Deep Domain Adaptation Regression for Force Calibration of Optical Tactile Sensors

Zhuo Chen[1], Ni Ou[1,2], Jiaqi Jiang[1] and Shan Luo[1]

***Abstract*— Optical tactile sensors provide robots with rich force information for robot grasping in unstructured environments. The fast and accurate calibration of contact forces holds significance for new sensors and existing sensors. However, the conventional nerual-netwrok-based force calibration method necessitates a large volume of force-labeled tactile images to minimize force prediction errors with the need for a time-consuming data collection process. To address this challenge, we propose a novel deep domain-adaptation force calibration method to transfer the force prediction ability from a calibrated optical tactile sensor to uncalibrated ones with various combinations of domain gaps, including marker presence, illumination condition, and elastomer modulus. Experimental results show the proposed unsupervised force calibration method can achieve lowest force prediction errors of 0.102N (3.4% in full force range) for normal force, and 0.096N (6.4%) and 0.062N (4.1%) for shear forces along the x-axis and y-axis, respectively.**

I. INTRODUCTION

Optical tactile sensors [1] with high sensitivity in perceiving object geometry, slip and position have now been widely studied. Among those touch sensations applicable for measurement by optical tactile sensors, force sensing stands out as it is crucial for monitoring dynamic contact status and providing feedback for robot control. However, force calibration in optical tactile sensing faces three primary challenges. Firstly, deep neural networks to estimate forces from tactile images necessitates a substantial volume of tactile images paired with labeled forces for model training. Secondly, alternations in sensor components, such as degradations of soft elastomers, can lead to inaccuracies in force predictions. Thirdly, differences in marker presence/distributions, illumination conditions and elastomer modulus in different optical tactile sensors prevent force prediction models from being adapted from calibrated sensors to uncalibrated ones.

To address these challenges, we propose a novel domain adaptation regression methods for models trained on existing sensors (source domain) to adapt to new sensors (target domain) with diverse domain gaps and unlabeled images, resulting in unsupervised force calibration with high accuracy. This approach can eliminate the need for costly force/torque measurement tools when calibrating new sensors or recalibrating old sensor, and significantly reduce the calibration time compared with fine-tuning [2] methods, which requires extensive labeled force information from both domains.

[1]Zhuo Chen, Ni Ou, Jiaqi Jiang and Shan Luo are with the Robot Perception Lab, Centre for Robotics Research, Department of Engineering, King's College London, London WC2R 2LS, United Kingdom. Emails: {zhuo.7.chen, shan.luo }@kcl.ac.uk.

[2]Ni Ou is with the State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing, 100081, China.

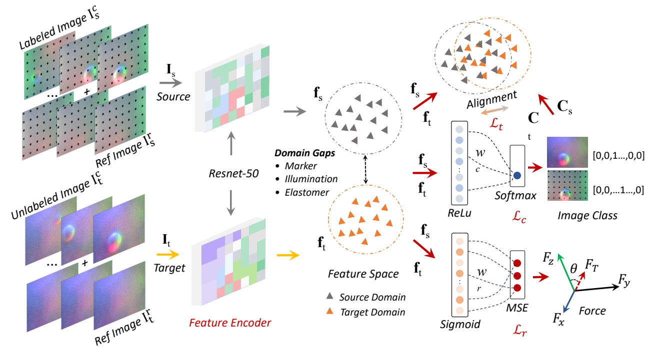

Fig. 1. Domain adaptation model for force calibration of tactile sensors.

