TransForce: Transferable Force Prediction for Vision-based Tactile Sensors with Sequential Image Translation

Zhuo Chen, Ni Ou, Xuyang Zhang and Shan Luo

***Abstract*— Vision-based tactile sensors (VBTSs) provide high-resolution tactile images crucial for robot in-hand manipulation. However, force sensing in VBTSs is underutilized due to the costly and time-intensive process of acquiring paired tactile images and force labels. In this study, we introduce a transferable force prediction model, TransForce, designed to leverage collected image-force paired data for new sensors under varying illumination colors and marker patterns while improving the accuracy of predicted forces. Our model effectively achieves translation of tactile images from the source domain to the target domain, ensuring that the generated tactile images reflect the illumination colors and marker patterns of the new sensors while accurately aligning the elastomer deformation. As such, a recurrent force prediction model trained with generated sequential tactile images and existing force labels is employed to estimate forces for new sensors with lowest average errors of 0.69N (5.8% in full work range) in x-axis, 0.70N (5.8%) in y-axis, and 1.11N (6.9%) in z-axis.**

I. INTRODUCTION

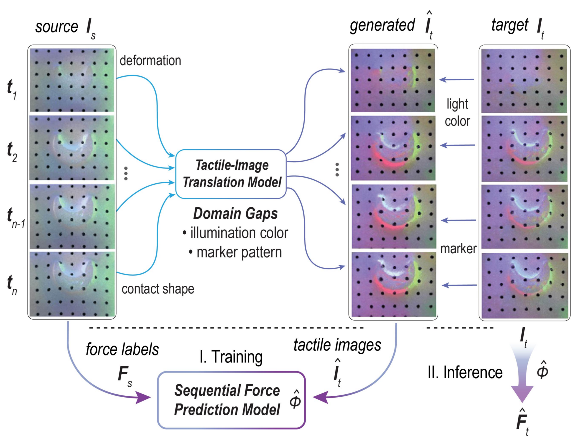

Fig. 1. Transferable force prediction model for VBTSs.

Vision-based tactile sensors [1] have now been widely used in robot manipulation. By measuring high-resolution tactile images, robots equipped with VBTSs are endowed with the sense of touch to complete in-hand tasks with human-like dexterity, such as slip detection [2] and dexterous manipulation.

Recently, deep learning models [3] are developed to map tactile images to force values without the aid of a physical model. Training these image-force mapping models requires a substantial amount of paired image-force data, along with costly calibration instruments like 6-DoF force/torque sensors. This labor-intensive data collection process must be repeated frequently due to wear and tear of soft elastomer and changes in sensor components, such as camera and LED. Therefore, there remains a demand for a transferable force prediction method with high-accuracy for VBTSs that can effectively predict both normal and shear forces.

In this study, we propose a novel transferable model, TransForce, to address the challenge of unsupervised adaptation in force prediction for VBTSs. As illustrated in Fig. 1, TransForce translates the style of tactile images from a source domain, where force labels are available, into the target domain, representing new sensors. This process successfully preserves the sequential deformation information from the source domain while adapting image properties, such as illumination color, intensity, and marker pattern, to the target domain. Consequently, a recurrent force prediction model can be trained using the generated sequential tactile images and existing force labels, enabling accurate force estimation with tactile images from new sensors.

II. METHODOLOGY

In this problem, we are given two sets of tactile data from two VBTSs shown in Fig. 1: one that has pairs of tactile images and forces $\{\mathbf{I}_s^i, \mathbf{F}_s^i\}_{i=1}^{n_s}$, named source domain $\mathcal{S}$, and the other one that only has tactile images $\{\mathbf{I}_t^i\}_{i=1}^{n_t}$, named target domain $\mathcal{T}$. The goal is to align the image styles of $\mathbf{I}_s$ and $\mathbf{I}_t$ so that we can map $\mathbf{I}_s$ to $\hat{\mathbf{I}}_t$, which shares the same style with $\mathbf{I}_t$. To this end, we can use the image-force pairs $\{\hat{\mathbf{I}}_t^i, \mathbf{F}_s^i\}_{i=1}^{n_s}$ derived from $\mathcal{S}$ to train a mapping function $\hat{\phi}$. This function $\hat{\phi}$ is able to estimate forces $\{\mathbf{F}_t^i\}$ in $\mathcal{T}$ using the tactile images $\{\mathbf{I}_t^i\}_{i=1}^{n_t}$.

