Integrating Model-based Control and RL for Sim2Real Transfer of Tight Insertion Policies

Isidoros Marougkas*,1, Dhruv Metha Ramesh*,1, Joe H. Doerr[1], Edgar Granados[1], Aravind Sivaramakrishnan[2], Abdeslam Boularias[1], and Kostas E. Bekris[3]

***Abstract*— Object insertion under tight tolerances ($< 1mm$) is an important but challenging assembly task as even small errors can result in undesirable contacts. Recent efforts focused on Reinforcement Learning (RL), which often depends on careful definition of dense reward functions. This work proposes an effective strategy for such tasks that integrates traditional model-based control with RL to achieve improved insertion accuracy. The policy is trained exclusively in simulation and is zero-shot transferred to the real system. It employs a potential field-based controller to acquire a model-based policy for inserting a plug into a socket given full observability in simulation. This policy is then integrated with residual RL, which is trained in simulation given only a sparse, goal-reaching reward. A curriculum scheme over observation noise and action magnitude is used for training the residual RL policy. Both policy components use as input the `SE(3)` poses of both the plug and the socket and return the plug's `SE(3)` pose transform, which is executed by a robotic arm using a controller. The integrated policy is deployed on the real system without further training or fine-tuning, given a visual `SE(3)` object tracker. The proposed solution and alternatives are evaluated across a variety of objects and conditions in simulation and reality. The proposed approach outperforms recent RL-based methods in this domain and prior efforts with hybrid policies. Ablations highlight the impact of each component of the approach. For more information please refer to the corresponding website.**

I. Introduction

This paper addresses object insertion under tight tolerances ($< 1mm$). Given visual tracking of the `SE(3)` object pose, this work proposes a strategy for learning a policy for tight insertion into a socket. A key feature of the proposed strategy is that it first defines a model-based control solution, which is then complemented with a residual policy trained via Reinforcement Learning (RL) in simulation to address the uncertainty arising from perception noise and contact dynamics. The policy trained in simulation is directly deployable on the real system without any fine-tuning.

Tight object insertion is applicable both in industrial and domestic setups, from product part assembly to plugging sockets of home devices. Thus, peg-in-hole challenges have long been the focus of robotics research [**?**], [1]–[3] as a contact-rich manipulation task. Nevertheless, the sub-millimeter precision required to complete such tasks and the uncertainty regarding the objects' states has limited the real-world deployment of developed solutions.

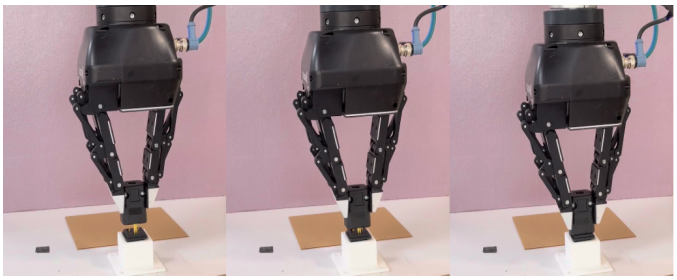


Fig. 1: Zero-shot transfer of the policy learned in simulation to a real forceful insertion of an unseen plug and socket.

Model-based efforts [3]–[6] have engineered control policies for insertion that can be effective for a well-instrumented workspace setup. Such solutions, however, are brittle to changes in the workspace, and do not generalize easily to new objects. Recent **data-driven approaches** have attempted to solve the problem by learning policies either from human demonstrations [7], [8] or from online interaction via RL [9]–[11]. They can generalize to new objects, nevertheless, they require significant demonstration effort, reward engineering, and incur high sample complexity. In particular, recent work that is closely related to this paper, IndustReal [9], demonstrated peg insertion accuracy at $\sim 85\%$ under varying initial conditions and perception noise through the use of RL in simulation given dense reward engineering and curriculum learning. While IndustReal is a state-of-the-art result in zero-shot transfer of the RL policy trained in simulation, success rate can be further improved, especially for workspace setups that are less carefully instrumented.

The **current work** seeks to achieve higher success rates in tight insertion tasks while minimizing engineering effort and enabling zero-shot transfer from simulation to reality. This is accomplished through the integration of model-based reasoning, RL and other key components:

1. The method begins with a straightforward potential field policy that operates over the `SE(3)` pose observations of the plug and socket. This *model-based policy* evaluates to near-perfect success rates under zero observation noise in simulation, but its success rate decreases drastically with observation noise.
2. A *residual policy* is trained using a reinforcement learning (RL) objective with only sparse, goal-reaching rewards. The output of the residual policy is added to the output of the model-based policy. This residual policy is trained in a physics-based simulator (IsaacGym [12]) with noisy pose

*The first two authors contributed equally to this paper. [1]Dept. of Computer Science, Rutgers University, NJ, USA. [2]A.S. is affiliated with Amazon.com Inc. [3]K. E. Bekris holds concurrent appointments as a Professor at Rutgers University and as an Amazon Scholar. This paper describes work performed at Rutgers and is not associated with Amazon. The work has been partially supported by NSF awards NRT-FW-HTF 2021628, FRR 2309866 and POSE 2346069. Opinions expressed here are those of the authors and do not reflect positions of the funding agency. Corresponding author e-mails: {im316,ab1544,kostas.bekris}@cs.rutgers.edu.

observations. The policy is trained to correct for errors introduced by perception noise and unforeseen contacts during real-world execution. A *curriculum-based training scheme* incrementally increases the difficulty of the task, adjusting the noise level of the plug's pose observations and the magnitude of the residual RL policy's actions. As noise increases, the policy relies more on the RL component and less on the model-based potential field, ultimately balancing out their contributions.

