



## King's Research Portal

[Link to publication record in King's Research Portal](#)

### *Citation for published version (APA):*

Chen, Z., Jiang, J., & Luo, S. (in press). Deep Domain Adaptation Regression for Force Calibration of Optical Tactile Sensors. In *The 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2024)*

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Deep Domain Adaptation Regression for Force Calibration of Optical Tactile Sensors

Zhuo Chen<sup>1</sup>, Ni Ou<sup>1,2</sup>, Jiaqi Jiang<sup>1</sup> and Shan Luo<sup>1</sup>

**Abstract**—Optical tactile sensors provide robots with rich force information for robot grasping in unstructured environments. The fast and accurate calibration of three-dimensional contact forces holds significance for new sensors and existing tactile sensors which may have incurred damage or aging. However, the conventional neural-network-based force calibration method necessitates a large volume of force-labeled tactile images to minimize force prediction errors, with the need for accurate Force/Torque measurement tools as well as a time-consuming data collection process. To address this challenge, we propose a novel deep domain-adaptation force calibration method, designed 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 effectiveness of the proposed unsupervised force calibration method, with lowest force prediction errors of 0.102N (3.4% in full force range) for normal force, and 0.095N (6.3%) and 0.062N (4.1%) for shear forces along the x-axis and y-axis, respectively. This study presents a promising, general force calibration methodology for optical tactile sensors.

## I. INTRODUCTION

Optical tactile sensors [1] with high sensitivity in perceiving object geometry, slip and position have now been widely studied. With the superiority of directly capturing tactile information with high-resolution images, optical tactile sensors like GelSight [2], GelTip [3] and Digit [4] significantly enhance the capabilities of robots, particularly when integrated with machine learning models. 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 [5] and in-hand manipulation [6].

Force calibration in optical tactile sensing faces three primary challenges. Firstly, deep neural networks are often preferred over physical models to estimate forces from tactile images, necessitating a substantial volume of tactile images paired with labeled forces for model training [7]. This demanding data collection process has to be repeated for new sensors. As a result, deploying force calibration models, especially for users of commercial tactile sensors

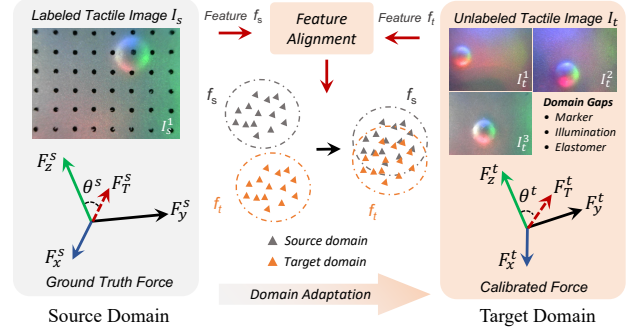


Fig. 1. Deep domain adaptation for force calibration of optical tactile sensors. Upon completing domain adaptation, the feature space of the unlabeled tactile images  $I_t$  from the target domain and the labeled tactile images  $I_s$  from the source domain are aligned. A shared regressor trained with ground truth forces  $F^s$  in source domain can be used to predict  $F^t$  in target domain.  $I_t^1, I_t^2$  and  $I_t^3$  denote tactile images in target domain with varying combinations of domain gaps.  $F_T$  represents the total force,  $F_z$  denotes the normal force,  $F_x$  and  $F_y$  denote the shear forces, and  $\theta$  represents the force angle, respectively.

lacking accurate calibration tools like the Nano17 F/T sensors, becomes challenging. 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.

Transfer learning presents a promising approach for models trained on existing sensors (source domain) to adapt to new sensors (target domain). However, current transfer-learning methods used for this task, such as fine-tuning [8], are supervised, which requires extensive labeled force information from both domains. Hence, there remains a significant demand for a dedicated model utilizing unsupervised transfer learning methods in the force calibration of optical tactile sensors.

In this study, we propose a novel domain adaptation regression model to address the unsupervised force calibration challenge in optical tactile sensors, illustrated in Fig. 1. This approach can eliminate the need for costly force/torque measurement tools in force calibration of optical tactile sensors and significantly reduce the calibration time through domain adaptation. Experimental results demonstrate the successful adaptation of pretrained models on force-labeled tactile images to sensors with diverse domain gaps and unlabeled images. To our best knowledge, this is the

\*This work was supported by the EPSRC project “ViTac: Visual-Tactile Synergy for Handling Flexible Materials” (EP/T033517/2).

<sup>1</sup>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.

<sup>2</sup>Ni Ou is with the State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing, 100081, China.

first work that utilizes the deep domain adaptation regression method to address this challenge. The contributions of this work are summarized as follows:

- 1) A novel domain adaptation regression method is introduced to address the challenge of unsupervised force calibration in optical tactile sensors;
- 2) The impact of different combinations of three key domain gaps on domain adaptation performance is investigated, and the effectiveness of our model is verified;
- 3) A dataset has been made publicly accessible for unsupervised force calibration of optical tactile sensors.

The rest of the paper is structured as follows: Section II provides an overview of related works; Section III introduces our methodology; Section IV details our data collection and implementation; Section V analyses the experimental results. Finally, Section VI presents the discussion and summarises the work.

