Tactile Sensing for Stable Object Placing

Luca Lach[*] IRI, CSIC-UPC

Niklas Funk[*] TU Darmstadt

Robert Haschke Bielefeld University

Helge Ritter Bielefeld University

Jan Peters TU Darmstadt

Georgia Chalvatzaki TU Darmstadt

Abstract

Placing objects on flat surfaces is a crucial skill to master for robots in household environments. Common object-placing approaches require either complete scene specifications or (extrinsic) vision systems, which occasionally suffer from occlusions. Rather than relying on indirect measurements, we propose a novel approach for stable object placing that leverages tactile feedback from an object grasp. We devise a neural architecture called PlaceNet that estimates a rotation matrix, resulting in a corrective gripper movement that aligns the object with the placing surface for the subsequent object manipulation. Our evaluation compares different sensing modalities to each other and PlaceNet to classical, non-learning approaches to assess whether a data-driven approach is indeed required. Applying PlaceNet to a set of unseen everyday objects reveals significant generalization of our proposed pipeline, suggesting that tactile sensing plays a vital role in the intrinsic understanding of robotic dexterous object manipulation.

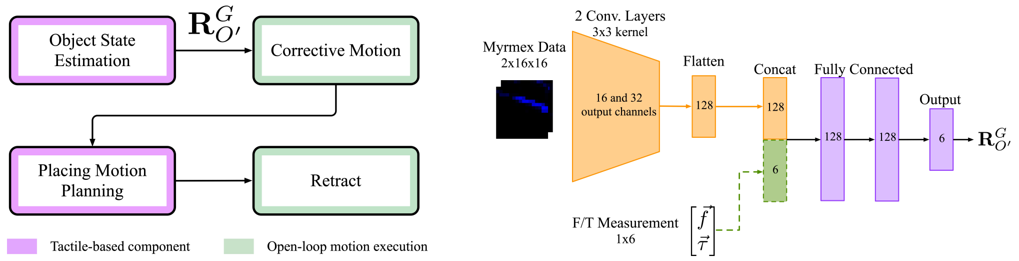
1 Introduction

This work studies the benefit of local tactile measurements between gripper and object for stable and reliable object placing. Stable object placement is an essential skill for any autonomous robotic system, particularly for capable assistive household robots. It forms the basis for many tasks, such as object rearrangement, assembly, sorting, and storing goods. While a large body of prior works exists on stable object placing (Jiang et al. [2012], Harada et al. [2012], Ma et al. [2018], Mitash et al. [2020], Manuelli et al. [2019], Haustein et al., Newbury et al. [2021]), none of those works investigate the contribution of tactile feedback in stable placing. Rather, they rely either on vision systems, which are prone to occlusions and require external sensors, or accurate scene descriptions, which demand cumbersome manual labor. We attempt to fill this gap by investigating the impact of tactile sensing in this simple yet challenging scenario. In this work, we utilize the TIAGo robot with a parallel-jaw gripper and Myrmex sensors (Schürmann et al. [2011]) as fingers.

Our method comprises a deep convolutional neural network called PlaceNet that predicts a corrective rotation action for the gripper from the Myrmex readings, which is then executed to align the object with the placing surface correctly. In an extensive evaluation, we compare tactile to F/T sensor readings and the learning-based approach to two classical baseline models. The main contributions are twofold; (i) the development and training of tactile-based policies for stable object placing without requiring any extrinsic visual feedback, and (ii) an open-source suite of our dataset, CADs, pretrained models, and the codebase of all methods (both classical and deep learning ones) from our extensive real-robot experiments. Overall, our study confirms that tactile sensing can be a powerful and valuable low-cost addition to robotic manipulators: their signals provide features that increase reliability and robot dexterity.

[*] Equal Contribution. Code available on `https://sites.google.com/view/placing-by-touching`

NeurIPS 2023 Workshop on Touch Processing: a new Sensing Modality for AI .



(a) Schematic overview of the four phases of stable object placing. Components relying on tactile data are highlighted, as well as those utilizing motion planning and execution.

(b) Overview of PlaceNet, our neural architecture. Tactile measurements from both sides of the gripper are stacked along the channel dimension of the convolutional layers.

