Visuotactile-RL: Learning Multimodal Manipulation Policies with Deep Reinforcement Learning

Johanna Hansen, Francois Hogan, Dmitriy Rivkin, David Meger,
Michael Jenkin, Gregory Dudek

***Abstract*— Manipulating objects with dexterity requires timely feedback that simultaneously leverages the senses of vision and touch. In this paper, we focus on the problem setting where both visual and tactile sensors provide pixel-level feedback for Visuotactile reinforcement learning agents. We investigate the challenges associated with multimodal learning and propose several improvements to existing RL methods; including tactile gating, tactile data augmentation, and visual degradation. When compared with visual-only and tactile-only baselines, our Visuotactile-RL agents showcase (1) significant improvements in contact-rich tasks; (2) improved robustness to visual changes (lighting/camera view) in the workspace; and (3) resilience to physical changes in the task environment (weight/friction of objects).**

I. INTRODUCTION

The synergy between the senses of vision and touch is fundamental to the manner in which animals interact with the world around them. While vision is informative of an object's pose and general shape, the sense of touch provides accurate feedback on the location of contact, interaction forces, and the object's material properties. Despite the well understood importance of visual and tactile feedback in human manipulation [1], robotic systems struggle to integrate both modalities with the ease displayed by humans.

In this paper, we develop model-free deep reinforcement learning agents that learn to utilize high-resolution visual and tactile information for manipulation tasks. Reinforcement learning approaches have shown an impressive ability to learn expressible controllers for robotic manipulation [2]. Most techniques are developed to use visual data from a third-person view optical camera [2] or combine camera observations with low-resolution tactile sensing [3]. With the recent development of modern pixel-based tactile sensors such as Gelsight [4], GelSlim [5], Omnitact [6] and STS [7], there is an opportunity to provide robots with high-resolution touch feedback. Here, we consider a challenging triad: dexterous manipulation tasks; high-resolution visual and tactile sensors; and reinforcement learning.

In over 200 experiments, we illustrate the benefits of the *Visuotactile-RL* paradigm and investigate the challenges of this high-resolution, multimodal observation space. In addition to evaluating state-of-the-art reinforcement learning algorithms, we propose training techniques and network architectures to overcome the difficulties faced in this multi-faceted sensing regime.

All authors are affiliated with the Samsung AI Research Center Montreal, 1000 Sherbrooke Street W., Suite 1100, Montreal, QC, Canada.

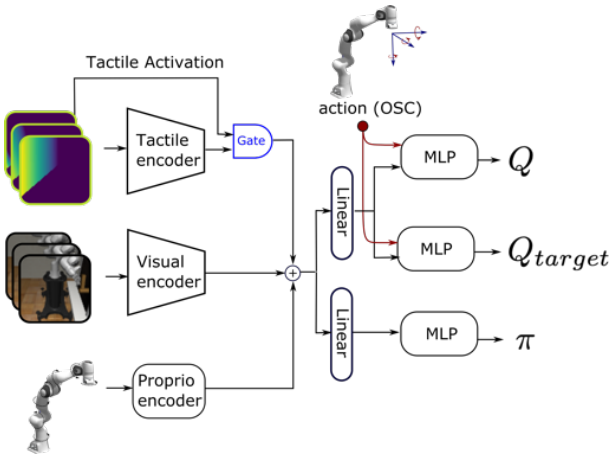
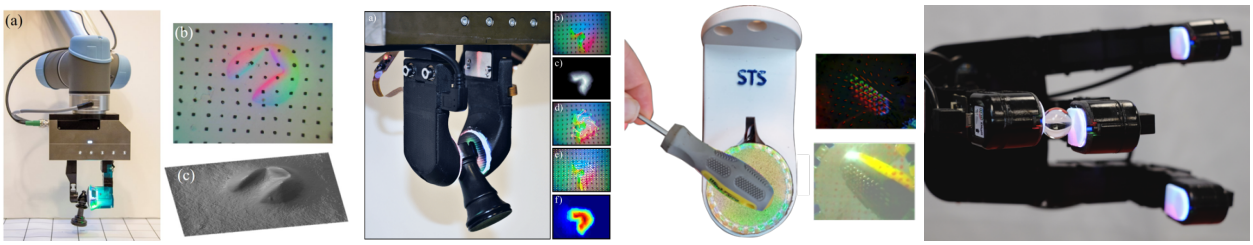


