

Force classification during robotic interventions through simulation-trained neural networks

Andrea Mendizabal, Raphael Sznitman, Stéphane Cotin

▶ **To cite this version:**

Andrea Mendizabal, Raphael Sznitman, Stéphane Cotin. Force classification during robotic interventions through simulation-trained neural networks. International Journal of Computer Assisted Radiology and Surgery, 2019, 14, pp.1601-1610. 10.1007/s11548-019-02048-3 . hal-02268782

HAL Id: hal-02268782

https://inria.hal.science/hal-02268782

Submitted on 21 Aug 2019

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.

International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-019-02048-3

Force Classification during Robotic Interventions through Simulation-trained Neural Networks

Andrea Mendizabal · Raphael Sznitman · Stephane Cotin

Received: 16 March 2019 / Accepted: 30 July 2019

Abstract

Purpose Intravitreal injection is among the most frequent treatment strategies for chronic ophthalmic diseases. The last decade has seen a serious increase in the number of intravitreal injections, and with it, adverse effects and drawbacks. To tackle these problems, medical assistive devices for robotized injections have been suggested and are projected to enhance delivery mechanisms for a new generation of pharmacological solutions. In this paper, we present a method aimed at improving the safety characteristics of upcoming robotic systems. Our vision-based method uses a combination of 2D OCT data, numerical simulation and machine learning to classify the range of the force applied by an injection needle on the sclera.

Methods We design a Neural Network (NN) to classify force ranges from Optical Coherence Tomography (OCT) images of the sclera directly. To avoid the need for large real data sets, the network is trained on images of simulated deformed sclera. This simulation is based on a finite element method and the model is parameterized using a Bayesian filter applied to observations of the deformation in OCT images.

Results We validate our approach on real OCT data collected on five *ex vivo* porcine eyes using a robotically-guided needle. The thorough parameterization of the simulations leads to a very good agreement between the virtually generated samples used to train the network and the real OCT acquisitions. Results show that the applied force range on real data can be predicted with 93% accuracy.

Conclusions Through a simulation-trained neural network, our approach estimates the force range applied by a robotically guided needle on the sclera based solely on a single OCT slice of the deformed sclera. Being real-time, this solution can be integrated in the control loop of the system, permitting the prompt withdrawal of the needle for safety reasons.

A. Mendizabal · S. Cotin
Inria, Strasbourg, France
E-mail: andrea.mendizabal@inria.fr, stephane.cotin@inria.fr

R. Sznitman
University of Bern, Bern, Switzerland
E-mail: raphael.sznitman@unibe.ch

Keywords Finite element modeling, Bayesian inference, Artificial neural networks, Force estimation in robotics

1 Introduction

Intravitreal injections are among the most common surgical interventions in ophthalmology with more than 4 million injections worldwide in 2014 alone [1]. This procedure is principally used in the treatment of diabetic maculopathy and for injecting vascular endothelial growth factor inhibitors in the treatment of age-related macular degeneration. Besides, we observe an increasing demand for such therapy due to the growing prevalence of diabetic patients and aging demographics. Intravitreal injections are mostly performed by doctors and the cost of such therapy has to be reduced. The increasing workload and the reduced reimbursement makes it difficult for hospitals to handle the situation. The time spent by the clinicians performing the injections has to be minimized while preserving the precision and the safety of the patient.

At the same time, robotic assistance in ophthalmology provides the ability to improve manipulation skills, along with shorter and safer surgeries [2]. To this end, Ullrich *et al.* [1] proposed a robotized intravitreal device capable of assisting injections into the vitreous cavity. However, designing such robotic systems requires to solve multiple challenges in terms of safety, cost and time efficiency and usability. The position accuracy and the orientation of the needle are of particular interest since the injection must be performed in a small region (the pars plana) of the eye. If the region is missed, damage of the eye lens or the retina might occur. In addition to accurate positioning, the ability to estimate or measure the force exerted by the needle on the sclera during the procedure could offer an important additional safety for the patient [3].

As a matter of fact the knowledge of the force plays a central role in the control loop of surgical robots for patient safety during robotic assisted interventions [4]. The force can be either measured through force sensors or estimated through vision-based methods. The use of force-based control algorithms allows for an improved human-machine interaction, more realistic sensory feedback and telepresence. Beyond this, force sensing or force estimation can facilitate the deployment of essential safety features [5]. A considerable amount of work relying on force sensors has previously been done, focusing on the development of miniaturized devices to ease their integration with actual systems. Force sensors usually need to meet several additional requirements, such as being water resistant, sterilizable and insensitive to changes in temperature [6]. The major limitation of conventional sensors is thus the associated cost since most surgical tools are disposable [5]. To overcome this point, alternative solutions have been proposed, such as qualitative estimation of forces based on images [7,6]. Mura *et al.* [7] introduced a vision-based haptic feedback system to assist the movement of an endoscopic device using 3D maps generated with a Shape-from-Shading method where the 3D shape of the surface is recovered from a 2D image of that surface. Haouchine *et al.* incorporated the use of a biomechanical model of the organ in addition to the 3D maps to estimate force feedback in robot-assisted surgery. This approach is, however, limited to the ability to evaluate tissue properties of the organ quantitatively [8].