III. DATA COLLECTION AND IMPLEMENTATION

As illustrated in Fig. 2a, the real-world setup for collecting tactile image and force pairs comprises five main components. Before data collection, we fabricate two GelSight sensors using the same silicone elastomer (XP-565, ratio A:B = 1:15, size 10×8 mm^2, thickness 3 mm), but with different marker patterns and illumination colors. We employed five contact points on the surface due to a larger size of 3D-printed indenter, which adequately contacts the entire surface of soft elastomer. The contact motion was divided into four stages: downward movement, horizontal movement, inverse horizontal movement, and upward movement. The normal force range was from -16 N to 0 N, while the shear force range was from -6 N to 6 N.

Zhuo Chen, Ni Ou, Xuyang Zhang and Shan Luo are with the Robot Perception Lab, King's College London, London WC2R 2LS, United Kingdom. Emails: {zhuo.7.chen, shan.luo}@kcl.ac.uk.

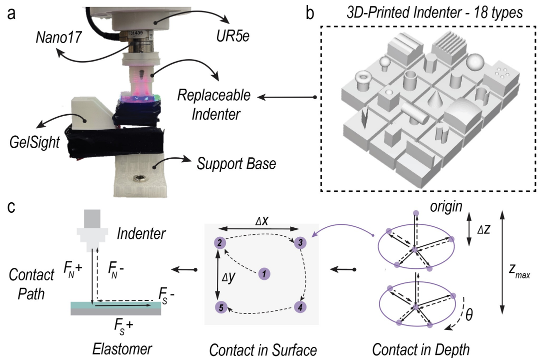

Fig. 2. Real-world setup for data collection.

IV. EXPERIMENTAL RESULTS

A. Image Translation Result

Fig. 3a and Table I demonstrates the image translation results. In *seen* group, the generated tactile images $\hat{\mathbf{I}}_\mathbf{t}$ closely match $\mathbf{I}_t$ in illumination colors and marker patterns, resulting approximately four times similarity on FID and KID when comparing $\mathbf{I}_s$ with $\mathbf{I}_t$ and comparing $\mathbf{I}_s$ with $\hat{\mathbf{I}}_\mathbf{t}$. To test the generalizability, we evaluate the model on the *unseen* group shown in Fig. 3a, which includes 6 indenters with varying aspect ratios and shapes. The results demonstrate that the model effectively translates marker patterns and illumination colors.

TABLE I
IMAGE TRANSLATION EVALUATION

group	$\mathbf{I}_s$ - $\mathbf{I}_t$		$\mathbf{I}_s$ - $\hat{\mathbf{I}}_t$		$\mathbf{I}_t$ - $\hat{\mathbf{I}}_t$	
	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓
seen	81.1	0.126	85.4	0.133	**18.3**	**0.019**
unseen	87.9	0.134	92.4	0.144	**21.3**	**0.021**

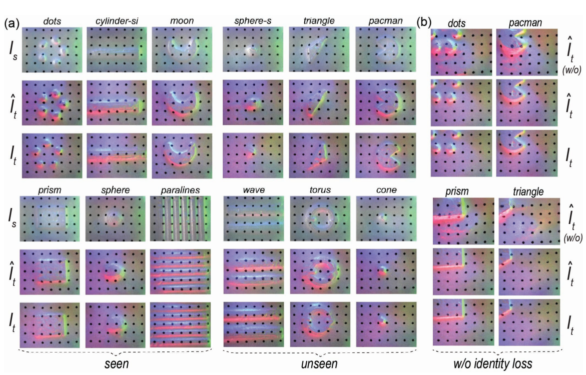

Fig. 3. Visualization of tactile image translation.