Figure 1: Schematic overview of our placing methods (left) with a detailed depiction of we estimate $\mathbf{R}^G_{O'}$ using a neural network

2 Related Work

Object placing. Stable object placing is a crucial skill for autonomous robotic systems. Many works have therefore focused on modeling or learning it, although most rely on vision as input, such as Harada et al. [2012], Jiang et al. [2012], Manuelli et al. [2019] and Mitash et al. [2020]. Haustein et al. presents a planning algorithm for stable object placement in cluttered scenes requiring a fully specified environment. Closely related to our work is that of Newbury et al. [2021], which presents an iterative learning-based approach for placing household objects onto flat surfaces but again using a system of three external depth cameras for input. Relying on image data might be problematic due to gripper-object occlusions in highly cluttered scenes, and tedious, time-consuming robot-camera calibrations. In contrast, tactile sensors directly provide the contact information between the object and gripper, independent of the surrounding environment.

In-hand object pose estimation. Due to the inherent difficulties of estimating a grasped object's pose and due to its importance for tasks like pick & place or in-hand manipulation, multiple methods for object-in-hand pose estimation have been developed. In this area, tactile-based models (Bimbo et al. [2016], Kelestemur et al. [2022]) and tactile-vision (Álvarez et al. [2017], Anzai and Takahashi [2020]) hybrids are more prominent. Doosti et al. [2020] and Sodhi et al. [2022] present end-to-end approaches based on RGB images. In contrast to many of these works, we are not interested in the full 6D object pose, but rather its orientation relative to a placing surface. This allows us to employ a small and efficient neural network that is not as demanding in terms of data,

3 Stable Object Placement

Fig. 1a shows the four phases of stable object placement, starting with the object state estimation in the first phase. We define the object's placing normal $\vec{z}_p = \mathbf{R}^W_O \mathbf{R}^O_{O'} \vec{z} = \mathbf{R}^W_{O'} \vec{z}$ to lie along the $\vec{z}$-axis in the so-called local object placing frame O' with the rotation matrix from a reference frame, e.g. "world", to the object placing frame given by $\mathbf{R}^W_{O'} \in SO(3)$. $\mathbf{R}^W_{O'}$ can be decomposed as $\mathbf{R}^W_{O'} = \mathbf{R}^W_G \, \mathbf{R}^G_{O'}$, where G denotes the gripper frame, whose pose w.r.t. worl is usually known from proprioception. It thus suffices to estimate $\mathbf{R}^G_{O'}$ to correct the object pose. This necessitates estimating the object's in-hand pose at least partially in order to generate the desired corrective motion. Assuming $\mathbf{R}^G_{O'}$ is known, an IK solution can then be computed that moves the gripper to a pose that aligns the two placing normals. Then, the placing motion is planned and executed that move the object towards the table while preserving its orientation to the world frame. Once table-object contact is detected, the object is released and the arm retracted.

Then, the placing motion is realized by executing a linear downward movement in Cartesian space. In our case, we simply use TIAGo's torso lift for this motion, while constraining IK solutions to be elbow-up such that the gripper will be the first part of the arm that will acquire table contact. During the motion, the grasping forces are monitored, where an unexpected spike is interpreted as table contact acquisition, which terminates the torso trajectory. Finally, we open the gripper and drive the torso up again to retract.

Table 1: Comparison of different placing methods on the two 3D-printed training objects (cf. Fig. 2a). The angular error is the angle between the predicted $\vec{z}_p$ and the one measured with OptiTrack, given in radians.

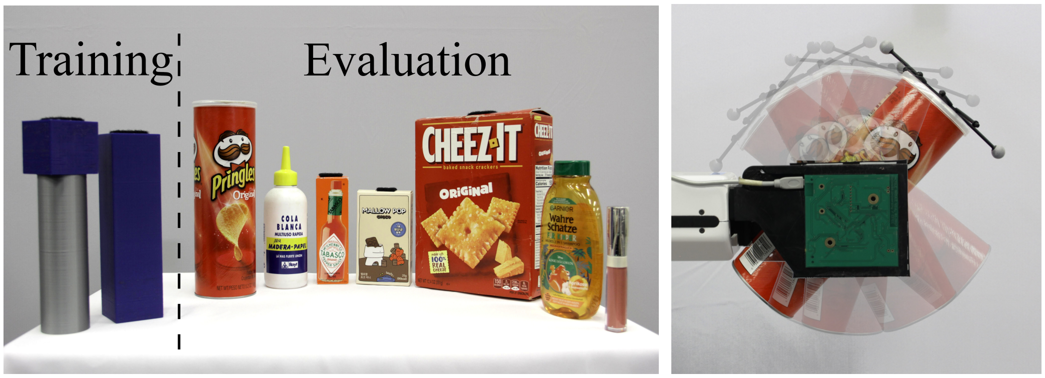
Method	Metric	Cylinder	Cuboid	Average
OptiTrack	% Suc.	90%	**95%**	**92.5**%
	Ang.Err.	-	-	-
Tactile PlaceNet	% Suc.	**95%**	85%	90.0%
	Ang.Err.	**0.06 ± 0.03**	0.17 ± 0.12	**0.11 ± 0.08**
Tactile + F/T PlaceNet	% Suc.	90%	85%	87.5%
	Ang.Err.	0.08 ± 0.04	**0.16 ± 0.19**	0.12 ± 0.08
F/T PlaceNet	% Suc.	15%	25%	20.0%
	Ang.Err.	0.38 ± 0.26	0.39 ± 0.32	0.38 ± 0.29
PCA (Tactile)	% Suc.	90%	10%	50%
	Ang.Err.	0.07 ± 0.02	0.83 ± 0.42	0.45 ± 0.22
Hough (Tactile)	% Suc.	80%	10%	45%
	Ang.Err.	0.09 ± 0.04	0.76 ± 0.37	0.42 ± 0.21