## II. RELATED WORK

### A. Optical Tactile Sensors

Optical tactile sensors are characterized by utilizing compact cameras to capture high-resolution images of deformed soft elastomers when contacting with objects. By exploiting image processing techniques or machine learning models, tactile images can be used for various downstream robotic manipulation tasks, including force estimation [7], slip detection [9], object recognition [10], and object localization [11]. Optical tactile sensors can be broadly categorized into marker-motion-based sensors like TacTip [3] and high-resolution image-based sensors like GeSight [1]. TacTip-like sensors consist of physical pins on the sensing elastomer and primarily detect shear information and contact position by capturing marker displacement. In contrast, GeSight-like sensors utilize RGB cameras to capture images with detailed textures and reconstruct contact height maps using photometric stereo methods. The black marker dots can also be painted on the gel elastomer for shear displacement estimation [12]. Over the past decades, GeSight-like sensors have been developed into different variants, such as GelSlim [13], GelWedge [14] and GelSvelte [15]. In this work, we primarily focus on force calibration using GeSight sensors, but it's worth noting that this method is also applicable to TacTip-like sensors and other types of optical tactile sensors.

### B. Force Calibration of Optical Tactile Sensors

Force calibration of optical tactile sensors involves establishing the mapping relationship between tactile images and contact forces [1]. This process can be achieved either by applying a physical-based model/mechanical calibration method to compute contact force, or by directly constructing an end-to-end image-to-force mapping using a deep neural network [7]. However, due to the hyperelastic properties of gel elastomer membranes, the mechanical calibration

method, which assumes linear elastic behavior, is only accurate for small deformations, and its force prediction performance significantly deteriorates with large deformations [16]. In contrast, data-driven deep neural networks exhibit impressive performance in non-linear regression tasks and can maintain accurate force prediction across the full working range of deformation [7]. Nevertheless, the neural network-based approach is data-driven and sensitive to the feature distribution of tactile images, necessitating data collection and model re-training whenever the sensor's physical factors change. There is a pressing need for a method that simplifies the cumbersome data collection process and expedites the calibration process on new or aging sensors by adapting trained models on existing sensors.

### C. Domain Adaptation Regression

Domain adaptation [17] aims at mitigating the distribution shifts between the labeled source domain and the unlabeled target domain. Adversarial domain adaptation techniques, such as DANN [18] and Deep-CORAL [19], are widely used to learn an embedding feature space where the source and target domains cannot be distinguished. While instance-based methods primarily correct the shift by re-weighting source instances or minimizing specific distances between the two distributions, such as KL-divergence [20] and Maximum Mean Discrepancy (MMD) [21]. However, these methods are typically applied to classification [22] and segmentation [23] tasks and are not directly applicable to regression problems. While recent methods like RSD [24] and DARE-GRAM [25] have been proposed to tackle domain adaptation regression problems, they have only been tested on a few benchmark datasets with simple domain gaps. In contrast, our domain adaptation regression task for force calibration will consider different combinations of domain gaps in optical tactile sensors, which is more complex and problem-specific.

## III. METHODOLOGY

### A. Problem Definition

In the domain adaptation regression problem for the force calibration of optical tactile sensors, we are provided with a labeled tactile image dataset  $\mathcal{D}_s = \{(\mathbf{I}_s^i, (\mathbf{F}_s^i, \mathbf{C}_s^i))\}_{i=1}^{n_s}$  including the source domain with  $n_s$  samples, and an unlabeled dataset  $\mathcal{D}_t = \{(\mathbf{I}_t^i)\}_{i=1}^{n_t}$  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, respectively.  $\mathbf{F}_s^i = (F_x^i, F_y^i, F_z^i)_s$  denotes the ground truth of the applied normal force vector  $F_z^i$  and shear forces vector  $F_x^i$  and  $F_y^i$  in source domain, while  $\mathbf{C}_s^i$  indicates the contact class of the tactile images in the source domain, as described in Section IV-A.

The primary challenge in this task is that  $\mathcal{D}_s$  and  $\mathcal{D}_t$  are sampled from different sensors or the same sensor with varying marker presence, illumination conditions, or elastomer modulus. Consequently, tactile images collected from sensors with different combinations of domain gaps exhibit distinct feature distributions, 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