B. Force Prediction Result

Results from the *seen* group are presented in Table II. Fig. 4 illustrates the half-violin distributions of force error within four force ranges by using Transforce model with LSTM in *seen* group. When using the source-only method, whether with or without the LSTM module ($\eta^0_{rm}(\mathbf{I}_t)$ and $\eta_{rm}(\mathbf{I}_t)$), the predicted forces exhibit significant errors, as

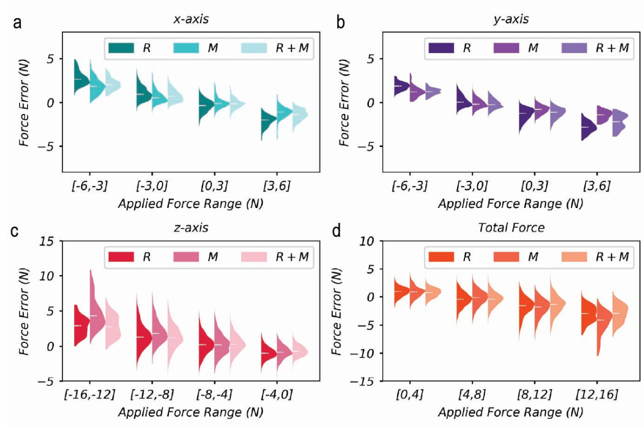

Fig. 4. Force prediction performance.

TABLE II
FORCE PREDICTION PERFORMANCE IN **SEEN** GROUP
NOTE: *error* IS CALCULATED WITH MAE WITHIN FULL FORCE RANGE

Method	Image Type	F_x		F_y		F_z		Total Force
		$error(N)\downarrow$	$R^2\uparrow$	$error(N)\downarrow$	$R^2\uparrow$	$error(N)\downarrow$	$R^2\uparrow$	$error(N)\downarrow$
Source-only ($lstm$)	R	1.528	-0.25	1.14	0.04	3.085	-0.25	2.869
	M	2.811	-2.28	1.267	-0.01	2.521	-0.71	2.771
	$R+M$	2.425	-1.56	1.092	0.19	3.623	0.12	3.606
TransForce ($nolstm$)	R	1.034	0.38	0.953	0.31	1.42	0.68	1.452
	M	**0.731**	**0.67**	**0.754**	**0.62**	1.462	0.64	1.495
	$R+M$	0.891	0.55	0.994	0.3	**1.312**	**0.71**	1.424
TransForce ($lstm$)	R	1.023	0.42	1.014	0.26	1.221	0.76	1.283
	M	**0.695**	**0.69**	**0.701**	**0.66**	1.34	0.68	1.384
	$R+M$	0.771	0.65	0.899	0.43	**1.112**	**0.79**	**1.197**

indicated by negative R^2 values in both normal and shear directions. In contrast, the TransForce model shows marked improvements. For the predicted force accuracy of $\hat{\phi}_{rm}(\mathbf{I}_t)$, the error in shear direction is 0.771 N (6.4% of the full range) in the x-axis and 0.899 N (7.5%) in the y-axis. For the normal direction, the force prediction error is 1.112 N (6.9%). These results validate the efficacy of the model in predicting forces accurately, especially for shear forces, even with expanded force ranges and diverse indenter types.

V. CONCLUSION

In this study, we propose a novel transferable model to address the challenge of unsupervised force prediction for VBTSs. Our approach effectively performs tactile image translation, transforming images from existing sensors to new sensors for leveraging collected force labels. A force prediction model that operates on sequential tactile images demonstrates superior force prediction accuracy over models trained with single image in both normal direction and shear direction. Owing to the generalization performance of generative model, this method is versatile and applicable to various types of VBTSs.

REFERENCES

[1] D. F. Gomes, Z. Lin, and S. Luo, "GelTip: A finger-shaped optical tactile sensor for robotic manipulation," in *IROS*, 2020, pp. 9903–9909.
[2] N. Ou, Z. Chen, and S. Luo, "Marker or markerless? mode-switchable optical tactile sensing for diverse robot tasks," *IEEE Robotics and Automation Letters*, 2024.
[3] Z. Chen, N. Ou, J. Jiang, and S. Luo, "Deep domain adaptation regression for force calibration of optical tactile sensors," *arXiv preprint arXiv:2407.14380*, 2024.