Fig. 1. **Visuotactile-RL network architecture diagram of *MP-DrQv2 with tactile gating*.** The network takes in raw visual, tactile, and proprioceptive observations and encodes them with modality specific modules. All training is done online with a TD error objective. We find that including a *tactile gate* to control the flow of tactile gradients through the network as a function of the contact state improves learning on tactile-rich tasks. The network's core is shared to infer the policy $\pi(\mathbf{a}_t|\mathbf{s}_t)$ and the Q-functions Q and Q_{target}, which include robot actions as input.

While visual sensing provides continuous feedback to the agent during interactions, tactile sensing only provides feedback when the sensor interacts with the environment physically. During contact, tactile sensing provides valuable information that is related to the interaction, namely the location, shape, and interaction forces at contact. For example, when opening a door, visual feedback drives the reaching phase, while our attention quickly shifts to tactile cues once contact is made with the handle to obtain more precise information on the moment of contact, the handle location, and the grasp stability.

The fundamentally discontinuous nature of physical interactions poses important challenges for policy learning methods. While this is acknowledged and well-studied within the model-based planning and control community [8], [9], [10], [11], it is often overlooked in reinforcement learning. Even as data-driven methods do not explicitly require reasoning over sensory discontinuities, they must cope with unbalanced datasets where only a small fraction of examples include tactile information, as illustrated in Fig. 4. This leads to a number of important questions regarding the applicability and effectiveness of reinforcement learning methods for policy learning. How do we prevent the agent from over-biasing its attention on visual feedback, which is more prevalent in



(a) GelSight (b) GelSlim (c) STS (d) DIGIT

Fig. 2. **Optical tactile sensors**. This class of high-resolution tactile sensors render pixel-level images that emphasize the contact geometry of objects as they interact with the robot finger. This paper investigates how best to exploit the type of feedback provided by these sensors within a reinforcement learning framework to learn expressive feedback policies.

the dataset? Do the discontinuities associated with contact interactions negatively impact learning stability?

While tactile-only feedback controllers have been showcased for tactile-centric tasks such as peg insertion [12], here we investigate a more general scenario where the object does not begin in contact with the desired location in the environment. In this paper, we investigate the limitations of existing RL algorithm for multimodal policy learning and propose novel perceptual architectures and training procedures to overcome them. The main contributions of this work are:

Analysis of Visuotactile-RL. We present an in-depth analysis of the performance of state-of-the-art reinforcement learning algorithms and various data augmentation and training strategies. We test a variety of perceptual frameworks and architectures to best fuse the visuotactile sensing modalities and present a numerical experiment study on three simulated manipulation robotic tasks in the Robosuite [13] simulation framework.

Multimodal Perception Architecture. We introduce *tactile gating*, a learning mechanism that addresses the intermittent nature of tactile feedback. A *tactile gate* in the tactile perceptual module prevents the flow of tactile feedback through the network in the absence of detected contact. We show that tactile gating can improve learning performance and help the agent better exploit tactile sensing when used in tandem with visual feedback on contact-rich tasks.

The long term objective of this research is to design control strategies for contact-rich robotic manipulation tasks that exploit multi-modal sensing. The rest of this paper is structured as follows. Section II reviews relevant work related to this manuscript. Section III provides background on Data Regularized Q (DrQ), a model-free reinforcement learning algorithm shown to be effective at learning low-level policies from image observations. In Section IV, we introduce *tactile gating*, a learning mechanism that controls the flow of tactile feedback through the agent's network shown to be effective for multi-modal contact-rich task. We present our experimental setup on three robotic manipulation simulation tasks in Section V and present our experimental results in Section VI.