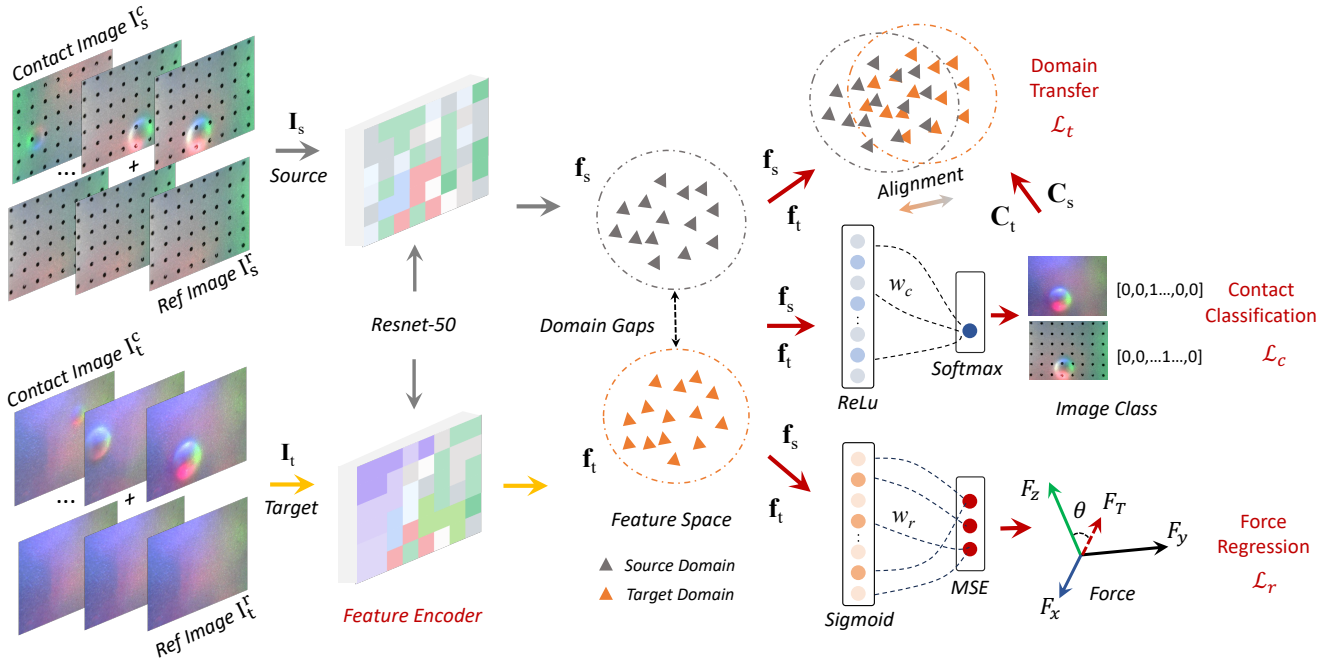


Fig. 2. Deep domain adaptation regression model for force calibration. The model comprises four components: a feature encoder, a contact classification head, a force regression head, and a domain transfer head. The model takes batches of tactile images from the source domain  $\mathbf{I}_s$  and the target domain  $\mathbf{I}_t$  as inputs. Both  $\mathbf{I}_s$  and  $\mathbf{I}_t$  consist of concatenated tactile images comprising a current contact image  $\mathbf{I}^c$  and a reference image  $\mathbf{I}^r$ . The domain transfer head accepts features ( $\mathbf{f}_s$  and  $\mathbf{f}_t$ ) and ground truth contact labels  $\mathbf{C}_s$  from the source domain, as well as the pseudo labels  $\mathbf{C}_t$  predicted from the classifier. The overall loss  $\mathcal{L}_h$  is weighted sum of the regression loss  $\mathcal{L}_r$ , classification loss  $\mathcal{L}_c$  and domain adaptation loss  $\mathcal{L}_t$ .

calibrated forces vector  $\mathbf{F}_t^i$  in the target domain. Therefore, our goal is to minimize the mean force prediction error in the target domain:

$$\arg \min_h \mathbb{E}_{(\mathbf{I}_t^i, \mathbf{F}_t^i)} \|\mathbf{h}(\mathbf{I}_t^i), \mathbf{F}_t^i\|_2^2 \quad (1)$$

However, since we only possess labeled images in the source domain, this optimization problem described in Equation 1 is reformulated to minimize the Mean Square Error (MSE) between the predicted force and ground truth forces on the labeled source samples. Additionally, we aim to minimize the domain distribution discrepancy between the source domain and target domain:

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

where  $\lambda_r \geq 0$ ,  $\lambda_t \geq 0$  represent the weights assigned to the regression loss  $\mathcal{L}_r$  and domain transfer loss  $\mathcal{L}_t$ .  $\hat{\mathbf{F}}_s^i$  denotes the predicted forces using source images, while  $\mathbf{f}_s$  and  $\mathbf{f}_t$  denote the features extracted from the source domain and target domain, respectively.  $\hat{d}_{\mathcal{H}}(\cdot, \cdot)$  is the estimation of Local Maximum Mean Discrepancy (LMMD) [22], with  $\mathcal{H}$  being the Reproducing Kernel Hilbert Space (RKHS). This equation aims to learn a shared feature space and a shared regressor simultaneously so that forces in images from both the source or target domains can be predicted.

To compute the domain transfer loss  $\mathcal{L}_t$ , we refer to the Deep Subdomain Adaptation Network (DSAN) [22], which has demonstrated the capability to capture fine-grained information for categories and align relevant domain distributions to learn a shared classifier. Hence, we try to reduce the discrepancy between source domain and target domain by optimizing the following function:

$$\arg \min_g \lambda_c \underbrace{\frac{1}{n_s} \sum_{i=1}^{n_s} J(g(\mathbf{I}_s^i), \mathbf{C}_s^i)}_{\mathcal{L}_c} + \lambda_t \hat{d}_{\mathcal{H}}(\mathbf{f}_s, \mathbf{C}_s, \mathbf{f}_t, \mathbf{C}_t) \quad (3)$$

where  $\lambda_c \geq 0$  represents the weight of classification loss  $\mathcal{L}_c$ ,  $g(\cdot)$  is a classifier for tactile images with different contact class, and  $J(\cdot, \cdot)$  denotes the cross-entropy error. Compared to other domain adaptation regression methods that use a single regression head, the classification head included in this equation not only provides essential pseudo contact class labels  $\mathbf{C}_t$  for the unlabeled images to calculate the LMMD but also benefits the regression task, as shown in [26].

Then, by combining Equations 2 and 3, we obtain the overall loss function  $\mathcal{L}_h$  of our model:

$$\mathcal{L}_h = \lambda_r \mathcal{L}_r + \lambda_c \mathcal{L}_c + \lambda_t \mathcal{L}_t \quad (4)$$

By optimizing the loss function in Equation 4, a shared regressor can be learned for both the source domain and target domain while maintaining a well-aligned feature space.



