NeuralTouch: Leveraging Implicit Neural Descriptor for Precise Sim-to-Real Tactile Robot Control

Yijiong Lin, Bowen Deng, Chenghua Lu, Max Yang, John Lloyd, Efi Psomopoulou, Nathan F. Lepora

I. INTRODUCTION

An everyday human ability that remains challenging for robots is our capacity to glance at an object to gauge its general location and then rely solely on touch to grasp it precisely. For example, after seeing where a plug is inserted into a socket, we can unplug it using just our sense of touch. In robotics, replicating this process typically involves two phases: (1) an initial coarse phase where vision captures global information essential for contact-rich tasks, and (2) a subsequent fine phase where touch refines the grasping pose, utilizing prior visual information about the object's position and geometry.

However, the current use of vision and touch in robotic manipulation faces several limitations: (1) it is restricted to scenarios where objects are already optimally positioned for grasping at the start; (2) the policies are limited to manipulating objects or contact features seen during training, lacking generalization to novel objects; (3) the independent use of vision and touch reduces their potential for synergy; and (4) multimodal policies developed in simulation often struggle to transfer effectively to real-world environments. This paper addresses these challenges by introducing a novel multimodal policy learning framework designed to overcome these limitations.

In this work, we introduce *NeuralTouch*, a tactile reinforcement learning (RL) policy framework augmented with neural descriptor fields (NDFs) [1]. The primary aim is to enhance the grasping accuracy of NDF-based methods by incorporating touch, while preserving broad generalizability across diverse object categories. Additionally, this framework removes the limitations of predefined contact geometries, enabling NDF-based tactile servoing to operate flexibly across various contact scenarios.

All authors are with the Department of Engineering Mathematics and Bristol Robotics Laboratory, University of Bristol, Bristol BS8 1UB, U.K. (email: {yijiong.lin, n.lepora}@bristol.ac.uk)

As shown in Fig. 2, we separate the robotic grasping task into two phases: the coarse phase and the fine phase. Note that while we structure this task similarly to [2], we do not rely on any specific methods from their proposed approach. In the coarse phase, we leverage the descriptor generated from NDFs to calculate the coarse target grasping pose. Then, in the fine phase, we apply a tactile RL policy [3], [4] to accurately grasp an object with a desired contact configuration represented by the NDF descriptor.

Specifically, we focus on learning a tactile RL policy that can be generalized to different target contact configurations for different objects or tasks with the help of implicit neural descriptors from NDFs. Specifically, the tactile RL policy should not only consider the local contact to achieve safe, gentle contact but also have a sense of its desired contact configuration with respect to the global shape of an object. Our method consists of three modules:

1) A PointNet Encoder [5] and Neural Pose Descriptor Fields [1] that learn implicit descriptors for various object shapes. These implicit representations are geometric-relevant, describing the relationships between a pose (local frame) and the corresponding local shape of an object.
2) Initial coarse grasping pose generation with pose regression using NDFs [1].
3) NeuralTouch RL, which learns a general tactile robotic policy that is conditioned on the implicit neural descriptors to achieve desired contact configuration (fine grasping pose)

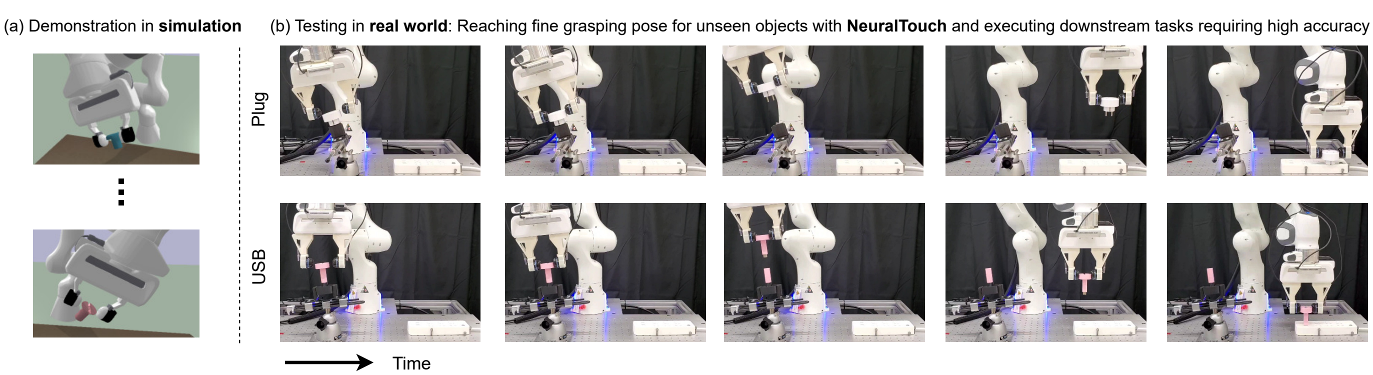

Figure 1. With just a few demonstrations of a manipulation task in simulation, our method *NeuralTouch* effectively generalizes to achieve accurate grasping poses for unseen objects in real-world settings, utilizing both vision and touch.

