Towards learning quasi-static models of markerless deformable linear objects for bimanual robotic manipulation

Michał Bidziński*, Piotr Kicki* and Krzysztof Walas

I. Introduction

In everyday life, people encounter and skillfully manipulate Defromable Linear Objects (DLOs), such as cables, ropes, threads, strings, and hoses. It would be beneficial to give robots similar skills to enable them to perform surgical suturing [1], knot tying [2], wiring harness assembly in the automotive sector [3], or threading the lace through the narrow hole [4]. A typical approach to manipulate DLOs is to use their models to plan the motion and control commands necessary to rearrange it to the desired state. In recent years, there were many attempts to develop DLO models, such as FEM-based ones [5], [6], using Cosserat rod theory [7] or dynamic B-splines [8]. However, these models make strong assumptions about the properties of the objects or require significant amounts of computations to evaluate. A response to these problems was the trial to develop a neural network-based model [9], which can mimic the behavior of a particular DLO with one end rigidly attached, based on the careful measurements of the cubic markers located along the DLO, which seems rather an unrealistic scenario in the typical problems of DLO manipulation.

In this paper, we present a neural network-based approach to develop a quasi-static model of the real DLO being manipulated by a pair of robots without the use of any markers (see Figure 1). DLO shape is being tracked with an RGBD camera, and the upgraded approach proposed in [10]. Based on these measurements, we learn a machine learning model able to predict the state of a DLO, given an actual DLO state and an initial and final pose of the end-effectors (EEs) holding the DLO.

Our main contribution is the first data-driven 3D model of a DLO, that can be used without any markers on featureless DLOs. Our proposed model works in a partially observable setting, as it is impossible to directly detect the bend and twist of the raw DLO using just RGBD data, and shows the ability to accurately predict the behavior of the DLO. We evaluate a neural model of a DLO proposed in [9] (INBiLSTM), a simple fully connected architecture (FC), and analyze the impact of different data representations on the model's accuracy. We introduce a simple but effective data augmentation procedure that significantly improves the accuracy of both architectures and enables the FC network to achieve accuracy similar to INBiLSTM while being substantially more computationally efficient. Moreover, we show that the knowledge gained for one DLO is transferable for the same DLO handled at different points (different lengths) and to DLOs of different physical parameters. Furthermore, we evaluate the possibility of retraining a model pre-trained on a different DLO setup. To the best of the authors' knowledge, this is the first analysis of the transferability of the neural DLO models. To facilitate the research on DLO modeling, we make our code and datasets publicly available at `https://github.com/PPI-PUT/neural_dlo_model`.

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 870133, REMODEL.

All authors are with the Institute of Robotics and Machine Intelligence, Poznan University of Technology, Poznan, Poland {name.surname}@put.poznan.pl

* authors contributed equally

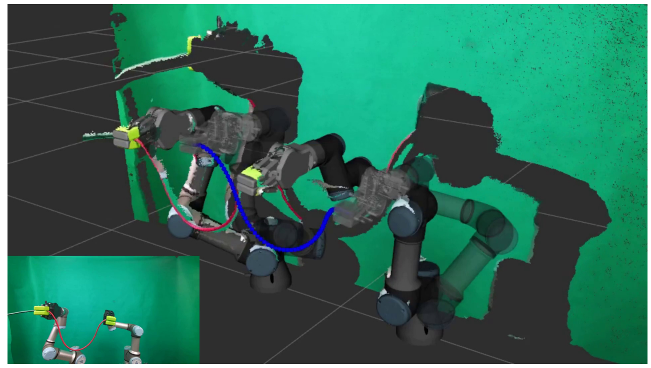

Fig. 1. Prediction of the markerless DLO shape in a different configuration of robotic arms.