## B. Deep Domain Adaptation Regression Network

Based on Section III-A, the deep domain adaptation regression network in this task can be divided into four components as depicted in Fig. 2: (1) feature encoder, (2) contact classification head, (3) force regression head and (4) domain transfer head. Firstly, the inputs from each domain consist of the concatenation of tactile images with contact  $I^c$  and reference tactile images without contact  $I^r$ , which are then fed into the feature encoder. The feature encoder with a ResNet-50 backbone and a linear bottleneck layer jointly extracts the deep representations  $f_s$  and  $f_t$  of tactile images from the source domain  $I_s$  and the target domain  $I_t$ , respectively. Next, the contact classification head composed of a single linear layer with a ReLU activation function classifies the contact cases of tactile images in the target domain and provides pseudo labels  $C_t$  for target images, which are then used for LMMD calculation with the ground truth labels  $C_s$  of source images. The force regression head using a linear layer with a Sigmoid activation function minimizes the MSE between the predicted forces  $\hat{F}_s$  and the ground truth forces  $F_s$ . The domain transfer head aligns two domains in the feature space by minimizing LMMD.

By using this model, tactile images  $I_s$ , contact forces  $F_s$  and contact class  $C_s$  from existing sensors can be leveraged as source domain, while adapting existing models to new sensors or aging sensors by just collecting unlabeled tactile images  $I_t$  as target domain. This process could eliminate the need for force measurement tools, such as F/T sensors, during the calibration stage. Furthermore, the new model can be trained with fewer epochs by adapting existing models, thereby reducing the time required compared to the traditional supervised force calibration process.

## IV. DATA COLLECTION AND IMPLEMENTATION

As reported in Section I, 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, as shown Table I. For data collection, three different elastomers and illumination conditions indexed by 0, 1, 2 are used. The elastomer  $b$  indexed with higher number is with higher hardness. To study the domain adaptation performance between sensors with different combinations of these variables, we obtain four types of labeled tactile images, i.e.,  $mb_0i_0$ ,  $wmb_0i_0$ ,  $wmb_1i_1$  and  $wmb_2i_2$ , and pair them into nine domain adaptation groups listed in Table I. For example, in the case of  $mb_0i_0 \rightarrow wmb_0i_0$ , tactile images  $I_s$  are with markers and collected with elastomer-0 and illumination-0, while  $I_t$  are without markers and collected with elastomer-1 and illumination-1. It is worth noting that, for data collection, we only need to collect  $mb_0i_0$ ,  $wmb_1i_1$  and  $wmb_2i_2$  in real GelSight sensors with corresponding physical properties. While  $wmb_0i_0$  is generated by applying an inpainting method [27] on the  $mb_0i_0$ , which ensures only the marker-presence gap exists between the two domains.

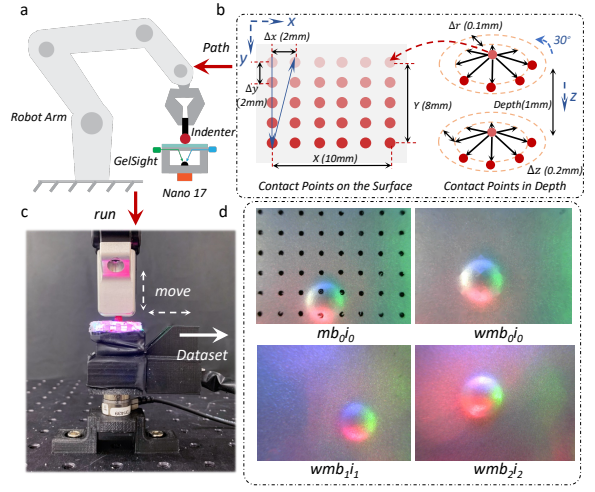


Fig. 3. (a) Data collection setup with a robot arm, a GelSight sensor, a sphere indenter and a Nano17 F/T sensor. (b) Programmed contact path for data collection. Contact pixel denotes a point for indentation. (c) Real-world setup. (d) Four groups of collected tactile images with different combinations of domain gaps.

TABLE I  
DOMAIN ADAPTATION GROUP

#Variable	Domain Adaptation Group
one	$mb_0i_0 \rightarrow wmb_0i_0$ , $wmb_0i_0 \rightarrow mb_0i_0$
two	$wmb_0i_0 \rightarrow wmb_1i_1$ , $wmb_0i_0 \rightarrow wmb_2i_2$ $wmb_1i_1 \rightarrow wmb_0i_0$ , $wmb_1i_1 \rightarrow wmb_2i_2$
three	$mb_0i_0 \rightarrow wmb_1i_1$ , $mb_0i_0 \rightarrow wmb_2i_2$ $wmb_1i_1 \rightarrow mb_0i_0$

\*  $m/wm$ : with/without markers

\*  $b_0, b_1, b_2$ : elastomer - 0, 1, 2

\*  $i_0, i_1, i_2$ : illumination condition - 0, 1, 2

### A. Data Collection

As shown in Fig. 3a, 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. The robot arm is programmed through MoveIt to follow a planned path, as depicted in Fig. 3b. The camera inside the GelSight sensor is synchronized with the Nano17 sensor to capture tactile images and contact forces simultaneously.