Deep learning has already been suggested to improve existent characteristics of robotic assisted surgeries such as instrument segmentation and detection [9], as well as force estimation. For instance, in [8], interaction forces in minimally invasive surgeries are estimated with recurrent neural networks using camera acquisitions combined with kinematic variables and deformation mappings. In a follow-up paper, Aviles *et al.* [10] used a neuro-vision based approach for estimating applied forces in the same context. In [11], authors used two neural networks to classify muscle force exertion levels to prevent musculoskeletal disorders based on features extracted from video data and blood volume changes. However, in all these approaches an intermediate step to determine the surface deformation is required.

Our method consists of estimating contact forces directly from an OCT image of the scleral deformation without the need for a specific image feature extraction method beforehand, as in [10,8]. This work is built on top of our previous paper [12]. The technique relies on an image classifier for estimating force quantiles during robotized intravitreal injections. An imperative requirement for machine learning algorithms to work is the huge quantity of data to train on. Currently since intravitreal injections are executed manually, there is no available information on the force induced by the needle. Hence, we propose to build a biomechanical model of the sclera to generate a very large number of force-OCT image pairs, which are then used for training any supervised machine learning method. We will show that this approach allows us to avoid the need for large data sets of real OCT images. The simulations are parameterized to match experimental results and compensate for the absence of real data. The parameterization can be carried out with different approaches. In our previous work [12], we manually matched the simulations with some real acquisitions performing a grid search. Alternatively, simple optimization algorithms can be used to minimize the errors between the output of the simulation and the OCT image over the set of possible parameters. However, in the view of an online estimation of the parameters, we choose to use a Reduced-Order Unscented Kalman filter [13] to estimate the stiffness of the scleras which is the main contribution of this paper compared to [12]. We then train a two-layer image classifier algorithm with the generated images and their corresponding forces. This solution allows a straightforward force estimation process to take place. This two-stage process makes it possible to classify force ranges with 93% accuracy while requiring very few experimental data, as demonstrated on several *ex vivo* porcine eyes undergoing robotically-controlled needle insertions.

2 Method

In this work, we first construct a numerical model of the interesting portion of the eye to synthesize images of the deformed sclera under needle-induced forces. To be as close as possible to the actual organ behavior, we propose to use Bayesian inference as a means of identifying material parameters. Then, a data generation process takes place in order to train a neural network.

2.1 Numerical simulation

Biomechanical model We simulate the deformation of the sclera under needle-induced forces by modeling the eye as an elastic half-sphere subject to Dirichlet

boundary conditions (see Fig. 1(a)). Since applied forces and resulting deformations remain small, we choose to describe the stress-strain relationship as linear, using Hooke's law (1):

$$\sigma = 2\mu\epsilon + \lambda tr(\epsilon)I \tag{1}$$

where ϵ is the strain, σ the stress, and λ and μ are the Lamé's constants derived from the Young's modulus E and the Poisson's ratio ν. The Young's modulus is a measure of the stiffness of the material while the Poisson's ratio characterizes the compressibility of the material. The linearity of Hooke's law leads us to a simple relation $\sigma = \mathbf{C}\epsilon$ where $\mathbf{C}$ is the constitutive matrix for an homogeneous and isotropic material. Then, Dirichlet boundary conditions are added to prevent rigid body motion of the sclera, while a constant pressure is applied to the inner domain boundary to simulate the intraocular pressure (IOP), as illustrated in Fig. 1(a). The intraocular pressure plays an important role in the apparent stiffness of the eye and its variability is well studied, as high eye pressure can be an indication for glaucoma. It is worth-mentioning that the common IOP measurement devices are influenced by the stiffness of the eye. The value measured by the tonometer is not absolute but linearly related to the Young's modulus [14]. The external force due to IOP is given by $\mathbf{f_p} = S \times P$ where S is the surface area of the eye (in m^2) and P the intraocular pressure, given in Pa. Note that $\mathbf{f_p}$ is normal to the surface. To simulate the needle pushing on the sclera, we apply a local force $\mathbf{f_n}$ in a small region of interest near the virtual needle tip. The external forces of the system are then formed by the IOP, the force induced by the needle and the gravitational force $\mathbf{f_g}$ applied in the negative y direction,.

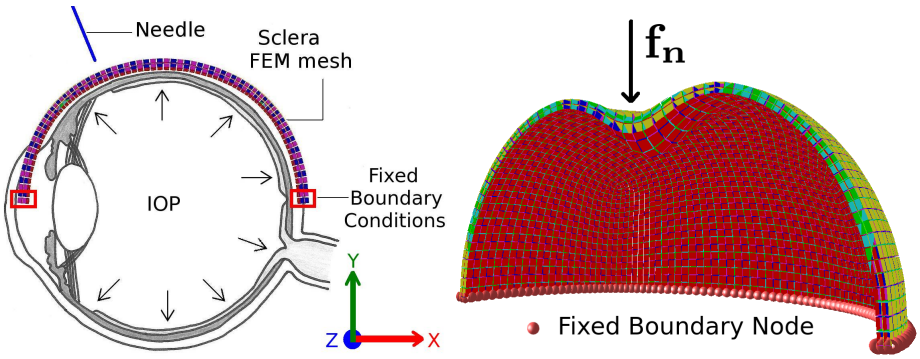


Fig. 1 (a) Simplified model of the sclera. (b) Deformed hexahedral mesh. Note that the colors of the hexahedra are just for visibility.