II. Neural network-based quasi-static model of a DLO

Our goal is to learn from data a model f for predicting the next DLO state $s_n = f(s_{n-1}, p_{n-1}, p_n)$, given the prior DLO state s_{n-1} and an actual p_{n-1} and next pose p_n of the robots EEs holding the DLO. We define the state of DLO as a sequence of N points in 3D space. The pose of the robots EEs consists of the positions t and orientations R of the left and right robotic arm TCPs i.e. $p = (t_L, R_L, t_R, R_R)$. We consider only a quasi-static DLO manipulation, as it is enough to accurately describe slow DLO reconfigurations between steady-states [11].

A very important aspect of the problem we consider is that we limit the perception system to a single RGBD camera, and we assume markerless and textureless DLO, unlike the most state of the art approches [9], [12], [13], [14], [15], [16]. This assumption is motivated by the fact that in typical bimanual robotic manipulation settings, we cannot stick any markers to the DLO we want to manipulate and that many DLOs do not have detectable visual or geometrical features that could enable us to track the twist along them.

A. Data representations

To efficiently learn a model, appropriate data representations are essential. Particularly, we must define how to represent the state of a DLO and the orientation of the EEs. Regarding the representation of orientation, we analyzed three concepts, i.e., quaternions, rotation matrix, and axis angle. In terms of the DLO state, we use a sequence of 64 3D points. This representation does not describe the DLO state fully, as it lacks information about the twist along the DLO. In markerless and textureless settings, it is impossible to detect the twist from a sequence of 3D points accurately. However, we expect that some notion of the twist can be seen in the geometry of a given DLO, given enough data to differentiate between the deformations innate for a given DLO instance and the twist along it.

To introduce a translation invariance of the DLO shape, we represent the DLO as a sequence of 3D points expressed in the coordinate system located in the middle of the gripper pad (right was selected) and orientation aligned with the coordinate system of the right manipulator base to maintain the direction of gravity vector w.r.t. the DLO. Another approach to achieve the translation invariance is representing the DLO as a sequence of difference vectors between points. Finally, one has to decide how to represent the motion of EEs, whether by the initial p_{n-1} and end p_n poses (IE), or by initial pose p_{n-1} and a change of the pose $\delta p_n = p_n - p_{n-1}$ (Diff).

B. Neural network architectures

Having the data representations defined, we can focus on how to process them to predict the DLO motion accurately. Similarly to [12], we propose to use a neural network (NN) to approximate the function f. In this paper, we considered two NNs: INBiLSTM proposed in [9], adjusted to comply with the considered task, and a simple, fully connected architecture (FC). Both networks were trained to predict the change in the position of the points on the DLO between EEs states.

III. Dataset

A. Dataset collection

For detecting the DLO shape in the RGBD image, we used a DLOFTBs algorithm [10] acting on the DLO mask extracted based on a hue-based segmentation. The output of this method is a 3D B-spline curve representing the shape of the DLO. However, we observed that in a bimanual manipulation setting, the ends of the DLO are often not clearly visible to the camera. This makes it harder to stably track the points on the DLO. Therefore, we propose to utilize the information about the grippers handling a DLO and include their TCPs in the list of points on the DLO. To achieve a stable representation, we decided to fit a B-spline to the points on the DLO and grippers TCPs, and then compute N equally distant points on the DLO (where the distance is computed along the B-spline curve). Moreover, we observed that using the Kinect Azure sensor, the quality of the DLO depth estimation is reduced when the depth of the background is close to the DLO, especially at the ends of the DLO, as they are close to the grippers. To address this issue, we decided to neglect the points on ends of a DLO that lie too close to the grippers.

To collect the dataset we utilize 2 UR3 robots with custom 3D printed grippers and an Azure Kinect camera with the above-described DLO detection algorithm. We were collecting EEs poses and DLO states in sequences of 20 random moves of both arms with enforced constraints to prevent ripping the DLO. For each sequence, the initial state of the system was chosen manually to cover a wide spectrum of system configurations. Finally, we generated data points by choosing pairs of system configurations from the same sequence. As a result, each sample from the dataset consists of the initial and end EEs poses p_{n-1}, p_n and DLO states s_{n-1}, s_n.