Regarding the indenter's contact,  $6 \times 5$  surface points are pre-determined in GelSight's  $10 \times 8$  mm<sup>2</sup> surface. For each contact point, the indenter is pressed into different depths ranging from 0 to 1 mm with a step of 0.2 mm. To collect shear forces, the motion of the robotic arm is divided into two stages: moving downwards and then moving horizontally shown in Fig. 3c. When moving at a specific depth, the indenter follows the motion rules in Algorithm 1. Specifically, it moves in a cyclical radial motion around a fixed radius towards 12 different angles. After the 12 angles of motions finished, the radius of the circle increases with a step of 0.1 mm from 0.1 mm to 0.6 mm, and the next cycle begins. There are  $k = 30$  contact points on the surface, and  $n = 361$  contact points in depth (including one reference

position before contact). 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. This is limited by the maximum thickness of gel elastomer (3 mm) and the non-slip displacement ( $<0.6$  mm) between the sphere indenter and the elastomer surface. As described in Section III-A, we introduce contact class  $\mathbf{C}_s^i$  in the loss function. Here, we allocate the contact class for each tactile image based on the index of the contact points in depth, i.e.,  $\mathbf{C}_s^i \in \mathbb{R}^{361}$  is a one-hot vector representing the label of  $\mathbf{I}_s^i$ , where  $C_s^{ij} = 1$  denotes  $\mathbf{I}_s^i$  labeled with class  $j$ . The contact class is only used in training stage, which will not hinder the prediction of continuous force values in test.

---

**Algorithm 1** Indenter’s Contact Motion

---

```

1: Input: Number of surface contact points  $k$ , surface contact points  $P = \{P_1, \dots, P_k\}$ , where  $P_i = \{p_1, \dots, p_n\}$  contains all contact points in depth,  $p_i = \{x, y, z\}$  denotes the coordinate of a contact point.
2: procedure MOVE TO CONTACT POINTS( $P, k$ )
3:   for  $i = 0 \dots k$  do
4:      $p_0 \leftarrow P_i[0]$   $\triangleright$  origin point  $p_0$  without contact
5:      $\text{moveTo}(p_0)$ 
6:     for  $j = 1 \dots n$  do
7:        $p_1 \leftarrow (p_0[0], p_0[1], P_{ij}[2])$ ,  $p_2 \leftarrow P_{ij}$ 
8:        $\text{moveTo}(p_1)$   $\triangleright$  Move downwards
9:        $\text{moveTo}(p_2)$   $\triangleright$  Move horizontally
10:       $\text{imageCapture}()$ ,  $\text{forceRecord}()$ 
11:       $\text{moveTo}(p_0)$   $\triangleright$  Move to origin
12:   end for
13: end for
14: end procedure

```

---

Fig. 3d illustrates the collected tactile images within four groups. It is evident that  $mb_0i_0$ ,  $wmb_1i_1$ , and  $wmb_2i_2$  exhibit distinctive illumination conditions, also corresponding to three real GelSight sensors with different elastomer modulus. The data group  $wmb_0i_0$  is directly generated from  $mb_0i_0$  via the inpainting method, which controls the variables of illumination and elastomer, and shares the same ground truth force with  $mb_0i_0$ . Notably, we have three groups ( $mb_0i_0$ ,  $wmb_0i_0$ ,  $wmb_1i_1$ ) that comprise a total of  $3 \times 10,830$  images collected by the path in Fig. 3b, along with 873 images of  $wmb_2i_2$ . The last group of  $wmb_2i_2$ , containing only 873 images, is collected along a sparser path to test the performance of our model when trained with a few unlabeled images as target domain, while the other three data groups with ten times more images are utilized to evaluate the model performance across different domains with the same data size. These four data groups hold promise for being leveraged as source domains to extend our domain adaptation model to new sensors by simply collecting unlabeled tactile images without using force measurement tools.

### B. Implementation & Evaluation Metrics

In the training stage, three models are first trained with data groups  $mb_0i_0$ ,  $wmb_0i_0$ , and  $wmb_1i_1$ , respectively,

where only the regression layer is used for source-domain supervision ( $\lambda_r = 1, \lambda_c = 0, \lambda_t = 0$ ), and an initial learning rate  $\eta_0 = 0.1$  for 20 epochs. This step aims at offering existing models for domain adaptation. In real-world use, we usually have an existing sensor with the trained model, and then the model and tactile images are used as source domain in this task for adapting to new/old sensors in target domain. For the second training stage, the corresponding pretrained model of source domain is selected and adapted by adding transfer loss and classification loss, i.e.,  $\lambda_r = 1, \lambda_c = 1, \lambda_t = 1$  with an initial learning rate  $\eta_0 = 0.01$  for 10 epochs. The learning rate in the backbone layer is always set as 10 times smaller than the initial learning rate. SGD optimizer is used with a momentum of 0.9, along with a learning rate scheduler  $\eta = \eta_0 \cdot (1 + 0.0003 \cdot i)^{-0.75}$ , where  $i$  is the number of iterations. The batch size is set as 32, and all experiments are trained on a NVIDIA RTX 3090 GPU.

All source-domain data are used in training stage, while the target-domain tactile images are split into train, valid and test dataset with a ratio of 0.6:0.2:0.2. The ground truth forces  $\mathbf{F}_s$  from the source domain are normalized using min-max normalization to the range of  $[0, 1]$  in the training stage. For testing, the model exclusively receives unlabeled tactile image batches from the target domain. The calibrated force  $\mathbf{F}_t$  is then derived from the regression head and subsequently subject to de-normalization for error calculation. The evaluation metrics used in Section V are the Mean Absolute Error (MAE) and the coefficient of determination  $R^2$ , while tSNE is used to visualize domain distances in the feature space.

## V. EXPERIMENT RESULTS & ANALYSIS

In this section, we demonstrate the domain adaptation performance of our model in force calibration of GelSight sensors across different domain adaptation groups as shown in Table I. 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. Two other representative domain adaptation methods, including DANN [18] and GRAM [25], are also compared with our method in subsections V-A, V-B and V-C. The force prediction errors in Tables II, III, and IV are calculated by the average of MAE with shear forces and normal forces.