To estimate the object rotation w.r.t the gripper frame $\mathbf{R}_{O'}^{G}$, we propose to use a convolutional neural network. Tactile data from the two gripper-mounted Myrmex sensors is represented as a 16×16 matrix of normal force readings that are normalized in $[0, 1]$. To process the tactile data, we first use two convolutional layers with a 3×3 kernel each, and 16 and 32 output channels respectively. The output of the last convolutional layer is then fed to a Multilayer Perceptron (MLP) consisting of two hidden layers with 128 neurons each and ReLU activation functions, followed by a dropout layer with a dropout probability of $p = 0.2$. F/T data can be optionally fed into the MLP as an additional input signal, which is concatenated with the tactile features. The general architecture of PN is visualized in Fig. 1b, including Myrmex sensor samples on the left. To smoothly represent the rotation matrix in the output layer, we use the 6D representation comprising the first two columns of $\mathbf{R}_{O'}^{G}$, as introduced in Zhou et al. [2019], as well as a slight variation of their loss function which ignores rotation errors around the placing normal.

4 Experimental Evaluation

We used two primitive 3D-printed objects for data collection as shown on the left in Fig. 2a. The in-hand object poses $\mathbf{R}_{O'}^{G}$ were measured by OptiTrack, an external camera-based tracking system. Data was collected for different arm postures and object in-hand poses, amassing 800 samples per object and 1600 in total. We trained every PlaceNet for 40 epochs and reserved 20% of the data for testing. The tactile-only PlaceNet (PN) and tactile with F/T PN achieved a test error of 0.03 rad and 0.05 rad respectively, indicating that they were able to estimate the object orientation with high precision. On the other hand, the F/T-only PN performed rather poorly with a minimum test error of 0.43 rad, indicating that F/T data alone is not sufficient to estimate the object state. To assess whether a data-driven model is required to solve this task robustly and to gauge its performance compared to other approaches, we introduce two baseline models for comparison. Given the nature of our sensory data, line-fitting methods naturally come to mind. We chose two popular methods from that field, Principal Component Analysis (PCA) (Pearson [1901]) and Hough transforms (Duda and Hart [1972]). For the real-world evaluation of placing success, we let each method place an object from 5 different arm poses and 4 different in-hand object poses, yielding 20 samples per object for each method (see Fig. 2b). We report the estimation accuracy by again using OptiTrack to measure the object pose and the success rate of the actual placing trials.

We first compare the OptiTrack baseline with all three PlaceNets (PNs) and the two line-fitting baselines on the training objects. Evaluating 6 methods on the two 3D-printed training objects and conducting 20 trials per object results in 240 placing trials overall. Table 1 shows the results. We can compare the OptiTrack baseline to the other methods based on success rate and the angular error between their predictions of the object normal and OptiTrack's ground-truth measurement for $\vec{z}_p^{G}$. A surprising result is that the tactile-only PN performed better in terms of success rate than the OptiTrack baseline, which can be attributed to occasional marker loss due to occlusions and imperfect IK solutions. Among the neural networks, the tactile-only PN performed best in almost all metrics. The evaluation also confirmed another indication from training, namely that the F/T-only PN did not succeed in estimating the object's placing normal sufficiently well for stable placing. The baselines perform well on cylindrical objects, and PCA even comes close to PN performance, while showing difficulties in estimating the cuboid object's pose, likely because cuboid objects create less distinct line patterns in the sensor image.



(a) Objects used during data collection (left) and for out-of-distribution evaluation (right).

(b) Variations of in-hand object poses used during evaluation.

Figure 3: Experiment objects and object poses used for evaluation and data collection.