Finite Element simulation We solve the equation for the constitutive model using a finite element method. The eyeball is considered as nearly spherical and is discretized using a quadrilateral surface mesh of a half sphere of radius 12 mm, discretized using the Catmull-Clark subdivision method. The obtained quadrilateral surface is extruded conforming to the scleral thickness in order to generate almost regular hexahedra (see Fig. 1(b)). Scleral thickness plays an important role in the deformation of the sclera. Under a given force, a thick sclera is less deformed than a thin one, therefore it is important to take this thickness into account. Finally,

the deformation is specified by the nodal displacements $\mathbf{u}$ and the nodal forces $\mathbf{f}$, according to the following equation:

$$\mathbf{Ku} = \mathbf{f_p} + \mathbf{f_g} + \mathbf{f_n} \tag{2}$$

The matrix $\mathbf{K}$ is the stiffness matrix, and can be computed thanks to the elastic parameters of the material E and ν similarly as in [15].

To compute accurately the deformation of the sclera, the finite element mesh needs to be sufficiently fine (to avoid discretization errors). In our simulation, we use a mesh composed of $14,643$ hexahedral elements, resulting in about 5 seconds of computation time to obtain the static solution of the deformation. All our experiments are run on a Dell Precision laptop equipped with an intel Core i7 2.90GHz, a Quadro M1200 Graphics Processing Unit and 16Go of RAM. Since this computation has to be repeated thousands of times to generate the training data set, we take advantage of the linearity of the model and pre-compute the inverse of $\mathbf{K}$ to speed up the generation of the training data. This leads to a substantial computational speed-up (x10).

2.2 Model parameterization

Our finite element simulations depend on parameters of the constitutive model and the geometry of the eye: Young's modulus, Poisson's ratio, intraocular pressure and scleral thickness. It is therefore essential to understand which parameters can be assumed constant, and which ones vary (and within what range) to properly generate training data.

Data from the literature [16] report a thickness for corneal-scleral limbus in porcine eyes ranging from 630 μm to 1030 μm. We perform thickness measurements on porcine eyes and get a variation in thickness of more than 35%. Hence, we choose to consider the thickness as a parameter in the training, and we simulated scleras with five different thicknesses: 400 μm, 500 μm, 600 μm, 700 μm and 800 μm.

The IOP is also known to vary from patient to patient, and can be measured using a tonometer. So if we were to apply our method on a patient, we could use this information to parameterize the simulations used for the training. However, this is different for the porcine eyes used in our study. The IOP decreases gradually with postmortem time [19] and is halved only three hours after death. The IOP we measured ranged from 1 to 4 $mmHg$ which is very small and does not affect much the deformation of the sclera. Therefore we considered the IOP constant and equal to 2 $mmHg$ (266 Pa) for our experiments and simulations.

The Poisson's ratio, $\boldsymbol{\nu}$ is a constant value for the sclera according to the literature [17] and can be set to $\boldsymbol{\nu} = 0.45$. This leads to a nearly incompressible behavior during the deformation. With this in mind, we can reasonably assume that our training data set can be generated from numerical simulations in which the thickness, Poisson's ratio and IOP are known and constant.

The value of Young's modulus E, however, is more difficult to estimate as it varies depending on the porcine breed and experimental conditions. Therefore, values from the literature are not directly applicable. On the other hand, measuring it using a "classical" experimental biomechanics approach would be burdensome. For this reason, we propose to use a Bayesian approach to estimate the value of

E using observations from our OCT images. Such an approach could also be used to perform data assimilation on actual patients.

The knowledge of the elasticity parameters is key to build a good model of the sclera. The Young's modulus defines the ratio of stress to strain. The goal of this work being the estimation of the force range based on the shape of the sclera, it is therefore very important to correctly estimate E to avoid introducing errors in the force prediction. Since the exact value of E is not a priori known (only its average value based on data from the literature), we describe it as a stochastic parameter associated to a Gaussian probability density function (PDF). Initially $E \sim \mathcal{N}(\mu_0, \sigma_0)$ with μ_0 the mean value of E reported in the literature, and σ_0 its standard deviation. The aim of the assimilation process is to reduce the standard deviation σ in order to find the most likely value for μ. For this, the PDF of the parameter E is then transformed based on observations. Although our stress-strain relation is linear, we choose to model the transformation of the PDF using a reduced-order Unscented Kalman filter (ROUKF)[13] which can handle non-linear processes, and is computationally efficient.

2.3 Data set generation

Once the model is correctly parameterized, we generate a synthetic data set $\{(\mathbf{f_n}_k, I_k)\}_{k=1}^{N}$ of N samples where $\mathbf{f_n}_k$ is the needle-induced force for sample k and I_k is the corresponding simulated image of the deformation. The forces range from 0.0 N to 0.06 N and are applied at random locations and normally to the scleral surface for each of the different thicknesses. For each sample, a 2D cross-section of the complete 3D mesh is extracted in order to simulate the OCT image I_k. The simulated images have to be informative of the scleral thickness and its deformation, hence a binary image suffices for our purpose. Therefore the images are post-processed with a contour detection algorithm (*findContours* OpenCV function) and binarized (*threshold* function).

2.4 Neural network image classification

We look into modelling a function that can estimate the range of forces applied by a needle when observing a single OCT image and the deformation within it. To do this, we make use of an artificial neural network (NN) to provide a robust and reliable function capable of estimating forces applied to the sclera. In our case, the input to the model is the cross-sectional OCT image that depicts anatomical information of the sclera, from its surface to 1 millimeter below. In order to provide high-frame rate imaging, we opt to use 2D OCT cross-sections as the imaging modality over 3D volumetric OCT scans that are more common but slower to acquire. In our set up, we use the low-cost solution introduced in [18] to image 2D OCT cross-sections.