3. The combined policy is then transferred directly from simulation to a real robot without fine-tuning. A vision-based pose estimation module detects the socket's configuration and tracks the plug's pose. The $\mathtt{SE(3)}$ plug transforms are returned by the policy and converted into joint controls for the robotic arm controller.

The accompanying experiments demonstrate that this integrated approach significantly improves insertion task success rates compared to alternatives, including IndustReal [9].

II. Related Work

This section reviews prior efforts on tight insertion, ranging from model-based to RL-based techniques.

Model-based Insertion Strategies Classical approaches for robotic tight insertion rely on model-based planning and control, that vary from integrating manipulation primitives for fine assembly [2] to assembly-by-disassembly [13] and continuous visual servoing [14]. Continuous object tracking has also been integrated with passively adaptive mechanical hardware for tight insertion [3]. In general, active and passive compliance can be beneficial for insertion [4], [5], [15]. Some efforts focus on contact-based search strategies, such as spiral and random motions. Various frameworks have been proposed to discover such solutions, including Finite-State Machine Controllers [16], Task-and-Motion Planning [17] and tactile-based behavior trees [6]. Search assembly strategies have been evaluated given position uncertainty estimation [18]. Socket-location probability distributions can be estimated to devise a search trajectory [19]. Deploying the aforementioned model-based strategies for insertion in the real world can be challenging due to pose uncertainties of the plug and the socket.

Machine Learning Insertion Strategies Data-driven controllers can help address the above challenges. Various methods which utilize multi-modal sensory input [20]–[22], learn robust insertion policies from human demonstrations [8], [23] and generalize over object geometries [24]. Large-scale, high-fidelity simulation [9] can capture the wide distribution of contacts that may be encountered in the real world. The learning process can be accelerated when a well-defined curriculum is used for the RL training [25]. A related effort learns motion primitives for insertion [26]. Contact-rich data can be exploited by training with tactile stimuli [27], force/torque measurements [28], or a representation of extrinsic contacts like Neural Contact Fields (NCF) [29].

Integrated Control and ML Insertion Strategies The aforementioned learning-based techniques, however, exhibit high data requirements, especially for tight tolerances. To improve sample efficiency, prior work integrates RL and classical control by using impedance controllers for assembly tasks [11], [30]. Using RL to learn a residual policy given a model-based policy can improve sample complexity[31], [32]. Furthermore, these methods can work with demonstration data, dynamic movement primitives [33], [34] and contact-aware, compliant feedback-based controllers [35]. A key advantage of the proposed approach is that it trains entirely in simulation using a simple model-based policy, achieving a high success rate in real-world tight tolerance insertion tasks.

III. Method

The robot is tasked to insert a grasped object (plug) into a receptacle (socket) with sub-mm tolerance placed firmly in the workspace. At every timestep $t \in [0, T]$, the state is defined as s_t^P, where $s_t^P \in \mathtt{SE(3)}$ is the plug's pose at timestep t. Given the socket's (static) pose $s^S \in \mathtt{SE(3)}$, the goal pose for the plug so that it is fully inserted into the socket is denoted by s_G^P. The available observations $o_t = \{o_t^P, o^S\}$ correspond to continuous estimates $o_t^P \in \mathtt{SE(3)}$ of the plug's pose and an estimate $o^S \in \mathtt{SE(3)}$ of the socket's pose given visual input. The 3D object models of both the plug and socket are known at the time of execution and are denoted by Γ^P and Γ^S respectively. The objective is to train a policy $\pi(o_t)$, which, during inference, given an observation o_t, outputs an action $a \in \mathtt{SE(3)}$ that corresponds to transformations of the plug's pose so that it eventually reaches s_G^P.

Fig. 2 outlines the components of the proposed approach for computing policy $\pi(o_t)$: (i) a simple model-based policy outputs an action a_t^{PF} using a potential field – this is computed at every timestep during training and inference in simulation and reality for the target geometries Γ^P and Γ^S; (ii) a residual RL policy's action a_t^{RL} is added to the model-based policy's action to provide the final output action a_t^T; (iii) training is performed in simulation over randomized conditions to learn a_t^{RL} so that a_t^T results in successful insertions given sparse rewards and a curriculum; and, finally, (iv) the resulting policy $\pi(o_t)$ is transferred to the real system to solve tight insertion tasks involving novel geometries relative to those seen during training.

A. Model-based Policy

The potential-field action a_t^{PF} is computed given the plug and socket's observed poses (o_t^P, o^S) and their geometries (Γ^P, Γ^S). a_t^{PF} is a combination of an action arising from an attractive potential a_t^{Att}, i.e., moving the plug towards the goal s_G^P, and an action arising from a repulsive potential a_t^{Rep}, i.e., pushing the plug away from collisions with the socket.