Method	Metric	Pringles	Glue Bottle	Tabasco	Mallow Pop	Cheez It	Shampoo	Lipstick	Average
Tactile PlaceNet	% Suc.	**90%**	85%	**85%**	**80%**	80%	**85%**	**90%**	**85%**
	Ang.Err.	**0.07 ± 0.03**	0.08 ± 0.04	**0.10 ± 0.13**	**0.16 ± 0.10**	0.10 ± 0.06	**0.06 ± 0.03**	-	**0.09 ± 0.07**
Tactile + F/T PlaceNet	% Suc.	85%	80%	70%	70%	**85%**	75%	80%	77%
	Ang.Err.	0.13 ± 0.20	0.09 ± 0.04	0.16 ± 0.10	0.14 ± 0.10	**0.05 ± 0.06**	0.10 ± 0.06	-	0.12 ± 0.10
PCA (Tactile)	% Suc.	**90%**	**90%**	15%	20%	80%	75%	85%	65%
	Ang.Err.	0.08 ± 0.06	**0.07 ± 0.03**	0.79 ± 0.61	0.70 ± 0.45	0.09 ± 0.05	0.08 ± 0.0	-	0.30 ± 0.24
Hough (Tactile)	% Suc.	85%	80%	60%	50%	70%	70%	80%	70%
	Ang.Err.	0.11 ± 0.06	0.10 ± 0.03	0.20 ± 0.30	0.35 ± 0.37	0.12 ± 0.08	0.16 ± 0.32	-	0.17 ± 0.17

Table 2: Experimental results for seven household testing objects (cf. Fig. 2a). All objects were unknown to the models before this evaluation, i.e., they were not present in the training set. Angular errors are reported in radians.

After confirming the training results on the real robot in our first experiment, we conducted a second evaluation on objects that were not present in the training data using the most promising models from the previous experiment. We only evaluated the two best PlaceNets (PNs) from the previous experiment, namely the tactile-only PN and the tactile with F/T PN, along with the baselines. We evaluated the 4 methods on 7 (see Fig. 2a on the right side) different household objects for 20 trials each, hence performing 560 placing trials in total, with results given in Table 2. We, again, report the success rate of correct placements and the angular error between the model's predictions and OptiTrack where possible (the lipstick was too small to attach a marker to). The PNs perform very well on most unknown objects, indicating that our method generalizes across object primitives of unknown dimensions. Similar to the results from the previous evaluation on known objects, the tactile-only PN showed superior performance in most cases, as it performed best on average in both metrics with a low variance in results. The tactile + F/T PN only performed better in terms of both metrics when placing the Cheez-It box. We hypothesize that the slightly worse performance might result from the network receiving rather non-informative F/T signals alongside the tactile data.

PCA and Hough both showed similar results to the first experiment. While Hough's performance does not show such a large difference between cuboids and cylinders as PCA, it did not manage to achieve satisfactory results consistently. Finally, all models showed reasonable to good performance on the Lipstick, an object that is difficult to place due to its small placing faces. The tactile-only network performed best among all methods, further underlining its generalizability.

5 Conclusion

Our results show that tactile data is a crucial source of information for predicting the object's placing normals, whereas F/T data has not been proven to be as informative for this task. Furthermore, our evaluation has shown that the PNs generalize to unknown objects with high success rates and precision. It also revealed that classical approaches can work reliably on a set of objects with specific characteristics (cylinders) while not being accurate enough on objects that lack these (boxes). Future work might extend our approach to incorporate active touch scenarios where the object is already in contact with a placing surface. This could make it applicable to a larger variety of objects, e.g. those that cover the whole sensor surface and render line fitting impossible.

Acknowledgments

This work was supported by the European Union's Horizon 2020 Marie Curie Actions under grant no. 813713 NeuTouch, the Horizon Europe research and innovation program under grant no. 101070600 SoftEnable, the DFG Emmy Noether Programme (CH 2676/1-1), the BMBF Project Sim4Dexterity, the BMBF Project Aristotle, the German Research Center for AI (DFKI), Hessian.AI, and the AICO grant by the Nexplore/Hochtief Collaboration with TU Darmstadt.

References

Yun Jiang, Marcus Lim, Changxi Zheng, and Ashutosh Saxena. Learning to place new objects in a scene. *The International Journal of Robotics Research*, 31(9), 2012.

Kensuke Harada, Tokuo Tsuji, Kazuyuki Nagata, Natsuki Yamanobe, Hiromu Onda, Takashi Yoshimi, and Yoshihiro Kawai. Object placement planner for robotic pick and place tasks. In *Proc. IROS*, 2012.

Jiayao Ma, Weiwei Wan, Kensuke Harada, Qiuguo Zhu, and Hong Liu. Regrasp planning using stable object poses supported by complex structures. *IEEE TR-CDS*, 11(2), 2018.