II. RELATED WORK

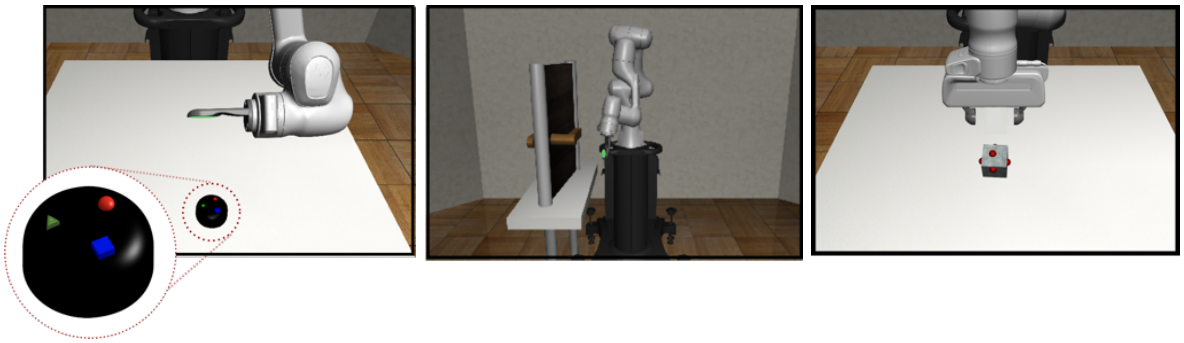
A. Tactile Sensing for Robotics

The sense of touch is a rich and critical source of feedback during robot manipulation. Traditional tactile sensing has included measurement of shape, texture and forces in various directions, among other key attributes [14], gathered using a wide variety of tactile measuring technologies [15]. A new generation of optical tactile sensing [4], [16], [5] employs cameras embedded in a compliant gel capable of imaging the contact surface at high resolution. These sensors capture the deformations of a reflective soft surface as it makes contact with the world. This enables high resolution reasoning about contact geometry as well as slip and contact forces.

Recently, there have been a number of works concerned with developing tactile policies that are able to exploit the rich information provided by optical tactile sensors for robotic manipulation. Tian et al. [17] develop a model-based tactile controller using pixel-level feedback to manipulate small objects. Hogan et al. [18] develop closed-loop tactile controllers are developed for a dual palm manipulation system able to manipulate objects with dexterity on a table top. Wang [19] showcases robotic system capable of swining up and stabilizing objects by using the rich feedback provided by optical tactile sensors to estimate the object's physical parameters. In Dong [20], model-free RL is used to develop a tactile policy to align and object and environment with a tactile-based feedback insertion policy. These studies focus on the development of tactile-only policies and do not integrate visual reasoning.

B. Visuotactile Manipulation

There are a number of recent works in the literature that explore how best to fuse visual and tactile feedback in the context of reinforcement learning. [21] shows that a Variational Autoencoder (VAE) perceptual architecture is effective to extract meaningful state representations from visual and tactile inputs on a simple manipulation task. This architecture is extended to multimodal control in [3] on a peg insertion task with visual and force-torque sensing. [22] use Proximal Policy Optimization (PPO) [23] with pixel-level visuotactile inputs to teach a simulated robot arm to perform several tactile-rich tasks in a simulation environment. Unlike the tasks that are the focus of this work, the tasks in [22]



(a) TactileReach (b) Door (c) TactileLift

Fig. 3. **Visuotactile-RL benchmark tasks**. We consider three robotic tasks implemented in the Robosuite simulation framework. In TactileReach, the task is to precisely make contact with one of the three textures (square, triangle, sphere). In Door, the task is to open a hinged door with a robotic palm. In TactileLift, the task is to grasp and raise an object with a robotic gripper to a minimum height.

feature sustained contact interactions between the surface of the tactile sensor and the manipulated object. Our work contrasts from these previous approaches by focusing on the learning performance in the presence intermittent tactile feedback, as well as evaluating the robustness of the learned policies to perturbations in physical and visual properties of the environment.