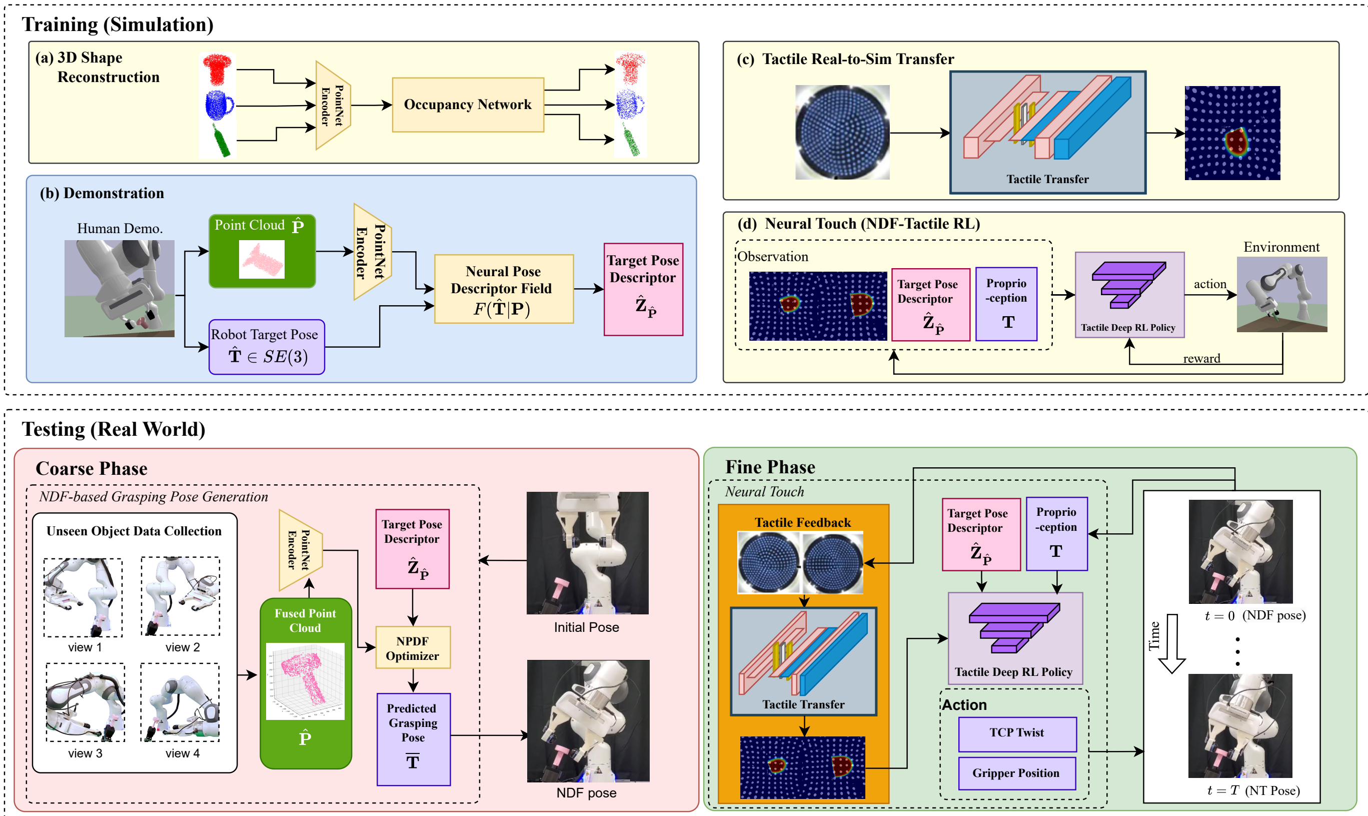

Figure 2. Overview of the NeuralTouch: In simulation, we first pre-train an occupancy network which is the core component of the Neural Pose Descriptor Fields. Secondly, we collect human demonstrations along with object point clouds and robot target grasping pose descriptors depending on the manipulation tasks. Thirdly, we train an RL policy with tactile and proprioceptive feedback, to achieve fine grasping poses implicitly specified by these collected descriptors. After obtaining the NPDF and a well-trained policy, our system is directly deployed in the real world with a real-to-sim tactile transfer to accurately grasp unseen objects, executing manipulation tasks such as unplugging a bolt-like USB and inserting it into a socket.

while maintaining safe, gentle physical interaction between a tactile robot and a manipulated object, given the tactile and proprioceptive feedback.

In general, we focus on achieving precise grasping with a tactile gripper, a task that involves two distinct phases. In the coarse phase, we leverage NDFs to generate an initial pre-grasping pose. The fine phase then employs in-hand tactile servoing across the gripper fingers, crucial for practical applications such as repositioning and reorienting the gripper to secure a specific grasp. This process is particularly challenging, as it requires an understanding of the object's underlying geometry and precise 6-DoF robotic control.
Preliminary experimental results show that NeuralTouch can serve as a powerful complement to the state-of-the-art vision-based grasping method (NDF), achieving the desired grasping pose with greater accuracy. Furthermore, extensive testing with zero-shot sim-to-real policy transfer and few-shot demonstrations underscores the adaptability of our approach in solving diverse downstream tasks for various real-world objects.

References

[1] Anthony Simeonov, Yilun Du, Andrea Tagliasacchi, Joshua B Tenenbaum, Alberto Rodriguez, Pulkit Agrawal, and Vincent Sitzmann. Neural descriptor fields: Se (3)-equivariant object representations for manipulation. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 6394–6400. IEEE, 2022.

[2] Edward Johns. Coarse-to-fine imitation learning: Robot manipulation from a single demonstration. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4613–4619, 2021.

[3] A. Church, J. Lloyd, R. Hadsell, and N. Lepora. Tactile Sim-to-Real Policy Transfer via Real-to-Sim Image Translation. In *Proceedings of the 5th Conference on Robot Learning*, pages 1–9. PMLR, October 2021.

[4] Y. Lin, J. Lloyd, A. Church, and N. Lepora. Tactile gym 2.0: Sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch. volume 7 of *Proceedings of Machine Learning Research*, pages 10754–10761. IEEE, August 2022.

[5] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.