Chaitanya Mitash, Rahul Shome, Bowen Wen, Abdeslam Boularias, and Kostas Bekris. Task-driven perception and manipulation for constrained placement of unknown objects. *IEEE RAL*, 5(4), 2020.

Lucas Manuelli, Wei Gao, Peter Florence, and Russ Tedrake. kpam: Keypoint affordances for category-level robotic manipulation. In *The International Symposium of Robotics Research*. Springer, 2019.

Joshua A Haustein, Kaiyu Hang, Johannes Stork, and Danica Kragic. Object placement planning and optimization for robot manipulators. In *Proc. IROS*. IEEE.

Rhys Newbury, Kerry He, Akansel Cosgun, and Tom Drummond. Learning to place objects onto flat surfaces in upright orientations. *IEEE RAL*, 6(3), 2021.

Carsten Schürmann, Risto Kõiva, and Robert Haschke. A Modular High-Speed Tactile Sensor for Human Manipulation Research. 2011 IEEE World Haptics Conference, 2011.

Joao Bimbo, Shan Luo, Kaspar Althoefer, and Hongbin Liu. In-hand object pose estimation using covariance-based tactile to geometry matching. *IEEE RAL*, 1(1), 2016.

Tarik Kelestemur, Robert Platt, and Taskin Padir. Tactile pose estimation and policy learning for unknown object manipulation. *arXiv:2203.10685*, 2022.

David Álvarez, Máximo A Roa, and Luis Moreno. Tactile-based in-hand object pose estimation. In *Iberian Robotics conference*. Springer, 2017.

Tomoki Anzai and Kuniyuki Takahashi. Deep gated multi-modal learning: In-hand object pose changes estimation using tactile and image data. In *IROS*. IEEE, 2020.

Bardia Doosti, Shujon Naha, Majid Mirbagheri, and David J Crandall. Hope-net: A graph-based model for hand-object pose estimation. In *Proc. CVF*, 2020.

Paloma Sodhi, Eric Dexheimer, Mustafa Mukadam, Stuart Anderson, and Michael Kaess. Leo: Learning energy-based models in factor graph optimization. In *CoRL*, 2022.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 1901.

Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1), 1972.

Siciliano B Slotine and B Siciliano. A general framework for managing multiple tasks in highly redundant robotic systems. In *proceeding of 5th International Conference on Advanced Robotics*.

Appendix

Stable object placing

As illustrated in Fig. 4a, we define the object's placing normal $\vec{z}_p = \mathbf{R}_O^W \mathbf{R}_{O'}^O \vec{z} = \mathbf{R}_{O'}^W \vec{z}$ to lie along the $\vec{z}$-axis in the so-called local object placing frame O' with the rotation matrix from a reference frame, e.g. "world", to the object placing frame given by $\mathbf{R}_{O'}^W \in SO(3)$. $\mathbf{R}_O^W$ & $\mathbf{R}_{O'}^O$, thus, describe the rotation matrices from the world to the object frame, and from the object frame to the object's placing frame, respectively. Note that we deliberately introduce this local object placing frame O', in addition to the object's pose frame O, for two reasons. On the one hand, there might exist multiple placing frames per object, and on the other hand, to highlight that knowing the object's pose might not be informative enough for stable placing, for instance, in the case of lacking precise information about the object's geometry. Since we only consider scenarios where the object was already grasped, $\mathbf{R}_{O'}^W$ can be decomposed into two distinct rotation matrices