In this way, we collected several datasets: (i) $50\,\mathrm{cm}$ long two-wire cable (40368 training / 6920 validation / 5736 test samples), (ii) $45\,\mathrm{cm}$ long two-wire cable (18950/4738/3220), (iii) $40\,\mathrm{cm}$ long two-wire cable (20596/3208/3628), (iv) $50\,\mathrm{cm}$ long solar cable (3262/298/474), and (v) $50\,\mathrm{cm}$ long braided cable (4198/530/812).

B. Data augmentation

In this paper, we propose a simple data augmentation technique that can be applied to introduce a critical inductive bias to the predictions of the neural network-based DLO model. We assumed that the considered DLO motions are quasi-static, and we would like to model them using a function $f(s_{n-1}, p_{n-1}, p_n)$. One of the critical properties we would like a function f to possess is to satisfy the following equality $f(s_{n-1}, p_{n-1}, p_{n-1}) = s_{n-1}$, i.e., if there is no change in the pose of the EEs then should be no change of the DLO state. Because this property is not automatically satisfied due to the architecture of our models, we propose to add to the dataset samples that represent the case of no motion of the arms at all DLO configurations from the dataset.

IV. Experimental evaluation

A. DLO shape prediction

We start our experiments with the comparison of FC and INBiLSTM DLO neural models and evaluation of all considered data representations on the dataset of $50\,\mathrm{cm}$ long two-wire cable. Moreover, we simultaneously want to evaluate the impact of the proposed data augmentation technique. The results of these experiments are shown in Figure 3. For comparison, we use the $\mathcal{L}_3$ error introduced in [10], related to the $\mathcal{L}_3$ error computed between initial and ground-truth cable pose after the move. We observed that the relative error obtained by the FC method without augmentation is higher than for INBiLSTM. However, when augmentation is used, then the performance of both models is very similar. Moreover, we see that augmentation improved the performance of all considered models such that there are almost no errors higher than 50%. Relative high error values may be caused by problems with an accurate estimation of

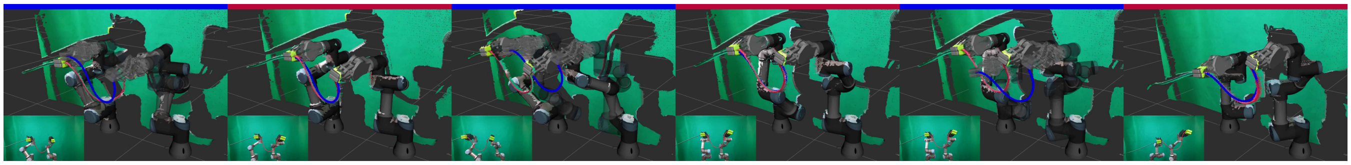

Fig. 2. Sequence of predictions (marked with blue) of the two-wire cable shape while being manipulated (movements marked with red) by two UR3 robotic arms.