Attractive Component A nominal, collision-free path for the plug is defined to connect the goal plug pose s_G^P to a pose with the same orientation above the socket along a straight retraction path, as in Fig. 3. This nominal retraction path is along the socket's medial axis, i.e., the locus of equidistant points from the socket's inner walls. Then, k *anchor* poses

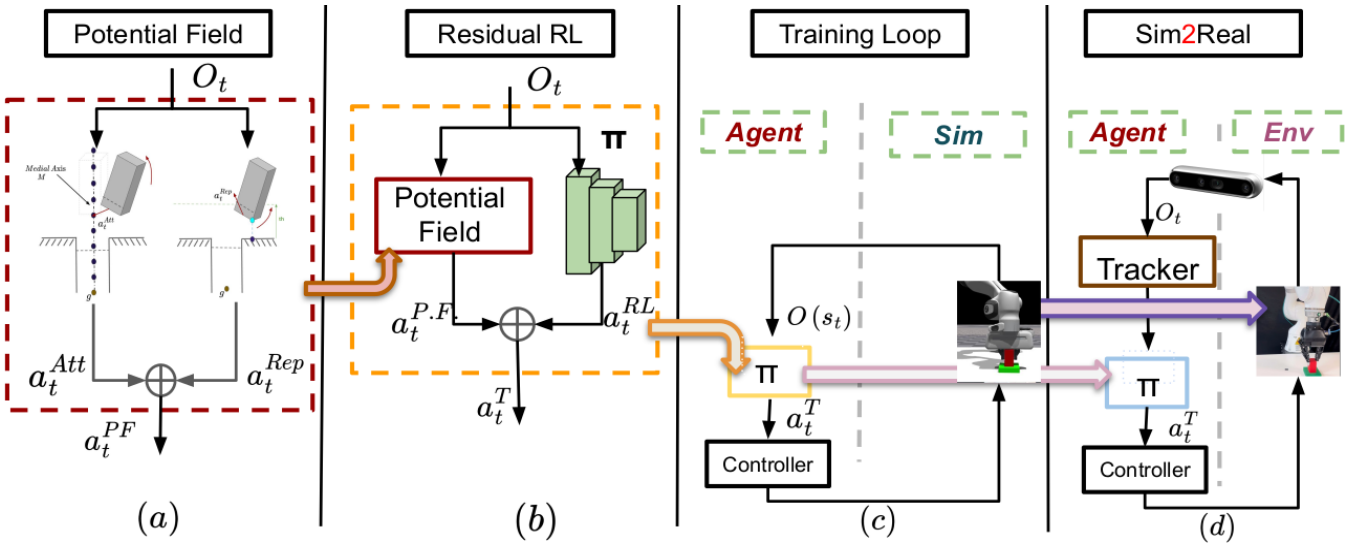


Fig. 2: From left to right: (a) A model-based policy is defined that generates a vector field under full observability. (b-c) An RL policy is trained in simulation given noisy pose observations to provide a residual action that is added to the output of the model-based policy. A sparse reward is provided only upon successful insertion (d) The final policy π is zero-shot transferred to the real world, where observations come from a pose tracking module given RGB-D data. A controller translates the policy into robot joint controls.

are defined along the nominal path by discretizing it. For every possible observation of the plug's pose o_t^P, the attractive potential computes the closest anchor pose on the nominal path s_{cl}. If the distance between o_t^P and s_{cl} is above a threshold, then the attractive potential returns $a_t^{\text{Att}} = s_{cl} - o_t^P$. If the distance s_{cl} is below a threshold, then the anchor pose s_{next} along the nominal path that is closer to the goal than s_{cl} is selected as the target. In this case, the attractive potential returns an action vector $a_t^{\text{Att}} = s_{next} - o_t^P$. Thus, the attractive field points towards the nominal path far from it and points more towards the goal pose close to the nominal path.

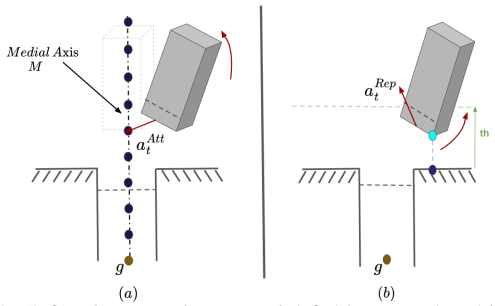


Fig. 3: (left) The attractive potential field moves the object towards a nominal, straight-line insertion trajectory that leads to the goal pose; (right) the repulsive component pushes the object away from making contact with the socket, only when the object is close to it.

Repulsive Component The closest pair of points on the plug p_t^P and the socket p_t^S are identified. If their distance $d_t = ||p_t^P - p_t^S||$ falls below a threshold th, a repulsive action is applied at the plug geometry's origin (which here is defined to be the plug's bottom-center point), to move the plug away from the socket. Otherwise, i.e., when $d_t > th$, the repulsive component is zero. To compute the repulsive action, a virtual 3D force vector $v_t = p_t^P - p_t^S$ is computed. A distance-normalized version of the virtual force vector v_t is defined as $\mathcal{N}(v_t)$ and corresponds to a division of the vector's magnitude with the distance d_t. This has the effect that the magnitude of $\mathcal{N}(v_t)$ increases as the peg approaches the socket. Then, the repulsive action is computed as $a_t^{\text{Rep}} = \mathbf{J} \cdot \mathcal{N}(v_t)$, where $\mathbf{J}$ is the Jacobian matrix that relates the coordinates of p_t^P to the plug's frame. This component moves the plug away from contact states that prohibit task success in tight setups. Overall, its use reduces the need to carefully tune the hyperparameters of the Attractive Field.

Potential Field The overall action combines the attractive and repulsive actions with a weighted sum, where weights $w^{Tr.}$ and $w^{Rot.} \in [0,1]$ (same for all objects) are applied to the translational and rotational components.