While we are interested in estimating the force applied by the needle, we choose to categorize the applied forces into three interval ranges. This reduces the need to be sensitive to exact force values, which not only is unnecessary in this instance, but also allows us to set up our inference task as a classification problem, whereby forces are grouped into ranges of clinical relevance. To do this, our NN consists

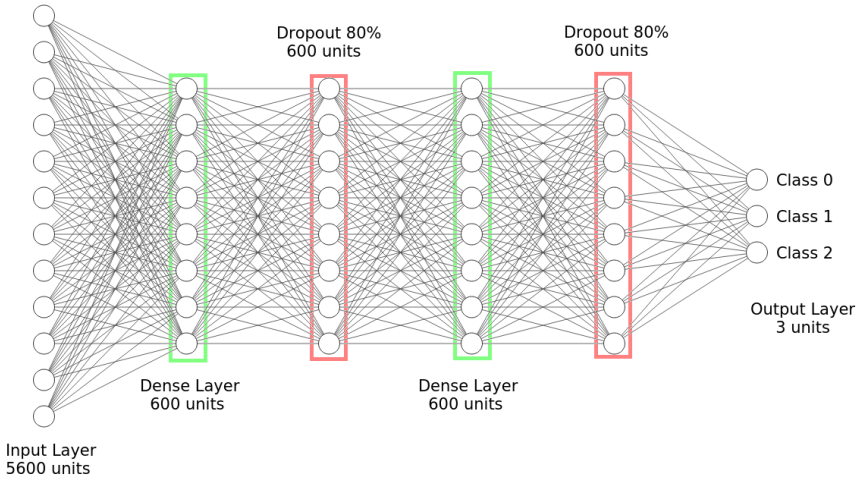


Fig. 2 Artificial neural network used for image classification. The input is an image of size 140×40 pixels resulting in an array of size 5600.

of two fully connected hidden layers and a classifier as output layer (see Fig. 2). The input layer has 5600 neurons (corresponding to the size of the input images 140×40) while the hidden layers have 600 neurons. These optimal values are found through a grid search. We use ReLu activations throughout the network and a softmax activation at the output layer. The network was implemented using Tensorflow and the Keras Python library.

Given the above simulation model presented in Sec. 2.1, we can train our NN with virtually an *infinite* amount of synthetically generated data. With the objective of stopping an incumbent needle from damaging the sclera, we use our simulation model to generate OCT images with forces between 0.0 N and 0.06 N applied on the Finite Element mesh of the sclera. This range was then divided into three ranges, or classes, consisting of:

- Class 0: force values smaller than 0.005 N
- Class 1: force values from 0.005 N to 0.03 N
- Class 2: force values bigger than 0.03 N

whereby class 0 corresponds to virtually no danger to the sclera, class 1 indicates a considerable force and class 2 is a dangerous force that should trigger a needle removal signal. The complexity of our problem makes it challenging to establish a clear cut between 'no danger to the sclera' and 'a needle removal is advised'. For this reason we introduced an intermediate class which can be seen as a gray area between the two bounding classes. If the force is classified in this gray area, it becomes the clinician's responsibility to stop or continue the insertion.

We train our NN from scratch using a gradient descent and the cross-entropy loss. A 0.8 dropout factor between layers was used to help with generalization.

3 Results

In this section, we first describe how data acquisition on *ex vivo* porcine eyes was performed. Then, we present the virtually generated data set, and associated training, with an emphasis on the parameterization of the biomechanical model. Finally, we validate the neural network on unseen real data to demonstrate the high level of accuracy of our approach at classifying contact forces.

3.1 Experimental set up

To validate the predictive accuracy of the NN on real data, we acquired a collection of $\{(\mathbf{f_n}_k, I_k)\}_{k=1}^{M}$ of M samples from *ex vivo* porcine eyes. Five porcine eyes were obtained from the local abattoir and transported to the experiment room while kept at low temperature. Experiments began within 3 hours of death, and we ensured they were completed less than 10 hours postmortem. During the tests, the eyes were moisturized with water and fixed with super glue on a 3D-printed holder to ensure fixed boundary conditions on the lower half of the eyeball. We choose super glue as it preserves the tightness of the eyeball. We then measured the IOP with a tonometer and obtained, for all the eyes, an IOP close to 2 $mmHg$ (about 266 Pa). The intraoperative pressure is lower than usual since it decreases dramatically after death [19]. However, we did not inject any fluid during the experiments to compensate for this low IOP.

A surgical robotic arm that was designed for high-precision drilling during robotic cochlear implantation [4] was used to measure the force applied by a needle. The robotic arm is fitted with a six-axis force/torque sensor (Mini40, ATI) at its wrist underneath a quick release for the end effector. After mounting a 22G Fine-Ject needle (0.7 x 30 mm) at the tip of the end effector, the robotic arm guided the needle forward along its axis. The needle was positioned as close as possible to the B-scan without intersecting it to bypass the generation of shadows in the OCT image. The margin between the needle and the B-scan was considered in the simulation. The robot was programmed to move towards the center of the eye along a path normal to the sclera. Contact and insertion forces, in the direction of motion, were continuously recorded during the insertion process. We also recorded the associated OCT images by storing B-scans over time. The custom-designed OCT device uses a 840 $nm \pm 40$ nm wavelength light source, with an A-scan rate of 50 kHz for a resulting 2D image of resolution 512×512 pixels with 12 bits per pixel, corresponding to an area of 15×4 mm^2. In these images, the black bands surrounding the OCT focus are removed leading to images of size 140×40 pixels.