C. RL for high-dimensional inputs

There have been a number of recent advances in the development of RL approaches that learn policies directly from pixel-level feedback. A popular approach has been to extract information from image observations using learned models as in SLAC [24], SAC-AE [25], PlaNet [26] and Dreamer [27]. However, DrQ [28] showed that state-of-the-art performance could be achieved in a model-free setting with a Soft Actor Critic (SAC) agent by employing data augmentation. The Data Regularized Q (DrQ) learning approach demonstrates that image-based RL agents have a tendency to overfit to observed data. This paper was followed by DrQv2 which exchanges SAC for a DDPG learner and multi-step Q updates. The idea that diverse data improves learning and robustness, a well-known concept in the field of computer vision, is also employed in robotic sim2real tasks. Domain randomization (DR), where a simulated environment is randomized during agent training [29] is a common tool for improving robustness in sim2real.

III. Background

We consider a Markov Decision Process (MDP) defined by the tuple $(\mathcal{S}, \mathcal{A}, p, r)$, where $\mathcal{S}$ is the set of continuous states, $\mathcal{A}$ is the set of continuous actions, $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}^+$. represents the probability density of the next state $\mathbf{s}_{t+1} \in \mathcal{S}$ given the current state $\mathbf{s}_t \in \mathcal{S}$ and the current action $\mathbf{a}_t \in \mathcal{A}$. A stochastic policy is a mapping $\pi : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}^+$. The environment returns a reward $r : \mathcal{S} \times \mathcal{A} \mapsto [r_{min}, r_{max}]$ at every state transition. The objective of reinforcement learning is to find an optimal policy π^* from the set of admissible policies Π that maximizes the total returns given a reward function r.

$$\pi^* = \text{argmax}_{\pi \in \Pi} \sum_t \mathbb{E}_{(\mathbf{s}_t, \mathbf{a}_t) \sim \rho_\pi} \left[r(\mathbf{s}_t, \mathbf{a}_t) \right]. \tag{1}$$

A. DrQv2

DrQv2 [2] is an off-policy actor-critic RL algorithm that efficiently learns a policy directly from pixels without a model. Due to its recent success in learning robotic control from images, we employ this approach as the basis for our Visuotactile-RL agents. DrQv2 yields state-of-the-art performance by using image perturbations to regularize the value function. It introduces an optimality-invariant state transformation $f : \mathcal{S} \times \mathcal{T} \rightarrow \mathcal{S}$ as a state mapping that preserves the Q-function

$$Q(\mathbf{s}, \mathbf{a}) = Q(f(\mathbf{s}, \boldsymbol{\nu}), \mathbf{a}) \forall \mathbf{s} \in \mathcal{S}, \mathbf{a} \in \mathcal{A} \text{ and } \boldsymbol{\nu} \in \mathcal{T}, \tag{2}$$

where $\mathcal{T}$ denotes the set of state transformations.

Adapting the DDPG algorithm, DrQv2 incorporates n-step returns, employs a decaying schedule for exploration noise, and computes the Q-function over image observations which undergo a randomly sampled image shift during training.

As DrQv2 owes much of its high performance to the augmentation of images during training, it is natural to ask whether this technique will translate from the stationary visual setting to the tactile paradigm. For instance, DrQv2 performs a random shift of the input image, which is a common data augmentation computer vision pipelines, but may not be applicable to visuotactile observations. We know that this operation will not preserve the underlying state of the tactile observation as it introduces an effective relative position shift between the observed scene and the tactile sensor. We further explore this concept and test the idea of **tactile augmentation** in Section VI.