$$\mathbf{R}_{O'}^W = \mathbf{R}_G^W \, \mathbf{R}_{O'}^G, \tag{1}$$

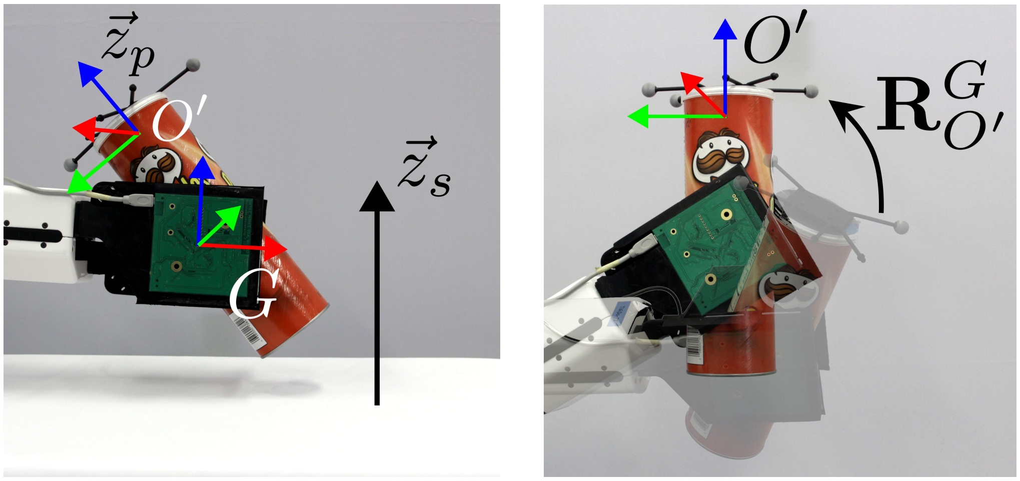
where $\mathbf{R}_G^W, \mathbf{R}_{O'}^G \in SO(3)$ describes the orientation of the gripper w.r.t. the world and the object's placing frame within the gripper, respectively. While the former can be reliably estimated via forward kinematics from proprioceptive feedback, the latter is usually unknown. Here, we make the assumption, that one suitable placing face of the object is oriented toward the ground, which is generally the case after grasping or human-robot object handovers of various household objects. Thus, $\mathbf{R}_{O'}^G$ has to be estimated based on sensor measurements as it is an indispensable ingredient for generating a motion that corrects the object misalignment and enables appropriate placing.

Next, we require a placing controller that, given $\mathbf{R}_{O'}^G$, generates two movements: a corrective arm movement that aligns $\vec{z}_p$ with $\vec{z}_s$ (cf. Fig. 4b) and a downward placing motion that stops on table contact and releases the object afterward. For the corrective movement, we first project $\vec{z}_s$ into the local gripper frame:

$$\vec{z}_s^G = \mathbf{R}_W^G \vec{z} \tag{2}$$

Then, we calculate the rotation $\mathbf{R}_{G'}^G$ that rotates the current gripper frame such that in the resulting one (G'), $\vec{z}_p^{G'}$ aligns with $\vec{z}_s^{G'}$. This is achieved by finding the rotation axis with $\vec{a} = \vec{z}_p^G \times \vec{z}_s^G$ and the rotation angle $\theta = \cos^{-1}(\vec{z}_p^G \cdot \vec{z}_s^G)$, with the vector cross-product $\times$. Using the rotation $\mathbf{R}_{G'}^G$, we can try to find an inverse kinematics (IK) solution that realizes the object reorientation.

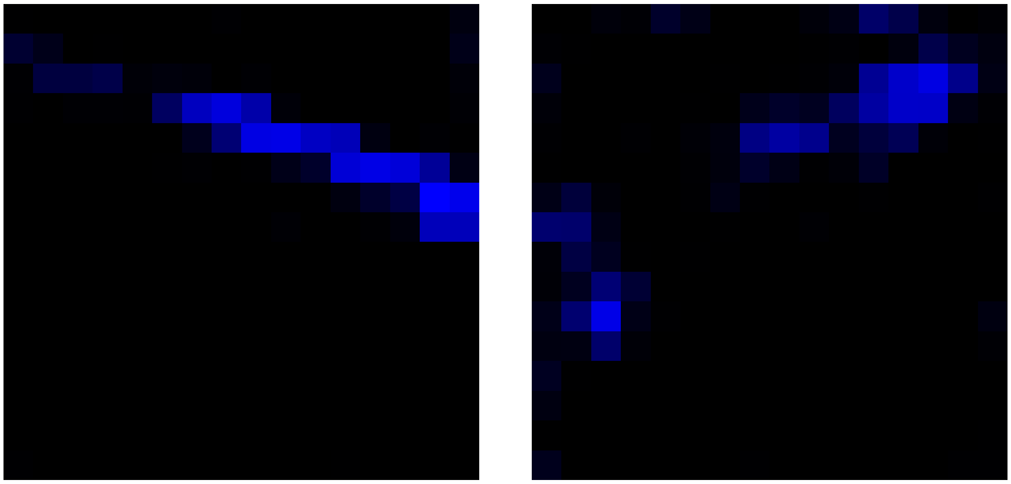
An IK solution that leads to a secure placing configuration should satisfy more constraints than the object reorientation alone. After reorientation, we will execute a linear downward movement in



(a) Frames and placing normal

(b) Corrective object motion

Figure 4: Illustration of the problem setting. (a) shows the gripper frame G, a possible object placing frame O' and the surface's placing normal $\vec{z}_s$. (b) Result of the corrective motion is to align the object with the placing surface based on the estimation of $\mathbf{R}_{O'}^G$.