B. Residual RL

The potential field actions succeed in insertion when the ground-truth poses of the plug and socket are available, e.g., in simulation. When these pose estimates are noisy, as in the real world, the efficacy of the potential field-based policy declines drastically (see Fig.4). To address this, complementary actions $a_t^{RL} \in \mathtt{SE(3)}$ are generated by a residual Deep RL policy and added to the potential field action. The RL policy accounts for uncertain estimations, enabling successful task completion. The combined action $a_t^T = a_t^{PF} + \beta a_t^{RL}$, where $\beta \in [0, 1]$ scales the contribution of the two action components.

Sparse Rewards The model-based policy enables the use of a sparse goal-reaching reward for training the residual Deep RL component. A fixed positive reward is provided if the plug is fully inserted into the socket. In addition, a negative reward is defined for object-object inter-penetration during contact as IssacGym allows for significant inter-penetration between objects as noted in IndustReal [9]. This discourages RL from exploiting the simulation during training.

Scaling and Noise Curriculum Training While training the RL policy in simulation, uniform noise is added to the ground-truth poses of both the plug and the socket. A

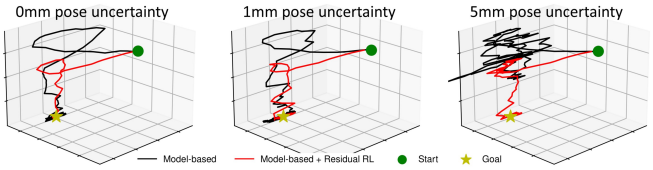


Fig. 4: Impact of observation noise on insertion trajectories: (left) Under no noise, both the model-based controller with and without the residual policy succeed. Residual RL helps to shorten trajectories. (center) With low observation noise, the performance of the model-based controller declines, but in combination with the residual policy output, the performance is preserved. (right) At high levels of noise, the model-based controller fails, while integrating the residual RL policy effectively compensates for the noisy pose estimate.

curriculum training strategy is implemented, where the plug observation noise ranges from 0mm/0° to n_{max}mm/n°_{max}, while the scaling parameter β ranges from 0 to 1. This curriculum adapts the difficulty of the training task by observing the success rate of insertion across multiple trials. It increases the difficulty if the success rate exceeds 75% and decreases it if the success rate falls below 50%. The noise ranges increment or decrement by st respectively. This adaptive approach trains the Deep RL policy to increase its contribution towards the combined action as observation uncertainty increases. During inference, the scaling parameter β is set to 1.

RL architecture An asymmetric actor-critic architecture is employed [36]. The actor network consists of a 3-layer Multi-Layer Perceptron (MLP) and a 2-layer Long Short-Term Memory (LSTM). The critic network consists of a 3-layer MLP. Both networks input the estimated plug and socket poses, and the critic additionally receives their ground-truth poses as privileged information. The architecture is trained with Proximal Policy Optimization (PPO) [37].

Training Randomization and Noise Conditions During training, the socket pose is randomized within a range of ± 10 cm in the x-y plane, 5 cm in the z-axis, and $\pm 5°$ in yaw. The initial plug pose varies within ± 10 mm in the x-y plane and $\pm 15°$ across roll, yaw, and pitch, while being positioned 10 mm above the socket's tip. To simulate uncertain observations, the 6D observation noise for both the plug and the socket is sampled i.i.d at each time-step from a uniform distribution. This noise is constrained to a maximum of ± 5 mm/$5°$ for the plug and ± 1 mm/$1°$ for the socket. The noise curriculum step-size is set to $st = 0.1$mm.

C. Sim2Real Transfer and Real-World Components

The policy was trained using IsaacGym [12] on a model of the Franka Emika Panda robot with a task impedance controller. The policy was then deployed zero-shot in the real world on a Kuka iiwa 14 manipulator with a Robotiq 3-fingered gripper, that uses a position controller. The successful transfer between the disparate simulation and real-world setup is facilitated by the definition of actions over the plug's `SE(3)` pose space.

RGB-D Pose Tracker and Pose Control M3T [38], an RGB-D-based pose tracker, provides an estimate of the socket's pose at the beginning of each trial as well as dynamically tracking the plug's pose across the trial (at a frequency of 30Hz). Tracking accuracy reduces as the plug nears and engages with the socket, due to increased occlusions. The combination of the deployed tracker with the proposed approach, results in a high insertion ratio. Task failures due to object-gripper slippage [9] are addressed as the policy reasons about the `SE(3)` pose of the plug and socket in a closed-loop manner.

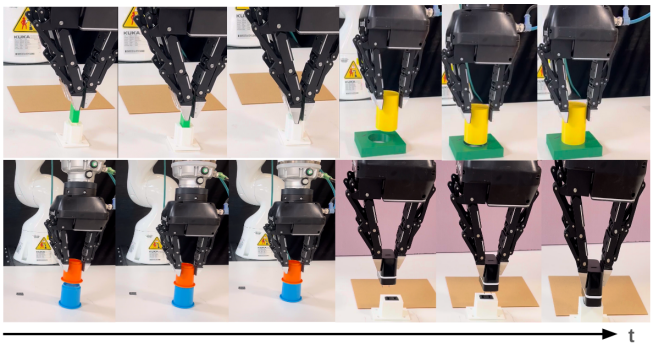


Fig. 5: Sim2Real policy transfer with 2 known (top) and 2 unknown at training objects (bottom).