IV. Methodology

The goal of this paper is to investigate the challenges of visuotactile-RL in order to develop agents capable of robustly fusing the senses of vision and touch for robotic manipulation. We focus on the problem setting where visual information is provided in the form of an RGB image and where tactile feedback is provided in the form of pixel-level

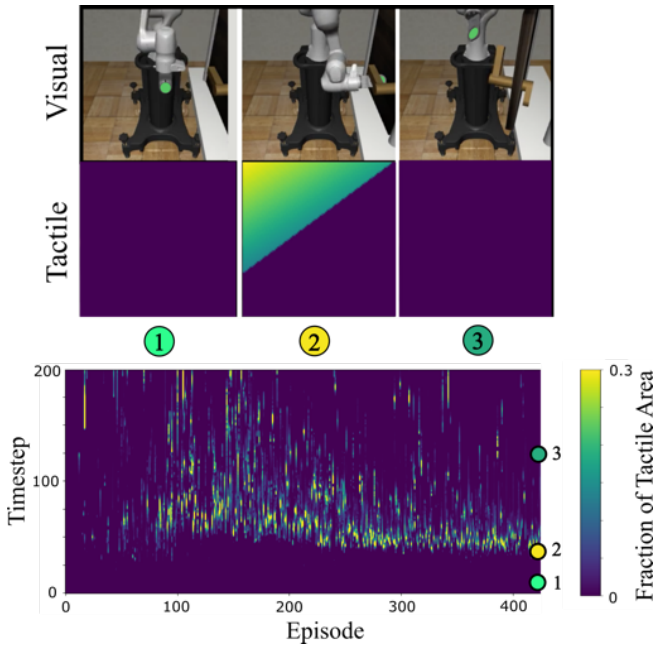


Fig. 4. **Tactile interactions vs. learning experience**. Tactile sensing provides intermittent feedback as the sensor interacts with the environment. As the agent learns to interact with the environment during a door opening task, it explores the door handle in three stages. Under 100 episodes, it makes very few tactile interactions resulting in sparse tactile observations. From 100 to 300 episodes, it obtains rich tactile observations as it is learning to turn the handle. Once the behavior is learned (around 300 episodes), the agent only receives tactile feedback during the handle turning phase that occurs around step 50.

measurements, as is typical for novel optical tactile sensors shown in Fig. 2.

One of the biggest challenges in achieving robust multimodal policies is that the tactile signal can be difficult to exploit when combined with visual feedback. We hypothesize that this occurs due to an unbalanced dataset where only a fraction of examples include tactile information. The distribution of tactile examples is also non-uniform over training. We illustrate this data imbalance problem in Fig. 4, which shows the intermittent feedback provided by tactile measurements as the sensor interacts with the environment. Early in the training process, the agent makes infrequent tactile encounters, but, as learning progresses and the tactile sensor is activated more often due to successful manipulation, the model must now deal with a newly useful tactile modality.

Perhaps as a result of the intermittent tactile signal, we found that baseline models become overly reliant on visual information and ignore the tactile input in tasks which do not explicitly require tactile (Door and TactileLift). In order to evaluate the capability of multimodal agents to utilize either modality, we propose the use of domain randomization to test the ability of agents to exploit either vision or touch. Our evaluation process will involve testing agents in a setting in which one of the modalities has been altered from the training environment and comparing to a baseline agent which is specifically trained under domain randomization.

A. *Tactile Gating*

Inspired by Long Short Term Memory recurrent neural networks [30] and Highway Networks [31], we introduce a gating mechanism that dynamically controls the flow of the information to the agent state at each time step based on the usefulness of the tactile signal. This technique, which we refer to as **tactile gating**, utilizes a hard gate which is activated during contact, detected by monitoring the depth image from the tactile observation. The gate remains closed when there is not tactile activation to prevent gradient propagation to the tactile encoder.