(a) Left Myrmex Sensor

(b) Right Myrmex Sensor

Figure 5: Raw sensor readings of both Myrmex sensors while holding the cylindrical object. Each reading of the 16×16 taxels corresponds to the currently measured normal force. The images were scaled up, and the force readings were amplified for visualization purposes.

Cartesian space to acquire table-object contact, and we need to ensure that the object will be the first point of contact with the table. Arm configurations suitable for placing thus need to ensure that the object frame (located in the gripper) has a lower z-coordinate in the world frame than the wrist. The commonly used hierarchy-of-tasks approach Slotine and Siciliano allows us to formulate such additional constraints for the IK. Since many IK solvers are sensitive to the initial arm configuration, we searched for solutions starting from 20 different initial poses, which are sensible placing configurations, and chose the one with the lowest error that is most similar to our current arm pose. The motion to reach this solution is generated by linear interpolation in joint space starting from the robot's current arm position.

For the placing motion, a trajectory is generated that moves the gripper linearly downward. This is realized using TIAGo's torso joint. During this motion, tactile measurements are continuously monitored. We interpret a spike in tactile sensation as making contact with the table. The torso controller is then signaled to halt execution. Finally, the object is released by opening the gripper and moving the torso upward again.

Object pose correction estimation with tactile sensing

As motivated previously, the key component for stable object placing is the estimation of the object placing frame w.r.t. the gripper frame ($\mathbf{R}^{G}_{O'}$). According to this transformation, the object is reoriented prior to placing. However, determining this quantity is difficult, as the object is handed over in an unknown pose, it is occluded by the gripper, and herein we also do not assume any prior knowledge about the object type. We, therefore, propose estimating $\mathbf{R}^{G}_{O'}$ from the signals of the tactile sensors inside the gripper. The Myrmex tactile data is represented as a 16×16 matrix of normal force readings that are normalized in $[0, 1]$. Fig. 5 shows a visualization of the tactile readings from the right and left sensors while holding an object. Next, we, first, introduce our proposed Neural Network, and, subsequently, explain line-fitting baselines for recovering $\mathbf{R}^{G}_{O'}$ from the tactile readings.

Since we require a solution that is flexible enough to deal with different objects, that can handle the sensors' noise, and convert the high-dimensional readings into a signal suitable for reorienting the objects, we propose to employ a neural network. The general network architecture is visualized in Fig. 1b. To process the tactile data, we first use two convolutional layers with a 3×3 kernel each, and 16 and 32 output channels respectively. The output of the last convolutional layer is then fed to a Multilayer Perceptron (MLP) consisting of two hidden layers with 128 neurons each and ReLU activation functions, followed by a dropout layer with a dropout probability of $p = 0.2$. F/T data can

	Tactile PlaceNet	F/T PlaceNet	Tactile + F/T PlaceNet
Test Loss (rad)	**0.03**	0.43	0.05

Table 3: Training results of networks with different input sensor modalities. We report the lowest test loss averaged over 10 batches.

be optionally fed into the MLP as an additional input signal, which is concatenated with the tactile features.

To smoothly represent the rotation matrix in the output layer, we use the 6D representation comprising the first two columns of $\mathbf{R}_{O'}^{G}$, as introduced in Zhou et al. [2019] and has been shown to exhibit superior properties for learning in $SO(3)$. Each estimate is then converted into an $SO(3)$ rotation matrix for the computation of the loss, which is defined as

$$\mathcal{L}(\mathbf{R}, \vec{z}_p^{\,gt}) = \cos^{-1}\Big(\mathbf{R}_G^W \, \mathbf{R} \, \vec{z} \cdot \vec{z}_p^{\,gt}\Big), \tag{3}$$

where $\mathbf{R} = \mathbf{R}_{O'}^{G}$ is the quantity of interest and the prediction of the network, and $\vec{z}_p^{\,gt}$ is the ground truth measurement of the object's placing normal in the world frame that is obtained through an OptiTrack motion capture system. By taking $\cos^{-1}$, the loss lies in the interval $[0, \pi]$, and can be interpreted as the angular distance error between predicting $\vec{z}_p$ using the network's output $\mathbf{R}$ and the ground truth. Note that our loss, contrary to e.g. the geodesic loss from Zhou et al. [2019], does not consider rotations about the placing normal since those are irrelevant in our problem definition. Table 3 shows the training results, and clearly shows that both PNs with tactile information were able to precisely estimate the object state, while the F/T-only PN was not.