IV. Results

Evaluated Object Categories The evaluation is performed across three plug-socket categories. The first category (Fig.6 (left)) includes small cylindrical and rectangular plugs of widths 8, 12, and 16 mm with tolerances of approximately $0.5-0.6$ mm, similar to NIST Taskboard Challenge benchmark [39] used for evaluation by IndustReal [9]. The second category (Fig.6 (middle)) encompasses larger cylindrical and rectangular sockets of 50mm width categorized into three difficulty levels based on their tolerance: Easy ($\sim$2 mm), Medium ($\sim$1 mm), and Hard ($\sim$0.1 mm). The last category of objects (Fig.6 (right)) is only used for real-world trials, and corresponds to five household objects that have not been seen during training: a 2-prong charger, a 3-prong charger, a HAN-type connector, two types of cups, and a marker with a marker holder.

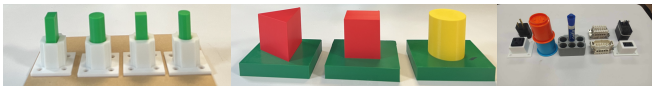


Fig. 6: (left) 3D printed objects from IndustReal [9] with $0.5-0.6$ mm tolerance. (middle) 3D printed custom objects with 2 mm (Easy), 1 mm (Medium), and 0.1 mm (Hard) tolerance. (right) Household objects not seen during training.

Evaluation in IsaacGym Table I evaluates the proposed method and the alternatives in simulation. The evaluation metric is the percentage of successful insertions for different levels of maximum perception noise n_{max}. Noise is sampled uniformly around the true states up to the value n_{max}. Three different values of n_{max} are considered for plug translational/rotational noise respectively: 0mm/$0°$, 1mm/$1°$, and 5mm/$5°$. For all scenarios with non-zero plug observation noise, a corresponding noise of 1mm/$1°$ was added to the socket.

The proposed method is compared against IndustReal [9], which also zero-shot transfers from Sim2Real but is an

TABLE I: Insertion Success Rates in Simulation

Method	Fig. 6 (left) Objects			Fig. 6 (middle) Objects		
	0mm/0°	1mm/1°	5mm/5°	0mm/0°	1mm/1°	5mm/5°
IndustReal [9]	92.40±2.30%	88.60±2.41%	n.a.	26.77±13.88%	27.09±13.61%	n.a.
PF + Res.RL + Curr. of [11]	98.65±0.87%	**98.44±0.55%**	**97.50±0.65%**	95.28±3.39%	92.36±4.35%	33.87±6.66%
Ours	**100.0±0.0%**	96.10±1.92%	96.25±1.22%	**99.09±0.91%**	**98.28±0.65%**	**95.88±1.86%**

exclusively RL approach. The code for AutoMate [24], an extension of IndustReal was not available while carrying out this evaluation. The proposed approach is also evaluated when using an alternative curriculum that transitions from model-based control to residual RL over time, instead of as a function of noise [11].

While evaluating IndustReal, observation noise of 1mm/1° is added only to the socket pose, as IndustReal operates over the `SE(3)` pose of the end-effector, whereas the proposed method operates over the `SE(3)` pose of the plug and socket. Thus, for IndustReal, the results reported with 1mm/1° noise are taken directly from the publication. IndustReal could not be evaluated with 5mm/5° plug noise scenario as applying high noise to the socket's pose artificially collapses its performance.

For the objects in Fig. 6 (left), a single policy was trained across all objects for a fair comparison with IndustReal. For the objects in Fig. 6 (middle), however, a dedicated policy was trained for each object instance to prevent the easier geometries from inflating the success rates while inserting the more challenging objects. The evaluation task is to insert these objects with the smallest tolerance. Similar to IndustReal, all policies are trained and tested over 5 random seeds and the mean and standard deviation of the insertion successes are reported.

IndustReal does not rely on 3D models of the plug and socket, whereas the proposed method requires them. To ensure a fair comparison, the plug and socket are approximated by their largest common bounding shape primitive (box, cylinder, etc.) This approximation allows the proposed method to operate without relying on specific instance 3D models while inserting the objects in Fig.6 (left), (Table I - left). The Potential Field (PF) policy generates the same actions across all geometric instances, using a 3D bounding box that approximates actual models. Since IndustReal trains a single policy across all these objects, this adjustment ensures alignment in model requirements for a fair comparison. Given this setup, the residual RL component should also compensate for the lack of a known 3D object model. This approximation applies only to the PF controller, in simulation, where no pose estimation is required. During Sim2Real transfer, the trackeruses the full 3D object models.

Overall, the proposed method consistently outperforms IndustReal in simulation. The time-based curriculum [11] also achieves high insertion success percentages for the objects of Fig.6(left), which verifies the efficacy of the designed model-based controller. The proposed success-based curriculum strategy surpasses the time-based curriculum for all for the objects of Fig. 6(middle), and a single object of Fig. 6 (left). As the difficulty of the task increases, it is observed that the difference in performance between the proposed method and the comparison points becomes more evident.

Note that the proposed residual RL policy requires only 25% to 33% of the training time/samples reported in the IndustReal work, without access to a handcrafted dense reward function. Our policy was fully trained in 2-3 hours using a single GPU, while IndustReal reports a corresponding time of 8-10 hours.

