing displacement information. Although this could be interpolated from a non-structured FEM mesh, we prefer here to rely on hexahedral mesh, as explained below.

2.2 Soft tissue deformation on hexahedral grids

To train our network, we need to generate many samples consisting of a volumetric output displacement field \mathbf{u}_v , given a nearly random input displacement \mathbf{u}_s . The simulation that generates this data needs to be as accurate as possible, yet computationally efficient to make it possible to generate data and train the network in less than a day (i.e. in the shortest amount of time available between preoperative data acquisition and surgery). While several choices of constitutive models can be made to describe soft tissue biomechanics, we want here to emphasize the choice of elements used in our FE implementation. Instead of using 4-node tetrahedral elements, as it is very often the case in the literature, we choose here to rely on 8-node hexahedral elements. This choice is governed by their better convergence, and lock-free behavior, especially with close to incom-

pressible materials (such as the liver) and strong shear stresses [19]. In addition, we combine this choice of elements with an immersed-boundary method, which allows us to create a hexahedral mesh directly from the segmented image and produces a sparse, regular grid of elements which matches the structure of the first layer of our U-net. Since most soft tissue exhibit an hyperelastic behavior,

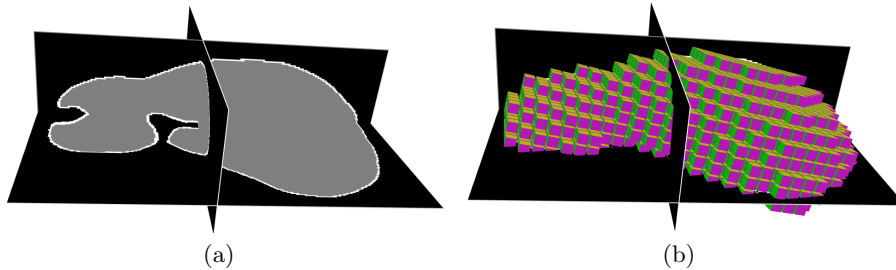


Fig. 3. Example of sparse grid discretization generated from a preoperative CT. Combined with an immersed boundary method, it allows the use of regular hexahedral meshes to simulate the deformation of the organ, and can be used with our CNN.

we chose to implement a Saint Venant-Kirchhoff model, for its relative simplicity and computational efficiency. This hyperelastic material model is just an extension of the linear elastic material model to the nonlinear regime. It is derived from Cauchy’s first law of motion, which states the conservation of linear momentum in a continuum. The second Piola-Kirchhoff stress tensor $\mathbf{S} = 2\mu\boldsymbol{\epsilon} + \lambda\text{tr}(\boldsymbol{\epsilon})\mathbf{I}$ conveys the amount of stress (N/m^2) sustained by the material undergoing a certain deformation, described by the strain tensor $\boldsymbol{\epsilon}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u}^T + \nabla\mathbf{u} + \nabla\mathbf{u}^T \cdot \nabla\mathbf{u})$. Here, λ and μ are the Lamé elasticity parameters and $\mathbf{u} = [u \ v \ w]^T$ is a displacement vector from position \mathbf{x}^0 in the undeformed state of the body to its deformed position $\mathbf{x} = \mathbf{x}^0 + \mathbf{u}$. Cauchy’s first law of motion can finally be translated into the system of partial differential equations $-\nabla \cdot \boldsymbol{\sigma} = \mathbf{f}_{ext}$ which pose our set of rules for the simulation process. The weak form of those rules, obtained from the principle of virtual work, brings forward the boundary terms and allows to solve the nonlinear system of equations of the Saint Venant-Kirchhoff model through a discretization of the domain into a finite set of hexahedral elements.

The choice of discretizing the initial domain with a sparse and regular grid brings several benefits. The usual tri-linear interpolation functions of an 8-nodes hexahedral element are reduced to a linear mapping, and, similar to 4-nodes tetrahedral elements, its jacobian remains constant inside the element. However, using such discretization requires particular care of the boundary elements. Volume integration of the displacement field inside these partially filled cells is carried on by recursively subdividing the cell into 8 sub-cells. The stiffness matrix inside a boundary cell is then accumulated from its sub-cells using the linear mapping, as in [16]. Since we are using a fine mesh of the domain, only one level of subdivision is enough to obtain an accurate approximation of the volume integral. Finally, the non-linearity of the strain tensor requires an iterative Newton-Raphson method to solve the non-linear system of equations approximating the unknown displacement. Since the convergence of the Newton-Raphson method

is only valid for a displacement near the solution, large external loads must be applied by small increments, which can in turn require a large number of iterations to converge. This is an important characteristic, as one would need to restrict those iterations during a simulation if such nonlinear model were to be used directly in the registration process. In our case, these computations are done in the training stage and do not need to meet real-time constraints.