In Fig. 3, a fragment of the collected data including 5 complete acquisitions is shown. To each trial corresponds one OCT image captured at a punctual time and a continuous flow of force values. The forces are filtered so that the noise is reduced. As the images were collected at a lower frequency than the forces, the corresponding forces were averaged over an interval of two seconds around the imaging time. The vertical lines correspond to the imaging times. The relaxation properties of soft tissue explain the slight decrease of the force a few seconds after the force is applied. The computed average force is used as ground truth for the following supervised learning. For each eye, one position near the corneal-scleral limbus was selected and the needle moves forward 0.5 or 1 mm in the same

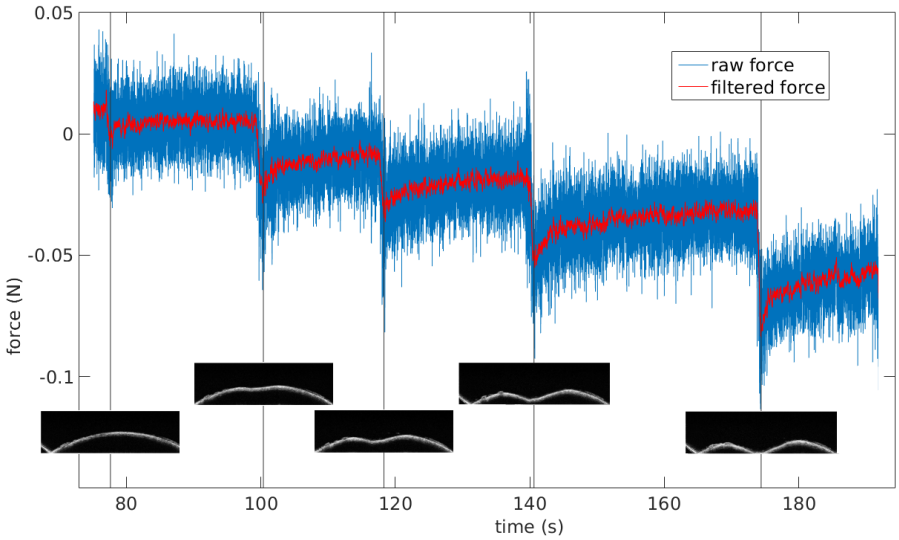


Fig. 3 Continuous acquisition of the robot force sensor and corresponding punctual OCT images for one eye. The vertical lines give the imaging times.

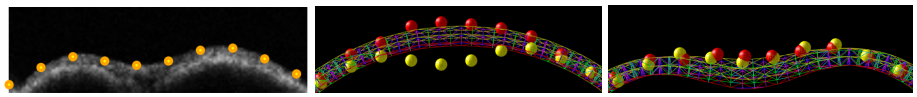
eye ID	1	2	3	4	5	Total
scleral thickness in μm	800	500	650	560	800	-
number of samples in class 0	0	1	2	2	2	7
number of samples in class 1	11	2	0	1	1	15
number of samples in class 2	14	2	3	2	2	23
Total	25	5	5	5	5	45

Table 1 Distribution of samples in the three force ranges and scleral thickness for each eye.

direction. Note that the pierced samples are excluded from the data set. Overall, 45 valid trials were performed on different eyes. The spread of each acquisition among the force classes is depicted in Table 1.

3.2 Stochastic identification of the Young's modulus

According to the collected porcine data, the Young's modulus E of the biomechanical model is estimated using the ROUKF. The PDE of E is transformed based on observations taken on the OCT images acquired during the experiment on an eye (see Fig. 4(a)). We consider the corresponding ground truth force (measured with the robot's force sensor) to run the simulations. According to [17], we set μ_0 and σ_0 to 0.49 MPa and 0.34 MPa respectively. We estimated the Young's modulus for each eye independently based on all OCT images where observation points can be extracted. We extract 10 observations on each OCT image for a given eye (see Fig. 4(a)). The finite element mesh is registered to the OCT image using rigid registration. To each observation in the OCT (yellow dots) corresponds a predicted observation attached to the model (red dots). During the assimilation, the known force is applied and the observations obtained are used by the filter



(a) (b) (c)

Fig. 4 (a) Observation points in the real OCT image. (b) Predicted observations (in red) and real observations (in yellow) for the first estimation of E and (c) for the final estimation of E.