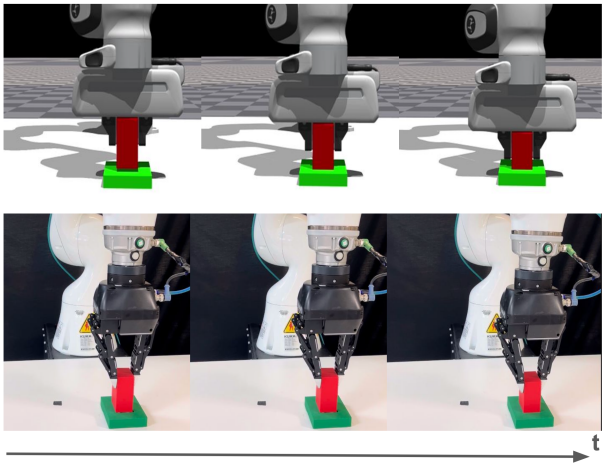


Fig. 7: Plug insertion in simulation (top) and real world (bottom).

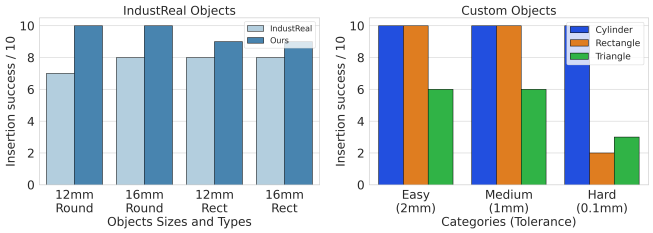
Ablation over Residual RL strategies: Table II evaluates variations in implementation of the proposed approach with progressively increasing $n_{\max}$ values to gauge the impact of noise levels. The first variation corresponds to applying only the Potential Field (PF) policy. Then, an alternative approach is tested where RL is used to learn the weights $w^{Tr.}$ and $w^{Rot.}$ that combine the attractive and repulsive components of the PF. Following this, variations of the proposed method are evaluated with and without the proposed success-based curriculum strategy. All variations of the proposed approach, where residual actions are output by the deep RL module, achieve high insertion rates for the tasks of Fig.6 (left). The full proposed method is the most successful while attemtping the harder insertion task of objects in Fig.6 (middle). Simultaneously applying the noise-action curriculum ensures that for every $n_{\max}$ value, the RL succeeds with the true pose as observations before progressing to more difficult task conditions. Combining this with the proposed success-based curriculum accelerates convergence to a insertion policy that achieves higher success rates than non-curriculum-based training schemes.

Real-world Experimental Setup: Experiments were conducted with the Kuka iiwa14 7-DOF manipulator in a minimally instrumented environment. An Intel RealSense D435 RGB-D camera was utilized to monitor the scene, capturing the manipulator as it reached for, grasped, and positioned a

TABLE II: Ablation Study

Method	Fig. 6 (left) Objects			Fig. 6 (middle) Objects		
	0mm/0°	1mm/1°	5mm/5°	0mm/0°	1mm/1°	5mm/5°
PF	98.91±0.89%	99.84±0.35%	3.28±1.69%	97.55±1.63%	97.81±1.58%	46.51±4.28%
PF + Learned Scaling w	90.47±3.51%	94.69±2.02%	11.72±2.45 %	89.21±2.44%	90.60±3.28%	7.76±3.71%
PF + Res.RL	99.53±0.70%	**100.0±0.00%**	**97.66±1.56%**	82.37±4.48%	63.48±5.52%	61.09±45.18%
PF + Res.RL + Learned Scaling β	98.28±0.65%	82.03±2.27%	84.69±4.12%	93.90±1.50%	92.34±2.67%	70.15±2.23%
Ours	**100.0±0.00%**	96.10±1.92%	96.25±1.22%	**99.09±0.91%**	**98.28±0.65%**	**95.88±1.86%**

target object at an randomized initial pose, 10mm above the socket. The control actions from policy then take over until the object is inserted (or time limits exceed). The socket was rigidly affixed to the table with heavy-load mounting tape to immobilize it.



(a) Comparison with IndustReal (b) Custom objects

Fig. 8: Real-world evaluation: Number of successful insertions over 10 trials for different objects.

Previously Seen Objects: Comparison with IndustReal in the real-world is shown in Fig. 8 (left). The statistics for IndustReal are derived from the corresponding publication. IndustReal achieves sim-to-real transfer for the same robot model, with a low-level task impedance controller and an instrumented setup. The proposed method achieves higher success rate across objects with less instrumentation and while training over a disparate setup in simulation.

Additional sim-to-real-world trials were performed on the more challenging objects of Fig. 6 (middle). The policy was trained only on the Hard objects in simulation and then evaluated in the real world on the Easy, Medium and Hard objects. Fig. 8 (right) shows that the proposed policy successfully inserted all instances of the Cylinder effectively (10/10 successful trials). Similar results are shown for the Easy and Medium Rectangular objects, while for the Hard case of $0.1mm$ tolerance, the recorded success rate drops. The Triangular objects were the most challenging with recorded success rates of 6/10, 6/10, 3/10 for the Easy, Medium, and Hardß cases respectively. Rotational alignment of the plugs with respect to the socket is a crucial factor for successful insertions in very low tolerance regimes.

Unseen household objects: The proposed method was also tested on 5 unseen real-world objects (Fig. 6 right). The policy was trained on the objects of Fig. 6 (left) and then tested directly on the household objects, starting from 10 randomized initial plug poses for each object. Fig. 9 shows that the proposed policy consistently achieved high success percentages for all test objects, demonstrating robust performance even in scenarios requiring significant force modulation (e.g., 2-prong, 3-prong, HAN-connector). These results show good generalization to novel geometries.