2.3 Data generation for U-Net training

To train the U-Net, a data set of pairs $(\mathbf{u}_s, \mathbf{u}_v)$ is generated, where the surface displacement corresponds to the mapped point cloud. We assume that, at most, half of the organ surface is visible from the camera. Hence, we uniformly sample 100 points on the visible surface mesh. Hundred simultaneous forces are then applied to this virtual point cloud with random amplitudes and directions. In order to generate a series of surface deformations (see Fig. 4(b)), these random forces are applied to their enclosing grid nodes using their barycentric coordinates. Patient-specific parameterization of the biomechanical model is not required since for homogeneous materials, the relation between the surface and the volumetric displacements is independent of the stiffness of the object [21], and only depends on the Poisson’s ratio. As a result, we set the Poisson’s ratio to 0.49 (soft tissues are generally incompressible).

3 Results

Results below are obtained using an *ex vivo* human liver data set, for which surface data was obtained with an RGB-D camera and ground truth data acquired at different stages of deformation using a CT scan. Markers were embedded in the liver to compute Target Registration Errors (TRE).

Predictions on controlled synthetic deformations To validate the performance of our network, we compare the predictions of the displacement field with the corresponding FE simulation, using the model described in section 2.2. We generated 1,000 samples and used $N = 800$ samples to train the network by minimizing Eq. (1). The minimization is performed using the Adam optimizer: a stochastic gradient descent procedure with parameter-wise adjusted learning rates. The remaining 200 samples are then used for validation. For each sample, the average target registration error (TRE) is computed over 10 virtual markers. The obtained average TRE over the validation set is $\overline{\text{TRE}} = 1.96 \pm 3.46 \text{ mm}$.

Application to augmented reality With a PyTorch implementation of the U-Net running on a GeForce 1080 Ti, the network can predict the volumetric deformation of the liver in only 3 *ms*. We can then apply this prediction each time the RGB-D camera generates a point cloud. Before computing the displacement field, the point cloud needs to be cropped to the portion of the surgical image that contains the liver. This is done by segmenting the associated color image, similarly as in [15]. The RGB-D point cloud is then interpolated onto the grid

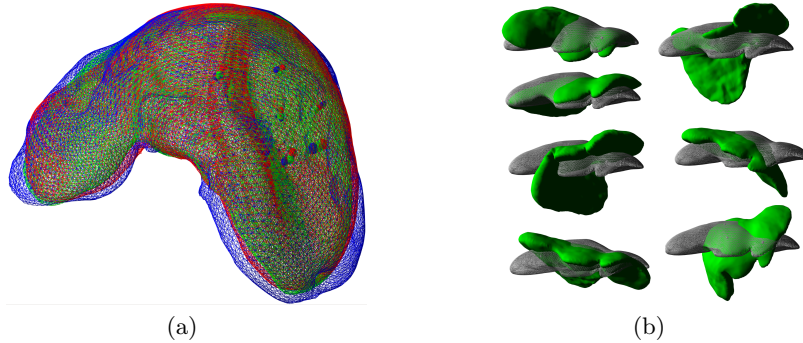


Fig. 4. (a) U-Mesh prediction in green, co-rotational FEM based registration in blue and ground truth in red. A mean TRE of 2.92 mm is achieved over the 10 markers in about 3 ms . (b) Samples of the generated deformations using our FE method. The rest shape is shown in gray.

to obtain per-node displacements on the surface (i.e. \mathbf{u}_s). Given this input, the network predicts the volumetric deformation, and the next point cloud can be processed. When compared to our *ex vivo* ground truth, the average TRE at the 10 markers is of only 2.92 mm with a maximal value of 5.3 mm . The same scenario, but this time using a co-rotational FE method, leads to an average TRE of 3.79 mm and is computed in about 25 ms . The solution of the Saint Venant-Kirchhoff model, used to train our model, gives nearly the same error (which was expected) but for a computation time of 1550 ms , even when using a very efficient linear solver (Pardiso³).