II. METHODOLOGY

In the domain adaptation regression problem, we are provided with a labeled tactile image dataset $\mathcal{D}_s = \{(\mathbf{I}_s^i, (\mathbf{F}_s^i, \mathbf{C}_s^i))\}_{i=1}^{n_s}$ as the source domain with n_s samples, and an unlabeled dataset $\mathcal{D}_t = \{(\mathbf{I}_t^i)\}_{i=1}^{n_t}$ as the target domain with n_t samples. Here, $\mathbf{I}_s^i$ and $\mathbf{I}_t^i$ represent the tactile images from the source domain and the target domain. $\mathbf{F}_s^i = (F_x^i, F_y^i, F_z^i)_s$ denotes the ground truth of the applied forces vector, while $\mathbf{C}_s^i$ indicates the contact class of the tactile images in the source domain. $\mathcal{D}_s$ and $\mathcal{D}_t$ are sampled from different sensors , i.e. $P(\mathbf{I}_s) \neq P(\mathbf{I}_t)$, whereas the objective is to learn a shared regressor $h : \mathbf{I} \rightarrow \mathbf{F}$ capable of directly mapping the i_{th} unlabeled tactile images $\mathbf{I}_t^i$ to calibrated forces vector $\mathbf{F}_t^i$ in the target domain. Our goal is to minimize the Mean Square Error (MSE) between the predicted force and ground truth forces on the labeled samples and the domain distribution discrepancy between both domains:

$$\arg\min_h \lambda_r \underbrace{\frac{1}{n_s}\sum_{i=1}^{n_s}\left\|\hat{\mathbf{F}_s^i} - \mathbf{F}_s^i\right\|_2^2}_{\mathcal{L}_r} + \lambda_t \underbrace{\hat{d}_{\mathcal{H}}(\mathbf{f}_s, \mathbf{C}_s, \mathbf{f}_t, \mathbf{C}_t)}_{\mathcal{L}_t} \quad (1)$$

where $\lambda_r \geq 0$, $\lambda_t \geq 0$ represent the weights assigned to the regression loss $\mathcal{L}_r$ and domain transfer loss $\mathcal{L}_t$. $\hat{d}_{\mathcal{H}}(\cdot,\cdot)$ is the estimation of Local Maximum Mean Discrepancy (LMMD) [3], with $\mathcal{H}$ being the Reproducing Kernel Hillbert Space (RKHS), and $\mathbf{f}_s, \mathbf{f}_t$ being the feature vectors. The overall loss function $\mathcal{L}_h$ of our model is designed as:

$$\mathcal{L}_h = \lambda_r\mathcal{L}_r + \lambda_c\mathcal{L}_c + \lambda_t\mathcal{L}_t \quad (2)$$

where $\lambda_c \geq 0$ represents the weight of classification loss $\mathcal{L}_c$ for calculation of LMMD. By optimizing the loss function in Equation 2 with the model illustrated in Fig. 1, a shared regressor can be learned for both the source domain and target domain while maintaining a well-aligned feature space.

III. EXPERIMENT RESULTS & ANALYSIS

The data collection setup comprises a UR5e robotic arm with a two-finger Robotiq gripper, a sphere indenter (d=3 mm), a flat-surface GelSight sensor, and a Nano17 F/T sensor. Marker presence, illumination condition, and gel elastomer are three essential domain variables of different GelSight sensors, which are denoted as wm/m, i and b respectively. For data collection, three different elastomers and illumination conditions indexed by 0, 1, 2 are used. We obtain four types of labeled tactile images, i.e., mb_0i_0, wmb_0i_0, wmb_0i_0 and wmb_1i_1, and pair them into nine domain adaptation groups. All the tactile images are collected from a force range of -3 N to 0 N for normal force in the z-axis and a range of -0.75 N to 0.75 N for shear forces in the x-axis and y-axis. We have three groups (mb_0i_0, wmb_0i_0, wmb_1i_1) that comprise a total of $3 \times 10,830$ images, along with 873 images of wmb_2i_2.

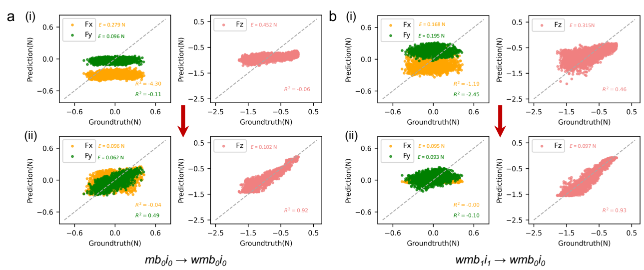

Fig. 2. Force prediction errors of domain adaptation groups featuring **one** variable (a) and **two** variables (b) by using source-only method (i) and our domain adaptation method (ii)