Object	**IndustReal**	**Ours**
2-Prong Charger	10/10	10/10
3-Prong Charger	7/10	10/10
Cups	-	10/10
Marker	-	10/10
HAN Connector	-	9/10

Fig. 9: Success rate for unseen household objects. Number of successful insertions over 10 real-world trials each.

V. Discussion

This paper proposes a hybrid approach for robotic insertion, which leverages the strengths of both model-based planning and data-driven methods, using a potential field as a guiding policy that works well in noise-free scenarios. RL enables the system to adapt to noise without requiring complex reward engineering. The policy is trained exclusively in simulation using sparse rewards and transferred zero-shot to the real world with good accuracy.

While in industrial setup 3D models may be available, the reliance on 3D models during inference may limit the method's applicability in service robotics. Future work will explore how to waive this requirement, while still achieving high accuracy and benefiting from model-based reasoning. Furthermore, occlusions prevented tests on small objects, indicating the need for fine-grained sensing.

Lastly, this work aims to inform how to solve general contact-rich manipulation tasks with tight tolerances, e.g. top-down insertion, while minimizing human engineering. As IndustReal demonstrated, for top-down insertion it is possible to design useful dense rewards. But it is not obvious how to define dense rewards that work across contact-rich manipulation tasks. The current effort indicates that the combination of a model-based policy and sparse reward residual RL can provide solutions in this domain. For other manipulation tasks, the model-based policy may be defined after a classical motion planner first generates successful (more complex) manipulation solutions under full observability. These model-based policies can still guide residual RL policies that work under partial observability and noise in the real-world.

REFERENCES

[1] T. Lozano-Perez, M. T. Mason, and R. H. Taylor, "Automatic synthesis of fine-motion strategies for robots," *The International Journal of Robotics Research*, vol. 3, no. 1, pp. 3–24, 1984. [Online]. Available: https://journals.sagepub.com/doi/10.1177/027836498400300101

[2] F. Suárez-Ruiz and Q.-C. Pham, "A framework for fine robotic assembly," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 421–426. [Online]. Available: https://ieeexplore.ieee.org/document/7487162

[3] A. S. Morgan, B. Wen, J. Liang, A. Boularias, A. M. Dollar, and K. Bekris, "Vision-driven compliant manipulation for reliable, high-precision assembly tasks," *arXiv preprint arXiv:2106.14070*, 2021. [Online]. Available: https://arxiv.org/abs/2106.14070

[4] S. Wang, G. Chen, H. Xu, and Z. Wang, "A robotic peg-in-hole assembly strategy based on variable compliance center," *IEEE Access*, vol. 7, pp. 167 534–167 546, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8906066

[5] A. S. Morgan, Q. Bateux, M. Hao, and A. M. Dollar, "Towards generalized robot assembly through compliance-enabled contact formations," 2023. [Online]. Available: https://arxiv.org/abs/2303.05565

[6] Y. Wu, F. Wu, L. Chen, K. Chen, S. Schneider, L. Johannsmeier, Z. Bing, F. Abu-Dakka, A. Knoll, and S. Haddadin, "1 khz behavior tree for self-adaptable tactile insertion." [Online]. Available: https://www.researchgate.net/publication/377983521_1_kHz_Behavior_Tree_for_Self-adaptable_Tactile_Insertion

[7] B. Wen, W. Lian, K. Bekris, and S. Schaal, "You only demonstrate once: Category-level manipulation from single visual demonstration," 2022. [Online]. Available: https://arxiv.org/abs/2201.12716

[8] A. Nair, B. Zhu, G. Narayanan, E. Solowjow, and S. Levine, "Learning on the job: Self-rewarding offline-to-online finetuning for industrial insertion of novel connectors from vision," 2023. [Online]. Available: https://arxiv.org/abs/2210.15206

[9] B. Tang, M. A. Lin, I. Akinola, A. Handa, G. S. Sukhatme, F. Ramos, D. Fox, and Y. Narang, "Industreal: Transferring contact-rich assembly tasks from simulation to reality," 2023. [Online]. Available: https://arxiv.org/abs/2305.17110

[10] J. Luo, Z. Hu, C. Xu, Y. L. Tan, J. Berg, A. Sharma, S. Schaal, C. Finn, A. Gupta, and S. Levine, "Serl: A software suite for sample-efficient robotic reinforcement learning," *ArXiv*, vol. abs/2401.16013, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267311834

[11] P. Kulkarni, J. Kober, R. Babuška, and C. Della Santina, "Learning assembly tasks in a few minutes by combining impedance control and residual recurrent reinforcement learning," *Advanced Intelligent Systems*, vol. 4, no. 1, p. 2100095, 2022.