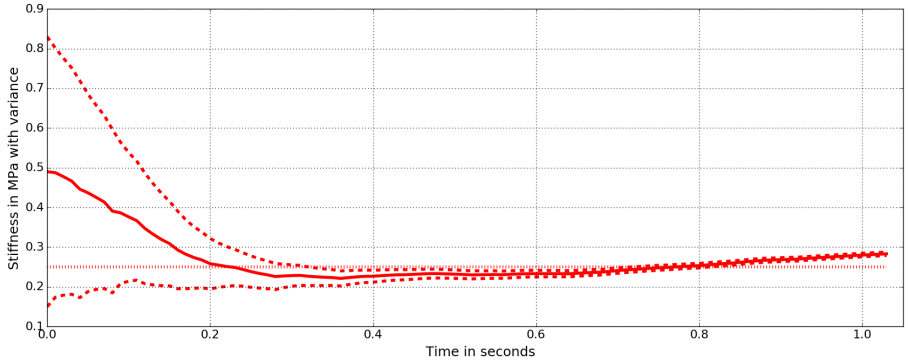


Fig. 5 Variation of μ and σ for the Young's modulus estimation using the ROUKF for 24 OCT images. The value of the parameter converges to 0.27 MPa.

correction phase to compute the Kalman gain. We employ the simplex version of the ROUKF, and since we only want to estimate one parameter, two evaluations of the model are performed in each prediction phase of the assimilation process (i.e. there are two sigma points). The evolution of μ and σ with the iterations of the ROUKF for eye 1 are shown in Fig. 5. After convergence, we see that the final value of μ is 0.27 MPa, as reported in first column of Table 2. When applied to other porcine eyes, estimated values of E vary only slightly, as seen in Table 2. Therefore we consider the Young's modulus E to be constant across all our experiments, with an average value of 0.25 MPa. Using this values leads to very good visual agreement between the OCT images and the simulations for the three force ranges for all the eyes (see Figs. 6 and 7) thus validating the assumption of linearity from section 2.1.

	eye 1	eye 2	eye 3	eye 4
μ in MPa	0.27	0.22	0.24	0.29
σ in MPa	$4.7e^{-3}$	$6.7e^{-3}$	$3.9e^{-3}$	$4.9e^{-3}$

Table 2 Mean and standard deviation of the estimated Young's modulus in 4 different porcine eyes after performing a data assimilation with a ROUKF. The fifth eye is not included since the observations required in the assimilation process are not consistent across the OCT acquisitions for this eye.

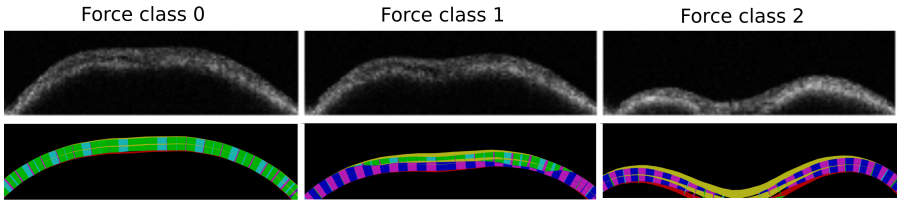


Fig. 6 Real OCT images of the eye used for the data assimilation and matching simulated images with $E = 0.25\ MPa$ for the three force ranges.

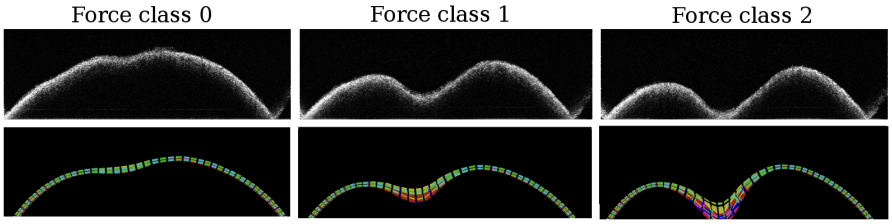


Fig. 7 Real OCT images of the sclera and their matching simulated images with $E = 0.25\ MPa$ for the three different force ranges.

3.3 Labelled data generation for neural network training

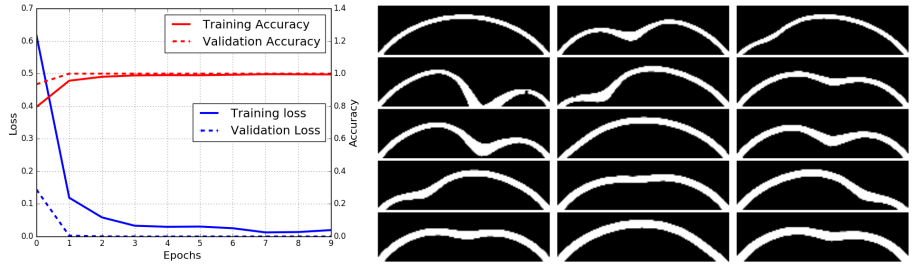
Data set generation To run our simulations we used the Simulation Open Framework Architecture[1]. We generated 3200 images of deformed scleras undergoing normal forces going from 0.0 N to 0.045 N for the smallest thickness. For each of the other four thicknesses, we generated 4000 images of deformed scleras where the forces vary from 0.0 N to 0.06 N at different random locations. For each thickness, the simulation took approximately half an hour. In Fig. 7 are shown different examples of the output of the simulation (bottom) matching the OCT images (top). Overall, we created a data set having 19200 synthetic images within 2 hours and a half (see Fig. 8(b)). In Fig. 7 below we show some synthetic images.

Neural network training The generated labelled data set is split such that 90% of the images is used for training and the remaining 10% is left for validation. All in all, 17280 images are used to train the artificial neural network that is validated on the other 1920 images. Fig. 8(a) depicts the accuracy and the loss of the neural network on both training and validation data sets over each epoch. The validation accuracy curve displays 100% accuracy when classifying unseen synthetic images. This curve is above the training accuracy curve probably because of the high dropout applied during the training.