We demonstrate the domain adaptation performance of our model. The baseline of this task is the source-only method, which directly predicts forces in the target domain using the pretrained models from source domain. As shown in Fig. 2a-i, if we employ the source-only method directly from mb_0i_0 to wmb_0i_0, the force prediction performance is unacceptable, with R^2 values of -4.3, -0.11, and -0.06 in the xyz-axis, respectively. The error E for normal force can even reach up to 0.452N (15% of the force range in the z-axis), while for the shear forces, the errors are 0.279 N (18.6%) and 0.096 N (6.4%), respectively. After applying our method, as depicted in Fig. 2a-ii, the errors E of normal force decrease to 0.102 N (3.4%), while the R^2 values increase to 0.92. This represents an increase in accuracy of more than 10% after domain adaptation. Although the R^2 values for shear force still appear low, the MAE errors in the x-axis decrease by 12.2% from 0.279 N (18.6%) to 0.096 N (6.4%) in the $mb_0i_0 \rightarrow wmb_0i_0$ transition, which is acceptable when compared with the shear force error of around 0.025 N in the supervised model. For two variables, as shown in Fig. 2b-i, the average errors using the source-only method remain high, around 0.285 N on average across all four groups. Upon utilizing our model, the average errors improve significantly, with notable improvements observed in the groups $wmb_1i_1 \rightarrow wmb_0i_0$ shown in Fig. 2b-ii, where the R^2 value increase from 0.46 to 0.93 while force error drops from 0.315N to 0.097N in z-axis.

Table I shows the average force prediction error of normal force and shear forces with three variables. All of the tested domain adaptation groups get improved for the force prediction error after using our methods, especially for $wmb_1i_1 \rightarrow mb_0i_0$ improved from 0.368N to 0.105N. Specially, for the shear force in $wmb_1i_1 \rightarrow mb_0i_0$, the force error get notably improved, with E decreasing from 0.296 N (19.7%) to 0.098 N (6.5%) in x-axis, from 0.362 N (24.1%) to 0.103 N (6.9%) in y-axis, and from 0.446 N (14.9%) to 0.114 N (3.8%) in z-axis. Fig. 3 further illustrates that the feature representations of source and target domains are successfully aligned in feature space after domain adaptation, compared with the distinctly separated feature space observed when using the source-only method.

TABLE I

AVERAGE FORCE PREDICTION ERROR WITH **THREE** VARIABLES (MAKER & ELASTOMER & ILLUMINATION, UNIT N)

Methods	$mb_0i_0 \rightarrow wmb_1i_1$	$mb_0i_0 \rightarrow wmb_2i_2$	$wmb_1i_1 \rightarrow mb_0i_0$	Avg
source-only	0.323	0.304	0.368	0.332
ours	0.160	0.195	0.105	0.153

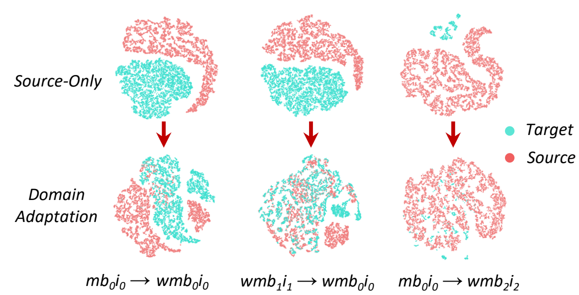

Fig. 3. Feature spaces visualized using tSNE.

IV. CONCLUSION

In this study, we propose a deep domain adaptation regression method for force calibration of optical tactile sensors. This method is promising to reduce the time consumed in the data collection process and eliminate the use of labeled tactile images collected from expensive force/torch sensors. We also study domain gaps of marker presence, illumination conditions, and elastomer modulus on domain transfer performance. We believe that this work not only provides a method for force calibration of GelSight sensors but also holds promise for enhancing the performance of various optical tactile sensors.

REFERENCES

[1] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[2] C. Sferrazza and R. D'Andrea, "Transfer learning for vision-based tactile sensing," in *IROS*, 2019, pp. 7961–7967.

[3] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, "Deep subdomain adaptation network for image classification," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 4, pp. 1713–1722, 2020.