[12] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021. [Online]. Available: https://arxiv.org/abs/2108.10470

[13] Y. Tian, J. Xu, Y. Li, J. Luo, S. Sueda, H. Li, K. D. Willis, and W. Matusik, "Assemble them all: Physics-based planning for generalizable assembly by disassembly," *ACM Trans. Graph.*, vol. 41, no. 6, 2022. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3550454.3555525

[14] R. Haugaard, J. Langaa, C. Sloth, and A. Buch, "Fast robust peg-in-hole insertion with continuous visual servoing," in *Conference on Robot Learning*. PMLR, 2021, pp. 1696–1705. [Online]. Available: https://proceedings.mlr.press/v155/haugaard21a.html

[15] Y. Liu, Z. Chen, X. Zhang, and J. Gao, "Compliant peg-in-hole assembly for components with grooves based on attractive region in environment," in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2021, pp. 919–924. [Online]. Available: https://ieeexplore.ieee.org/document/9536140

[16] C. Wang, H. Luo, K. Zhang, H. Chen, J. Pan, and W. Zhang, "Pomdp-guided active force-based search for robotic insertion," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 10 668–10 675. [Online]. Available: https://ieeexplore.ieee.org/document/10342421

[17] H. Chen, J. Li, W. Wan, Z. Huang, and K. Harada, "Integrating combined task and motion planning with compliant control: Successfully conducting planned dual-arm assembly motion using compliant peg-in-hole control," *International Journal of Intelligent Robotics and Applications*, vol. 4, pp. 149–163, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s41315-020-00136-1

[18] S. R. Chhatpar and M. S. Branicky, "Search strategies for peg-in-hole assemblies with position uncertainty," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, vol. 3. IEEE, 2001, pp. 1465–1470. [Online]. Available: https://ieeexplore.ieee.org/document/977187

[19] H. Kang, Y. Zang, X. Wang, and Y. Chen, "Uncertainty-driven spiral trajectory for robotic peg-in-hole assembly," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6661–6668, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9780009

[20] O. Spector and D. Di Castro, "Insertionnet-a scalable solution for insertion," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5509–5516, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9420246

[21] O. Spector and D. D. Castro, "Insertionnet - a scalable solution for insertion," *IEEE Robotics and Automation Letters*, vol. 6, pp. 5509–5516, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233443845

[22] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Learning multimodal representations for contact-rich tasks," *CoRR*, vol. abs/1907.13098, 2019. [Online]. Available: http://arxiv.org/abs/1907.13098

[23] B. Wen, W. Lian, K. Bekris, and S. Schaal, "You only demonstrate once: Category-level manipulation from single visual demonstration," in *Robotics: Science and Systems (RSS)*, 2022. [Online]. Available: https://www.roboticsproceedings.org/rss18/p044.pdf

[24] B. Tang, I. Akinola, J. Xu, B. Wen, A. Handa, K. Van Wyk, D. Fox, G. S. Sukhatme, F. Ramos, and Y. Narang, "Automate: Specialist and generalist assembly policies over diverse geometries," in *Robotics: Science and Systems*, 2024.

[25] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, "Accelerating robot learning of contact-rich manipulations: A curriculum learning study," *arXiv preprint arXiv:2204.12844*, 2022. [Online]. Available: https://arxiv.org/abs/2204.12844

[26] X. Zhang, S. Jin, C. Wang, X. Zhu, and M. Tomizuka, "Learning insertion primitives with discrete-continuous hybrid action space for robotic assembly tasks," 2021. [Online]. Available: https://arxiv.org/abs/2110.12618

[27] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," 2021. [Online]. Available: https://arxiv.org/abs/2104.01167

[28] O. Azulay, M. Monastirsky, and A. Sintov, "Haptic-based and $se(3)$-aware object insertion using compliant hands," *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 208–215, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9963587

[29] C. Higuera, J. Ortiz, H. Qi, L. Pineda, B. Boots, and M. Mukadam, "Perceiving extrinsic contacts from touch improves learning insertion policies," *ArXiv*, vol. abs/2309.16652, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:263143002

[30] S. A. Khader, H. Yin, P. Falco, and D. Kragic, "Stability-guaranteed reinforcement learning for contact-rich manipulation," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 1–8, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9211756

[31] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine, "Residual reinforcement learning for robot control," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 6023–6029, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:54464184

[32] G. Schoettler, A. Nair, J. Luo, S. Bahl, J. A. Ojea, E. Solowjow, and S. Levine, "Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards," *CoRR*, vol. abs/1906.05841, 2019. [Online]. Available: http://arxiv.org/abs/1906.05841

[33] T. Davchev, K. S. Luck, M. Burke, F. Meier, S. Schaal, and S. Ramamoorthy, "Residual learning from demonstration: Adapting dmps for contact-rich manipulation," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:237500490

[34] J. Carvalho, D. Koert, M. Daniv, and J. Peters, "Adapting object-centric probabilistic movement primitives with residual reinforcement learning," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, 2022, pp. 405–412.

[35] S. Brahmbhatt, A. Deka, A. Spielberg, and M. Müller, "Zero-shot transfer of haptics-based object insertion policies," 2023. [Online]. Available: https://arxiv.org/abs/2301.12587

[36] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," *arXiv preprint arXiv:1710.06542*, 2017. [Online]. Available: https://roboticsproceedings.org/rss14/p08.pdf

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017. [Online]. Available: https://arxiv.org/abs/1707.06347

[38] M. Stoiber, M. Sundermeyer, W. Boerdijk, and R. Triebel, "A multi-body tracking framework-from rigid objects to kinematic structures," *arXiv preprint arXiv:2208.01502*, 2022. [Online]. Available: https://arxiv.org/abs/2208.01502

[39] K. Kimble, K. Van Wyk, J. Falco, E. Messina, Y. Sun, M. Shibata, W. Uemura, and Y. Yokokohji, "Benchmarking protocols for evaluating small parts robotic assembly systems," *IEEE robotics and automation letters*, vol. 5, no. 2, pp. 883–889, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8957300