### A. Marker Presence

The GelSight sensors with markers offer significant advantages in slip detection and force visualization, while the sensors without markers show advantages in object classification. Given the wide use of both types of sensors, it is desired to directly transfer the trained force prediction model from one of those to its counterpart. However, if we employ the source-only method directly from  $mb_0i_0$  to  $wmb_0i_0$ , as depicted in Fig. 4a-i and Fig. 4b-i, 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.

Conversely, in the transition from  $wmb_0i_0$  to  $mb_0i_0$ , the  $R^2$  values are -0.13, 0.05, and 0.13 in the  $xyz$ -axis, and the MAE errors are high. These results confirm the existence of significant gaps of marker presence, leading to substantial errors of force prediction across different domains.

After applying our method, as depicted in Fig. 4a-ii and Fig. 4b-ii, the errors  $E$  of normal force in both groups decrease to 0.102 N (3.4%), while the  $R^2$  values increase to 0.92 and 0.91 respectively. 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. One possible reason for the low  $R^2$  values in the shear force could be that the collected shear forces are too small, mostly lying in the range of -0.75 N to 0.75 N, compared with the normal forces in -3 N to 0 N. Therefore, the static tactile images with small deformations contain negligible information for the shear forces, resulting in poorer transfer performance than normal forces. Regarding the comparison results with DANN and GRAM, our method shows an advantage in the overall average error (0.086 N), as shown in Table II.

TABLE II  
AVERAGE FORCE PREDICTION ERROR WITH **ONE** VARIABLE  
(MARKER, UNIT N)

Method	$mb_0i_0 \rightarrow wmb_0i_0$	$wmb_0i_0 \rightarrow mb_0i_0$	Avg
source-only	0.276	0.196	0.236
DANN	0.098	0.118	0.108
GRAM	<b>0.082</b>	0.091	0.087
ours	0.086	<b>0.086</b>	<b>0.086</b>

### B. Illumination & Elastomer

In this subsection, we study the influence of two domain variables, i.e., illumination condition and elastomer modulus on four domain adaptation groups listed in Table I. As shown in Table III, 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 in all groups, with notable improvements observed in the groups  $wmb_1i_1 \rightarrow wmb_0i_0$  and  $wmb_0i_0 \rightarrow wmb_1i_1$ , where the errors decrease from 0.226 N to 0.095 N and from 0.3 N to 0.138 N respectively. The total average error of 0.145 N outperforms other models.

It is noteworthy that when  $wmb_1i_1$  is used as the source domain, the force prediction errors in the source-only method are smaller compared to the other two groups. This is because the sensor with elastomer  $b_1$  is harder than the counterpart with elastomer  $b_0$ , resulting in larger contact forces with the same contact depth. Hence, more feature changes induced by larger normal forces, such as elastomer's deformation and light intensity, contains in the source domain, leading to

better adaptation performance. Additionally, when  $wmb_2i_2$  serves as the target domain, the transfer performance is the poorest, shown in Table III and Fig. 4, with the  $R^2$  values for the normal force in  $wmb_1i_1 \rightarrow wmb_2i_2$  improving from 0.47 to 0.67, compared to the improvement from 0.46 to 0.93 in  $wmb_1i_1 \rightarrow wmb_0i_0$ . This discrepancy can be attributed to the data size, as  $wmb_2i_2$  has ten times fewer images than the other three groups, as mentioned in Section IV-A. This indicates that the size of both target and source data also influences domain adaptation performance.

TABLE III  
AVERAGE FORCE PREDICTION ERROR WITH **TWO** VARIABLES  
(ELASTOMER & ILLUMINATION, UNIT N)

Methods	$wmb_1i_1 \rightarrow wmb_2i_2$	$wmb_1i_1 \rightarrow wmb_0i_0$	$wmb_0i_0 \rightarrow wmb_1i_1$	$wmb_0i_0 \rightarrow wmb_2i_2$	Avg
source-only	0.230	0.226	0.300	0.384	0.285
DANN	0.180	0.120	<b>0.132</b>	<b>0.162</b>	0.149
GRAM	0.242	0.102	0.143	0.186	0.168
ours	<b>0.167</b>	<b>0.095</b>	0.138	0.180	<b>0.145</b>

On the other hand, although the  $R^2$  values for shear forces still remain relatively low, the improvement over the baseline source-only method is significant. In the transition from  $wmb_1i_1$  to  $wmb_0i_0$ , the shear force errors decrease from 0.168 N (11.2%) to 0.095 N (6.3%) in  $F_x$  and from 0.195 N (13%) to 0.093 N (6.2%) in  $F_y$ . Similarly, in the transition from  $wmb_1i_1$  to  $wmb_2i_2$ , the shear force errors decrease from 0.148 N (9.8%) to 0.129 N (8.6%) in  $F_x$  and from 0.180 N (12.0%) to 0.136 N (9.0%) in  $F_y$ . The improvement in normal force is also notable, decreasing from 0.315 N (10.5%) to 0.097 N (3.2%) in  $wmb_1i_1 \rightarrow wmb_0i_0$ , and from 0.361 N (12.0%) to 0.236 N (7.9%) in  $wmb_1i_1 \rightarrow wmb_2i_2$ .