Line-fitting baselines

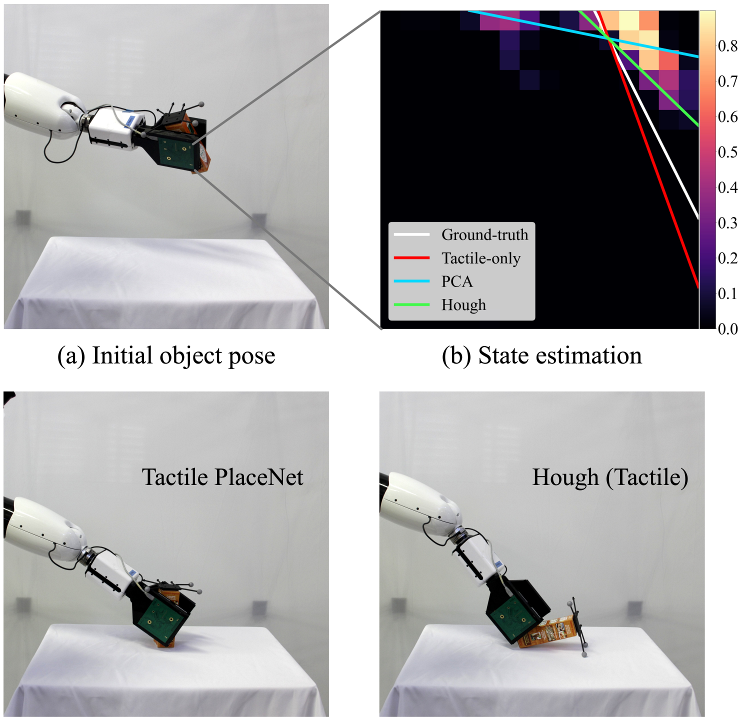
To assess whether a data-driven model is required to solve this task robustly and to gauge its performance compared to other approaches, we introduce two baseline models for comparison. Given the nature of our sensory data and the goal of finding the object's main axis within it, line-fitting methods naturally come to mind. We chose two popular methods from that field, Principal Component Analysis (PCA) Pearson [1901] and Hough transforms Duda and Hart [1972]. As both methods work on individual images, we combine the two sensor readings into one frame by flipping one of the sensor images to account for symmetry. Therefore, the input to the baselines contains the information of both sensors. This should increase robustness as the sensors might be differently affected by noise.

For the PCA baseline, we treat the force readings as a bi-variate, uni-modal Gaussian, and estimate its mean, standard deviations, and covariance matrix $\mathbf{C}$. To obtain the orientation of the object's main axis, we calculate the first principal component of $\mathbf{C}$ using PCA. Assuming that an object's main axis lies along its largest grasping surface, the first principal component should constitute a decent estimation for said axis. We then calculate the angle of the line relative to the sensor and transform it into the rotation $\mathbf{R}_{O'}^{G}$, allowing us to generate a corrective motion that aligns the objects.

Hough transform is a common tool in image processing for finding lines in images. Typically, a raw input image undergoes some preprocessing where edges are extracted and finally, the Hough transform is applied to the resulting binary image. Empirically, we have found the Hough transform to show better performance if we simply create a binary image by assigning a value of 1 to taxels with a force above a noise threshold and 0 otherwise. We only consider the resulting line with the most votes (the most confident estimate). Lines are parametrized by the angle ψ between the x-axis and a line normal that intersects the origin and the distance to the origin of said normal. From ψ, we calculate the angle between the x-axis and the line itself $\phi = \pi - \psi$ and again calculate $\mathbf{R}_{O'}^{G}$ as with the other baseline.

Model trials and visualizations

Fig. 6 shows placing trials of the box-like Tabasco object, where the tactile-only model successfully placed the object whereas both classical models' predictions led to failed placements. In Fig. 6(b), the normalized sensor image is shown which is used as input for PCA and Hough and serves well for visualization purposes as well. Each taxel value is normalized between 0 (= no force) and 1 (= highest measurable force), and then colorized accordingly. Then, the ground truth data from OptiTrack and



(a) Initial object pose

(b) State estimation

(c) Successful placement

(d) Failed placement

Figure 6: Placement sequence comparing the tactile-only PlaceNet (PN) with the baselines. The PN performed best with an angular error of 0.11 and placed Tabasco correctly, while Hough and PCA both failed with estimation errors of 0.34 and 0.91 respectively. (b) depicts the raw, normalized force readings for each taxel along with each model's prediction of the object's placing normal and the ground -truth obtained from OptiTrack.

the model output from the performing network and from the two baselines are superimposed over the resulting image. It is clear that the tactile-only PN was able to predict the object state precise enough to perform a successful placement (see Fig. 6(c)), while the most accurate baseline estimate (in this case, from Hough), was not good enough for the object to be placed stably (see Fig. 6(d)).