B. *Visual Degradation*

We also investigate whether degrading the visual signal can improve multimodal performance. Motivated by the incentive to better exploit tactile feedback, during training, we reduce the quality of the visual measurements to encourage the system to focus its attention on tactile cues. In the **image dropout** training paradigm, the signal from the camera is randomly removed from the observation for a fraction of the interactions.

V. Experimental Setup

We investigate Visuotactile-RL on a suite of tasks where the agent must learn to exploit visual information to find and establish contact with an object and then use its tactile sensor to interact with it.

Experiments are performed in the Robosuite [13] simulation framework, which uses MuJoCo [32] as a physics engine. In all scenarios, the RL agent controls the agent using an Operation Space Controller (OSC) [33] operating at 20 Hz on end-effector pose. Our results are reported on the Panda robot arm which uses proprioceptive feedback as well at least one other pixel-based sensing modality to complete the task.

We simulate the output of the tactile sensor by rendering the contact geometry relative to the perspective of the robot manipulator. While there are a number of available simulators for optical based tactile sensors such as TACTO [34], Tactile-Gym: RL [22], and Geometric Contact Rendering [35]. We use Geometric Contact Rendering as detailed in [35] to simulate the tactile imprint. This technique consists in clipping the depth image obtained from the perspective of the robot's fingertip to a threshold value corresponding to the half width of the silicone membrane. The main motivation to use this technique is that it results in faster simulation speeds that translate to lower agent training times. Note that there is a well established procedure based on photometric stereo that allows to reconstruct the depth image from the raw tactile imprint from optical tactile sensors, as depicted in Fig. I.

In order to quantify agent robustness to physical changes in the environment, we employ domain randomization on physical dynamics (friction, weight, etc) for objects in the scene. This environment setup is referred to as **DR Dynamics**, as shown in Table I, and is employed in some training experiments and for evaluation in relevant tasks.

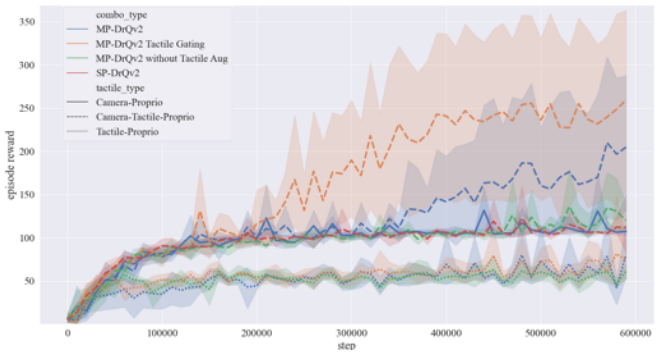


Fig. 5. **Tactile Gating improves learning speed on tactile-critical tasks.** This plot demonstrates the evaluation reward by training step for TactileReach. Tactile gating (orange) significantly improves the speed at which the agent learns to solve this visuotactile task when compared to the multimodal baseline, MP-DrQv2 (blue).

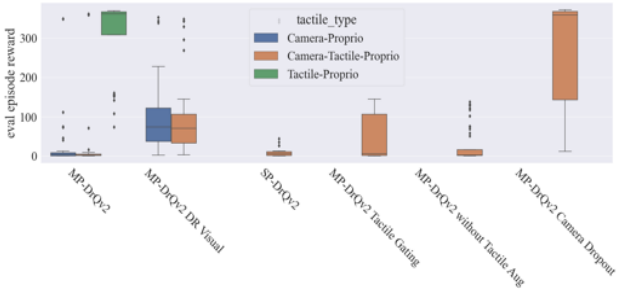


Fig. 6. **Agents that exploit tactile information perform well under strong visual changes compared to visual-only agents.** This plot depicts the evaluation of Door agents under Visual DR, testing the scenario where agents must act under unusual visual perturbations. The visual-only agent completely fails in this scenario, while the tactile-only agent remains robust as it does not observe the randomization. Comparing the multimodal models, we find that Tactile Gating performs similar to the baseline (MP-DrQv2) without requiring intentional degradation of the visual sensor during training (DR Visual and Camera Dropout).