3.4 Validation on unseen real data

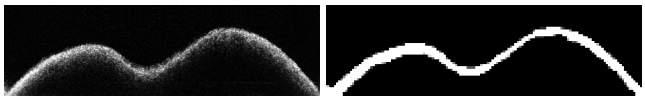
Our work aims at classifying force levels using only OCT data of the deformed scleras as input. All the OCT images collected during the experiments are filtered to

[1] https://www.sofa-framework.org/



(a) (b)

Fig. 8 (a) Loss and Accuracy curves for training and validation sets. (b) Fragment of the training data set generated by our numerical simulation.



(a) (b)

Fig. 9 (a) Unprocessed OCT image. (b) Filtered OCT image.

obtain inputs similar to the synthetic ones (see Fig. 9(a), Fig. 9(b) and Fig. 8(b)). Once the OCT images have been processed to look like the simulated ones, we can use them as input to the NN and perform force predictions. For each OCT acquisition, the force measured by the robot is converted into a class label (0, 1 or 2) and is taken as the ground truth (target class). A class with label 0 corresponds to a minimal force for which there is no risk of damaging the sclera. A class with label 1 corresponds to forces ranging from $0.005N$ to $0.03N$ and indicates that the sclera is being considerably deformed. A force class with label 2 means that the sclera is potentially being damaged and that a withdrawal of the needle is advised.

The performance of the classification is reported in the confusion matrix in Table 3. Each row of the table gives the instances in a target class, and each column gives the instances in a predicted class. For each category, we highlighted the correct decisions in red and show that the overall accuracy of the classifier is very high (93%). For all the experimental data set, the lowest score of the NN was obtained for the force class 0 with 71% accuracy. For target class 2, the precision reached 100%. In Fig. 10 are shown the raw force values (measured with the robot's force sensor) and the force range thresholds. The plot shows that the forces measured are uniformly spread throughout the force ranges. Note that the three misclassified samples are close to the upper bound of the force range and they are overestimated. We believe that misclassification of contact forces in the lowest ranges has a limited impact since the risk of damaging the sclera with such forces is almost null. On the other hand, it is essential that the forces of range 2 are predicted correctly, which appears to be the case on our (limited) data set.

Target \ Prediction	Class 0	Class 1	Class 2	Precision
Class 0	5	2	0	0.71
Class 1	0	14	1	0.93
Class 2	0	0	23	1.00
Recall	1.00	0.88	0.96	

Table 3 Confusion matrix

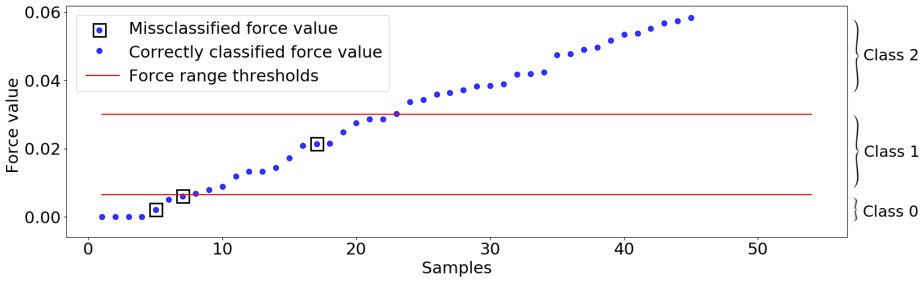


Fig. 10 Filtered force values corresponding to the 45 OCT acquisitions. The three misclassified samples are overestimated. Note that we only kept the samples under $0.06N$ which is the point of rupture of the thinnest scleras. To correctly account for the elements above that point of rupture, the puncturing process should be included in our model which would complicate the simulation without bringing any gain.

4 Conclusion and Discussion

This paper introduces a method for improving the safety features of robotized intravitreal injections. We show that our vision-based method, which combines numerical simulation and neural networks, can accurately predict the level of force applied by a needle, using only 2D OCT images of the scleral deformation. By being real-time, this classification can lead to an immediate withdrawal of the needle once it reaches a certain alarm threshold. To cope with the issue of performing a large number of experiments to populate our training data sets, the NN is trained on synthetic images generated from simulations of sclera deformations. To automatically parameterize the simulations from the experimental data, we use a Kalman filter which performs data assimilation using the sequence of OCT images.

In order to use our method in clinical practice (e.g. *in vivo* eyes) further simulation efforts should be performed. For instance, anatomically reasonable boundary conditions should be considered such as muscle fixation and eye movement should equally be included. It also seems essential to include the intraocular pressure as an additional input in the data set. Indeed, IOP is known to play a role in the deformation of the sclera and cornea of patients. To perform the prediction, a measure of the IOP would simply need to be performed on the patient using a tonometer. The reason we have not done it yet comes from the inaccuracy of the IOP measurement which is influenced by the stiffness of the eye. This coupling makes it more challenging to parameterize our simulations. Another improvement of the method would be to use adaptive force thresholds depending on the scleral

thickness and stiffness as for thinner or softer scleras, the puncture force is lower. The simulated images could also be improved to match actual surgical scenarios better. In particular, adding the shadows generated by the needle in the OCT image would be an essential feature. We also propose to address the sensitivity of the NN predictions to image framing and scaling by randomly cropping each simulated image, and augmenting the data set with these additional images. Note that in this paper, the optical interference phenomenon happening in OCT is not simulated. Instead, the synthetic and real images are binarized. This skeletonization is a way of filtering the complicated aspects of the the OCT images that are *a priori* non-informative for our purpose. Training a network directly from the OCT images without the skeletonization would lead to a significantly more complex problem requiring the simulation of light propagation and the use of Convolutional Neural Networks.