### C. Marker Presence & Illumination & Elastomer

We finally combine all three domain variables and study three adaptation groups listed in Table I. As shown in Table IV, it is evident that with the increase of domain variables, the average force prediction error rises, from 0.236 N (one variable, in Table II) to 0.285 N (two variables, in Table III) to 0.332 N (three variables) when using the source-only method. This trend implies the increased complexity of domain adaptation performance with more domain gaps. Despite this, our method also outperforms the other two methods and reduces the average error from 0.332 N to 0.153 N. Fig. 4 demonstrates the transfer performances in the three axis. We observe that, even when combined with three variables, the  $R^2$  value in normal force can be improved from -0.39 to 0.89 in  $wmb_1i_1 \rightarrow mb_0i_0$ . Although the  $R^2$  value in  $mb_0i_0 \rightarrow wmb_2i_2$  only shows changes from -0.59 to 0.47, the force error decreases from 0.637 N (21.2%) to 0.299 N (9.9%). 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.

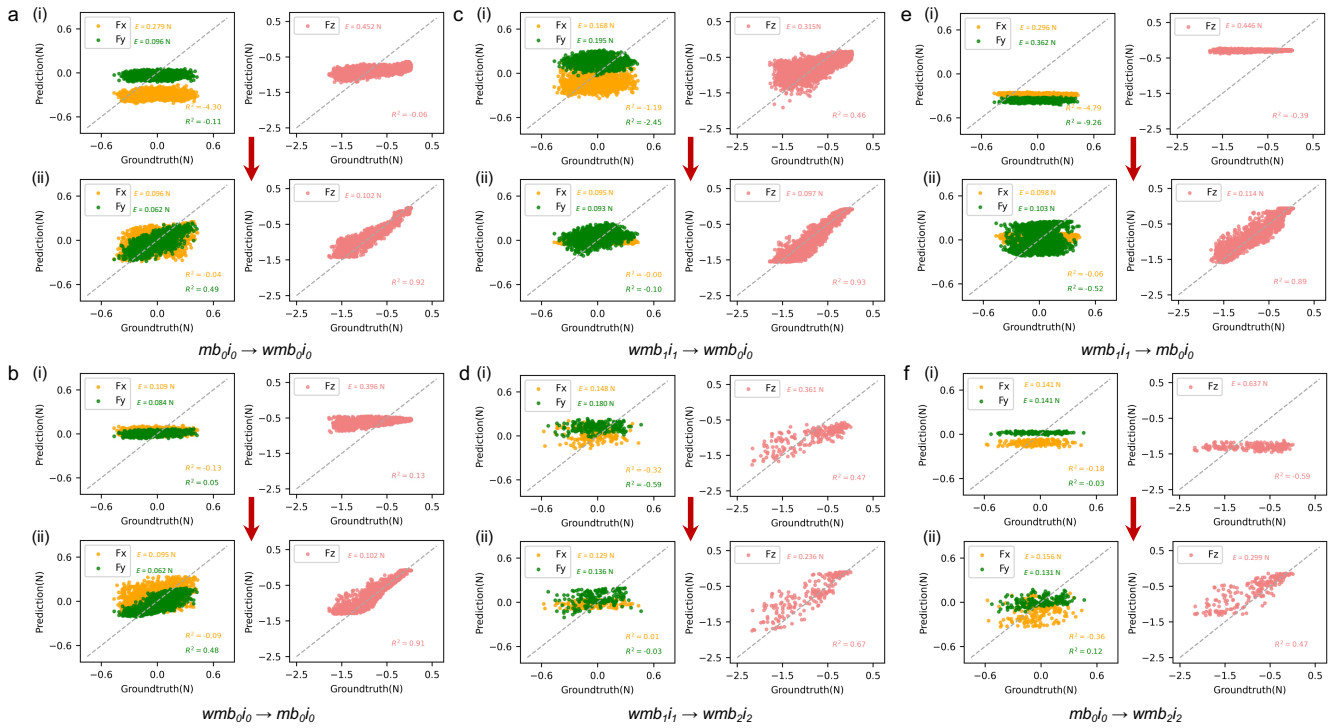


Fig. 4. Force prediction errors compared among domain adaptation groups featuring one variable (a-b), two variables (c-d), and three variables (e-f). This comparison is conducted using both the source-only method (i) and our domain adaptation method (ii).

TABLE IV  
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
DANN	<b>0.143</b>	0.238	0.185	0.189
GRAM	0.159	0.198	0.129	0.162
ours	0.160	<b>0.195</b>	<b>0.105</b>	<b>0.153</b>

#### D. Discussion

The experimental results in Sections V-A, V-B, and V-C have validated the feasibility of domain adaptation regression methods in the force calibration of optical tactile sensors. Particularly noteworthy is our method’s highest error percentage improvement in group  $wmb_0i_0 \rightarrow mb_0i_0$ , with improvements of 13.2% and 17.2% in shear forces, and 11.1% in normal force. Additionally, in the  $wmb_0i_0 \rightarrow mb_0i_0$  transition, the prediction error can be reduced to lowest among nine groups as 0.102N (3.4%) in normal force, and 0.095N (6.3%) and 0.062N (4.1%) in shear forces. Fig. 5 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. As such, considering the real-world deployment

of force prediction in robotic manipulation, the prediction accuracy achieved by domain adaptation methods, especially for normal forces, is accurate enough for force feedback control by detecting the threshold of contact forces.

However, our method still exhibits inferior performance in shear force calibration compared with the supervised method. This phenomenon could be attributed to the use of the sphere indenter with low friction coefficient, resulting in slip occurring over a very small moving distance ( $<0.5$  mm) thus cannot apply large shear forces on the elastomer. This determines the collected tactile images containing negligible shear features, as they are applied with small lateral forces. Furthermore, the method primarily focuses on accepting static contact images with reference images, which inherently contain poor information about dynamic shear forces. Promising methods to overcome these drawbacks may involve introducing sequential tactile images to provide temporal information about shear displacements.