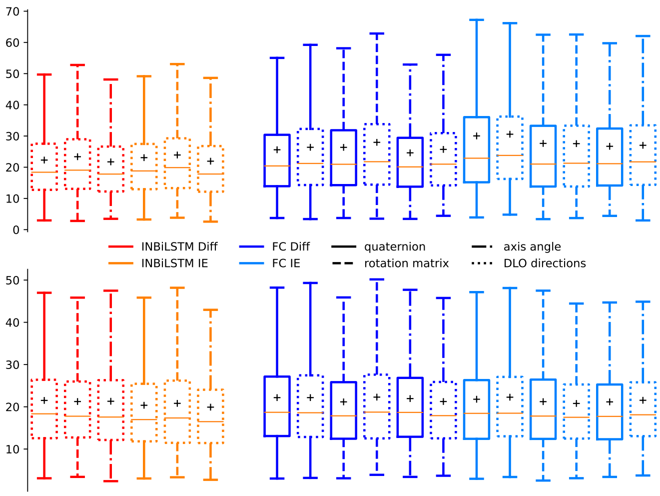

Fig. 3. Statictics of the relative error [%] of the neural DLO models (INBiLSTM – left, FC – right) with different data representations and augmentation (lower plot). Lighter colors denote IE representation, while darker the Diff representation. Whiskers denote the rotation representation: solid – quaternion, dashed – rotation matrix, dash-dotted – axis angle. In turn, the box style represents if the DLO is represented by the coordination of the points on it (solid) or by the differences of these points (dotted).

the cable shape in cases when the cable shape does not change much. While, in general, the way of representing the data does not have a major impact on the results achieved, we observe that representing the movements by initial and end configurations of grippers is slightly more effective for both models. Note that while both INBiLSTM and FC models achieve similar accuracy, the inference time is two orders of magnitude lower for FC. In Figure 2, we present the DLO shape predictions of the best-performing FC model for a sequence of bimanual robotic manipulation.

B. Generalization

1) DLO length: One of the commonly overlooked generalizations regarding the neural network-based DLO modeling is the cable length. In this experiment, we assess the impact of the pretraining on the 50 cm two-wire cable dataset on the performance on 40 cm and 45 cm long segments of the same cable. The results of this analysis are presented in Figure 4. One can see that even the model that was not trained on the shorter DLOs achieves relatively reasonable performance (about 35% of mean relative error). However, retraining on the specific training sets further improves it. While training on 0.1% of the 40 cm and 45 cm training sets (20 and 18 samples) is not enough to achieve significant improvement, the use of 1 or 10% results in the median of less than 20% of relative error. We can also see that the pretraining is beneficial regarding the relative error up to 10% of the training set. Nevertheless, in all cases, pretraining speeds up the training by at least 15 up to 100 times.

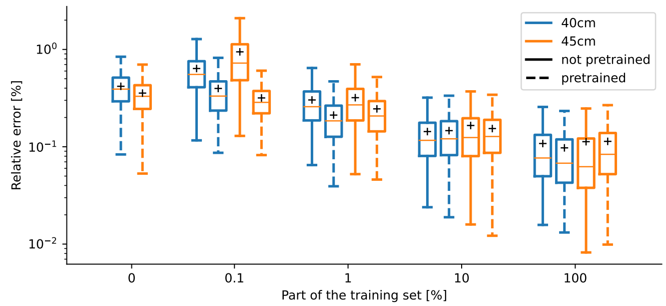

Fig. 4. Relative error statistics for 40 cm and 45 cm of two-wire cable datasets for different sizes of the training set. Dashed whiskers denote results obtained with the model pretrained on 50 cm dataset.

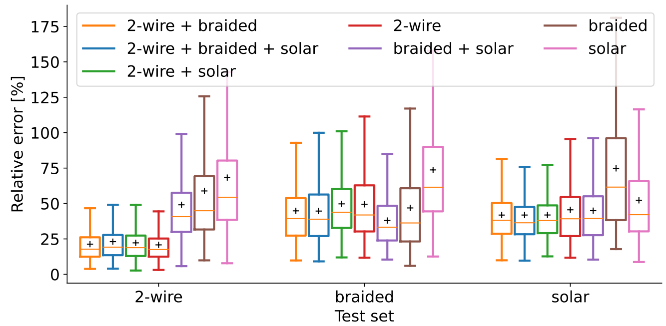

Fig. 5. Relative error statistics for models trained using combinations of 2-wire, braided, and solar cables datasets. Results for different DLO types

2) Different DLOs: Another very important ability of the model is to handle multiple DLOs types at once or even generalize to previously unseen ones. We evaluated these capabilities by training the NN on combinations of 3 datasets (2-wire, braided and solar cables) and evaluating the models on each of them (see Figure 5). One can see that, while in general, training on the other DLO type gives a rough estimate of the DLO behavior, to achieve the best performance, the training on the specific DLO seems inevitable. Nevertheless, for braided and solar cables, models trained on a dataset that does not include them obtained results close to the optimal ones.