Finally, the objective of this paper was to estimate a force range, but it would also be pertinent to estimate the location and the angle of the force as a slight error in the latter might damage the retina or the eye lens. In this context we could augment the data set by including various needle insertion angles and locations.

Acknowledgements The authors would like to thank Jan Hermann and Tatiana Fountoukidou for the help in the experimental data gathering, and Igor Peterlik for his advice and code on Bayesian filtering.
Conflict of Interest: The authors declare that they have no conflict of interest. This article does not contain any studies with animals performed by any of the authors.
Informed consent: This articles does not contain patient data.

References

1. Ullrich, F., Michels, S., Lehmann, D., Pieters, R. S., Becker, M., Nelson, B. J. (2016). Assistive Device for Efficient Intravitreal Injections. Ophthalmic Surgery, Lasers and Imaging Retina, 47(8), pp. 752-762.
2. Meenink, H. C. M., Hendrix, R., Naus, G. J. L., Beelen, M. J., Nijmeijer, H., Steinbuch, M., van Oosterhout, E.J.G.M., de Smet, M. D. (2012). Robot-assisted vitreoretinal surgery. In Medical Robotics, pp. 185-209.
3. Jagtap, A. D., and Riviere, C. N. (2004, September). Applied force during vitreoretinal microsurgery with handheld instruments. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Vol. 1, pp. 2771-2773). IEEE.
4. Weber, S., Gavaghan, K., Wimmer, W., Williamson, T., Gerber, N., Anso, J., Bell, B., Feldmann, A., Rathgeb, C., Matulic, M., Stebinger, M.,Schneider, D., Mantokoudis, G., Scheidegger, O., Wagner, F., Kompis, M., Caversaccio, M. (2017). Instrument flight to the inner ear. Science robotics, 2(4).
5. Haidegger, T., Beny, B., Kovcs, L., Beny, Z. (2009, August). Force sensing and force control for surgical robots. In 7th IFAC Symposium on Modeling and Control in Biomedical Systems 7(1), pp. 413-418.
6. Haouchine, N., Kuang, W., Cotin, S., Yip, M. C. (2018). Vision-based Force Feedback Estimation for Robot-assisted Surgery using Instrument-constrained Biomechanical 3D Maps. IEEE Robotics and Automation Letters.
7. Mura, M., Abu-Kheil, Y., Ciuti, G., Visentini-Scarzanella, M., Menciassi, A., Dario, P., Dias, J., Seneviratne, L. (2016). Vision-based haptic feedback for capsule endoscopy navigation: a proof of concept. Journal of Micro-Bio Robotics, 11(1-4), pp. 35-45.
8. Aviles, A. I., Marban, A., Sobrevilla, P., Fernandez, J., Casals, A. (2014, October). A recurrent neural network approach for 3d vision-based force estimation. International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1-6. IEEE.

9. Pakhomov, D., Premachandran, V., Allan, M., Azizian, M., Navab, N. (2017). Deep residual learning for instrument segmentation in robotic surgery. arXiv preprint arXiv:1703.08580.
10. Aviles, A. I., Alsaleh, S., Sobrevilla, P., Casals, A. (2015, April). Sensorless force estimation using a neuro-vision-based approach for robotic-assisted surgery. In Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on, pp. 86-89. IEEE.
11. Aggarwal, V., Asadi, H., Gupta, M., Lee, J. J., and Yu, D. (2018). Covfefe: A Computer Vision Approach For Estimating Force Exertion. arXiv preprint arXiv:1809.09293.
12. Mendizabal, A., Fountoukidou, T., Hermann, J., Sznitman, R., Cotin, S. (2018, September). A Combined Simulation and Machine Learning Approach for Image-Based Force Classification During Robotized Intravitreal Injections. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 12-20.
13. Moireau, P., Chapelle, D. (2011). Reduced-order Unscented Kalman Filtering with application to parameter identification in large-dimensional systems. ESAIM: Control, Optimisation and Calculus of Variations, 17(2), pp. 380-405.
14. Hamilton, K. E., Pye, D. C. (2008). Youngs modulus in normal corneas and the effect on applanation tonometry. Optometry and Vision Science, 85(6), pp. 445-450.
15. BroNielsen, M., and Cotin, S. (1996, August). Realtime volumetric deformable models for surgery simulation using finite elements and condensation. In Computer graphics forum (Vol. 15, No. 3, pp. 57-66). Edinburgh, UK: Blackwell Science Ltd.
16. Olsen, T. W., Sanderson, S., Feng, X., Hubbard, W. C. (2002). Porcine sclera: thickness and surface area. Investigative ophthalmology & visual science, 43(8), pp. 2529-2532.
17. Asejczyk-Widlicka, M., Pierscionek, B. K. (2008). The elasticity and rigidity of the outer coats of the eye. British Journal of Ophthalmology, 92(10), pp. 1415-1418.
18. Apostolopoulos, S., Sznitman, R. (2017). Efficient OCT volume reconstruction from slit-lamp microscopes. IEEE transactions on biomedical engineering, 64(10), pp. 2403-2410.
19. Balci, Y., Basmak, H., Kocaturk, B. K., Sahin, A., Ozdamar, K. (2010). The importance of measuring intraocular pressure using a tonometer in order to estimate the postmortem interval. The American journal of forensic medicine and pathology, 31(2), pp. 151-155.