4 Conclusion

We have presented a deep neural network approach that can learn complex elastic deformations of a liver and generate its deformed state about 500x faster than a reference FE solution. Since the network takes as input a regular sparse grid where displacements are imposed, we have shown an efficient FE immersed-boundary method based on the same hexahedral discretization from which thousands of deformed configurations are generated to train the network. Based on a U-net architecture, which emulates a model order reduction technique, complex and accurate deformation of a preoperative organ model are computed in only a few milliseconds. Driven by surface displacement data, it makes this approach an ideal candidate for providing augmented reality during surgery.

³ <https://pardiso-project.org>

References

1. Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI, pp. 234-241
2. Plantefeve, R. et al. (2016). Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery. *Ann Biomed Eng.* 44 (1), pp. 139-153
3. Clements, L. W., Chapman, W. C., Dawant, B. M., Galloway, R. L., and Miga, M. I. (2008). Robust surface registration using salient anatomical features for image-guided liver surgery: Algorithm and validation. *Medical physics*, 35(6Part1), 2528-2540.
4. Haouchine, N., Dequidt, J. et al. (2013). Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. ISMAR, pp. 199-208
5. Suwelack, S., Röhl, S., Bodenstedt, S., Reichard, D. et al. (2014). Physics-based shape matching for intraoperative image guidance. *Med Phys.* 41(11):111901
6. Alvarez, P., et al. (2018). Lung deformation between preoperative CT and intraoperative CBCT for thoracoscopic surgery: a case study. *Medical Imaging*, Vol. 10576D.
7. Modrzejewski, R., Collins, T. et al. (2018). Soft-body registration of pre-operative 3d models to intra-operative RGBD partial body scans. MICCAI, pp. 39-46
8. Peterlík, I., Duriez, C., Cotin, S. (2012). Modeling and real-time simulation of a vascularized liver tissue. MICCAI, pp. 50-57
9. Petit, A., Cotin, S. (2018). Environment-aware non-rigid registration in surgery using physics-based simulation. ACCV - 14th Asian Conference on Computer Vision
10. Meier, U., López, O., Monserrat, C. et al. (2005). Real-time deformable models for surgery simulation: a survey. *Comput Methods Programs Biomed.*, 77(3), 183-197
11. Marchesseau, S., et al. (2010). Multiplicative jacobian energy decomposition method for fast porous visco-hyperelastic soft tissue model. MICCAI, pp. 235-242
12. Miller, K., Joldes, G., Lance, D., Wittek, A. (2007). Total Lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. *Communications in numerical methods in engineering*, 23(2), 121-134.
13. Collins, T., Pizarro, D., Bartoli, A., Canis, M. and Bourdel, N. (2013). Real-time wide-baseline registration of the uterus in monocular laparoscopic videos. MICCAI
14. Heiselman, J. S. et al. (2017). Characterization and correction of intraoperative soft tissue deformation in image-guided laparoscopic liver surgery. *J Med Imag*, 5(2)
15. Petit, A., Lippiello, V., Siciliano, B. (2015). Real-time tracking of 3D elastic objects with an RGB-D sensor. IROS, pp. 3914-3921
16. Düster, A. et al. (2008). The finite cell method for three-dimensional problems of solid mechanics. *Comput Methods Appl Mech Eng.* 197(45-48), pp. 3768-3782
17. Niroomandi, S. et al. (2008). Real-time deformable models of non-linear tissues by model reduction techniques. *Comput Methods Programs Biomed.* 91(3), pp. 223-231
18. Cifuentes, A. et al. (1992). A performance study of tetrahedral and hexahedral elements in 3-D finite element structural analysis. *Finite Elem Anal Des* 12 313-318
19. Benzley, S. E. et al. (1995). A comparison of all hexagonal and all tetrahedral finite element meshes for elastic and elasto-plastic analysis. 4th IMR, (17) pp. 179-191
20. Wang, E., Nelson, T., Rauch, R. (2004). Back to elements-tetrahedra vs. hexahedra. Proceedings of the 2004 international ANSYS conference
21. Miller, K., Lu, J. (2013). On the prospect of patient-specific biomechanics without patient-specific properties of tissues. *J Mech Behav Biomed Mater.* 27, pp. 154-166