V. Conclusions

In this abstract, we proposed a solution to quasi-static neural network-based markerless DLO modeling for bimanual robotic manipulation. We proposed a simple but efficient data augmentation procedure that allowed a very simple and computationally cheap neural-based DLO model to achieve the SotA performance. The proposed solution was evaluated on the introduced datasets with real-world measurements from an RGBD camera. We analyzed the potential of the method for generalization in terms of the DLO length and type.

REFERENCES

[1] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *2016 IEEE Int. Conf. on Rob. and Aut. (ICRA)*, 2016, pp. 4178–4185.

[2] M. Saha and P. Isto, "Manipulation planning for deformable linear objects," *IEEE Transactions on Robotics*, vol. 23, no. 6, pp. 1141–1150, 2007.

[3] P. Kicki, M. Bednarek, P. Lembicz, G. Mierzwiak, A. Szymko, M. Kraft, and K. Walas, "Tell me, what do you see?—interpretable classification of wiring harness branches with deep neural networks," *Sensors*, vol. 21, no. 13, 2021.

[4] W. Wang, D. Berenson, and D. Balkcom, "An online method for tight-tolerance insertion tasks for string and rope," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2488–2495.

[5] J. Sanchez, C. M. Mateo, J. A. Corrales, B. Bouzgarrou, and Y. Mezouar, "Online shape estimation based on tactile sensing and deformation modeling for robot manipulation," in *2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2018, pp. 504–511.

[6] A. Petit, V. Lippiello, and B. Siciliano, "Real-time tracking of 3d elastic objects with an rgb-d sensor," in *2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2015, pp. 3914–3921.

[7] J. Spillmann and M. Teschner, "Corde: Cosserat rod elements for the dynamic simulation of one-dimensional elastic objects," in *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '07, 2007, p. 63–72.

[8] G. Palli, "Model-based manipulation of deformable linear objects by multivariate dynamic splines," in *2020 IEEE Conf. on Industrial Cyberphysical Systems (ICPS)*, vol. 1, 2020, pp. 520–525.

[9] Y. Yang, J. A. Stork, and T. Stoyanov, "Learning differentiable dynamics models for shape control of deformable linear objects," *Robotics and Autonomous Systems*, vol. 158, p. 104258, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889022001518

[10] P. Kicki, A. Szymko, and K. Walas, "Dloftbs – fast tracking of deformable linear objects with b-splines," 2023.

[11] T. Bretl and Z. McCarthy, "Quasi-static manipulation of a kirchhoff elastic rod based on a geometric analysis of equilibrium configurations," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 48–68, 2014.

[12] Y. Yang, J. A. Stork, and T. Stoyanov, "Learning to propagate interaction effects for modeling deformable linear objects dynamics," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1950–1957.

[13] Y. Lai, J. Poon, G. Paul, H. Han, and T. Matsubara, "Probabilistic pose estimation of deformable linear objects," in *2018 IEEE 14th Int. Conf. on Aut. Sci. and Eng. (CASE)*, 2018, pp. 471–476.

[14] W. H. Lui and A. Saxena, "Tangled: Learning to untangle ropes with rgb-d perception," in *2013 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013, pp. 837–844.

[15] A. Borum, D. Matthews, and T. Bretl, "State estimation and tracking of deforming planar elastic rods," in *2014 IEEE Int. Conf. on Rob. and Aut. (ICRA)*, 2014, pp. 4127–4132.

[16] Y. Huang, C. Xia, X. Wang, and B. Liang, "Learning graph dynamics with external contact for deformable linear objects shape control," *IEEE Robotics and Automation Letters*, pp. 1–8, 2023.