#### VI. 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/touch 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



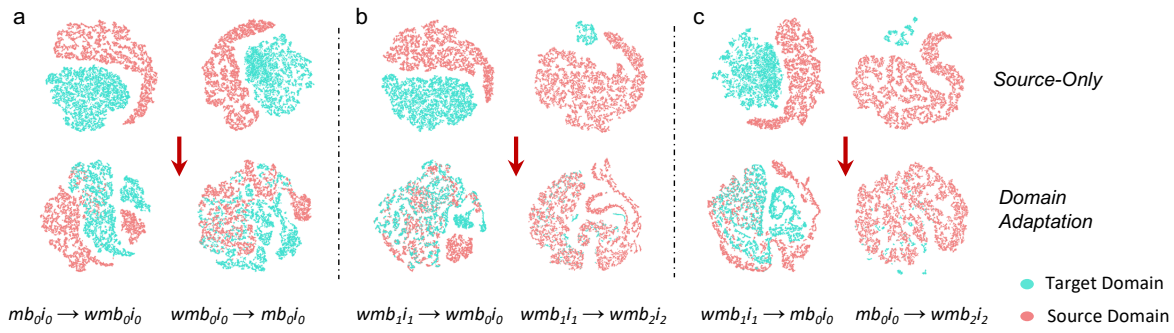


Fig. 5. Feature spaces visualized using tSNE from the source-only method and our domain adaptation method in (a) one domain variable, (b) two domain variables, (c) and three domain variables.

but also holds promise for enhancing the performance of various optical tactile sensors and tactile sensors based on different sensing principles.

## 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] R. Li, R. Platt, W. Yuan, A. Ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using GelSight tactile sensing," in *IROS*, 2014, pp. 3988–3993.
- [3] D. F. Gomes, Z. Lin, and S. Luo, "GelTip: A finger-shaped optical tactile sensor for robotic manipulation," in *IROS*, 2020, pp. 9903–9909.
- [4] M. Lambeta, P.-W. Chou, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *RA-L*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [5] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, and Y. Shen, "Soft magnetic skin for super-resolution tactile sensing with force self-decoupling," *Science Robotics*, vol. 6, no. 51, p. eabc8801, 2021.
- [6] Z. Lu and H. Yu, "Gtac-hand: A robotic hand with integrated tactile sensing and extrinsic contact sensing capabilities," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 5, pp. 2919–2929, 2023.
- [7] H. Sun, K. J. Kuchenbecker, and G. Martius, "A soft thumb-sized vision-based sensor with accurate all-round force perception," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 135–145, 2022.
- [8] C. Sferrazza and R. D'Andrea, "Transfer learning for vision-based tactile sensing," in *IROS*, 2019, pp. 7961–7967.
- [9] W. Yuan, E. Adelson *et al.*, "Measurement of shear and slip with a GelSight tactile sensor," in *ICRA*, 2015, pp. 304–311.
- [10] G. Cao, J. Jiang, D. Bollegala, and S. Luo, "Learn from incomplete tactile data: Tactile representation learning with masked autoencoders," in *IROS*, 2023, pp. 10 800–10 805.
- [11] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik, "General in-hand object rotation with vision and touch," in *Conference on Robot Learning*, 2023, pp. 2549–2564.
- [12] W. Kim, J. Kim *et al.*, "Uvtac: Switchable uv marker-based tactile sensing finger for effective force estimation and object localization," *RA-L*, vol. 7, no. 3, pp. 6036–6043, 2022.
- [13] I. H. Taylor, S. Dong, and A. Rodriguez, "GelSlim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *ICRA*, 2022, pp. 10 781–10 787.
- [14] S. Wang, Y. She, B. Romero, and E. Adelson, "GelSight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *ICRA*, 2021, pp. 6468–6475.
- [15] J. Zhao and E. H. Adelson, "GelSight Svelte: A human finger-shaped single-camera tactile robot finger with large sensing coverage and proprioceptive sensing," in *IROS*, 2023, pp. 8979–8984.
- [16] S. K. Melly, L. Liu, Y. Liu, and J. Leng, "A review on material models for isotropic hyperelasticity," *International Journal of Mechanical System Dynamics*, vol. 1, no. 1, pp. 71–88, 2021.
- [17] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
- [18] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *ICML*, 2015, pp. 1180–1189.
- [19] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *ECCV Workshop*, 2016, pp. 443–450.
- [20] M. Sugiyama, S. Nakajima *et al.*, "Direct importance estimation with model selection and its application to covariate shift adaptation," *Advances in neural information processing systems*, vol. 20, 2007.
- [21] J. Huang, A. Gretton, K. Borgwardt, B. Schölkopf, and A. Smola, "Correcting sample selection bias by unlabeled data," *Advances in neural information processing systems*, vol. 19, 2006.
- [22] 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.
- [23] Y. Liu, W. Zhang, and J. Wang, "Source-free domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1215–1224.
- [24] X. Chen, S. Wang *et al.*, "Representation subspace distance for domain adaptation regression," in *ICML*, 2021, pp. 1749–1759.
- [25] I. Nejjar, Q. Wang, and O. Fink, "Dare-gram: Unsupervised domain adaptation regression by aligning inverse gram matrices," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11 744–11 754.
- [26] S. L. Pintea, Y. Lin, J. Dijkstra, and J. C. van Gemert, "A step towards understanding why classification helps regression," in *ICCV*, 2023, pp. 19 972–19 981.
- [27] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004.