High performance under DR Dynamics suggests the agent may be capable of utilizing the tactile sensor for feedback when changes in the environment are not obvious in the visual sensor. Our evaluations are done on all objects in the scene aside from the robot itself. We also test the agents under **DR Visual**, which perturbs the lighting conditions and camera location of the visual image. This randomization significantly alters the viewpoint of the camera, often causing the object of interest to be absent from the scene. We employ the default Robosuite DR wrapper for randomization, and sample a new environment configuration for each episode.

A. Tasks

We focus on three simulated robot tasks: TactileReach, Door, and TactileLift (see Figure 3). Each task is evaluated on three sensor combinations: 1) camera-proprio, 2) camera-tactile-proprio, and 3) tactile-proprio, where camera denotes a third person view on the environment, tactile refers to an image-based imprint of the contact, and proprio is the position and velocity of the robot joints.

1) TactileReach: In TactileReach, shown in Fig. II-A, the agent is tasked with touching a tactile feature on the surface of the cylinder in the presence of two distracting tactile features with a palm end-effector. We design this task to test the performance of visuotactile controllers across both modalities. The robot must use vision to reach out from a starting position to the cylinder, which is randomly initialized on the workspace. Since the tactile features are not observable by the camera, tactile feedback is necessary to intentionally align the palm with the target tactile feature and receive full reward. The reward schedule is the same as the Reach component of the Robosuite Lift task [13], with the exception that success is defined as the visuotactile sensor making precise contact with the target texture.

2) Door: We consider a the standard Robosuite door opening task, but outfit the robot arm with a palm end-effector. The task, shown in Fig. II-A, requires the agent to turn an articulated handle to open a door with a randomly generated door frame position and rotation offset.

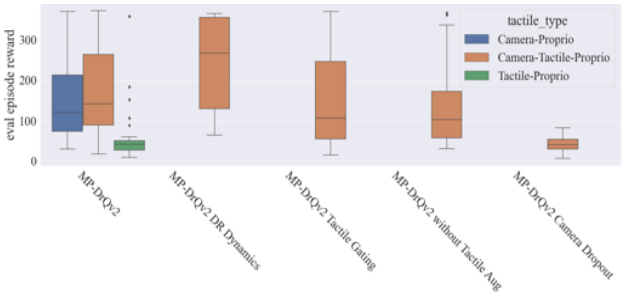


Fig. 7. **Multimodal policies improve robustness to dynamics changes in TactileLift (weight, friction of the box).** We find that tactile is critical to solving this evaluation, with multimodal agents performing better than visual-only or tactile-only policies when faced with randomized dynamics. The agent which was trained on domain randomization performs best overall, but Tactile Gating produced the single highest performing agent on this evaluation as shown in Table I.

3) TactileLift: The TactileLift task, shown in Fig. II-A, is adapted from the standard Robosuite Lift task to test tactile sensing. This task requires the agent to successfully grasp and raise a randomly positioned box to a minimum height. We equip the robot with a parallel jaw gripper which has two tactile sensors. Images from the two tactile sensors are stacked on the channel axis and treated as one observation. At the start of each episode, we randomly generate small spherical protrusions on the surface of the box. The spheres are made to be tactile obstacles by reducing the simulated friction on their surface so as to make grasping the box more difficult.

VI. Experiments

In this section, we demonstrate that Visuotactile-RL is powerful in scenarios involving 1) rich contact interactions 2) visual randomizations, and 3) perturbed dynamic parameters.

All RL experiments are performed using the default DrQv2 hyperparameters [2]. We change the size of the replay buffer 600,000 to accommodate the longer training time needed to learn manipulation. In addition to our analysis using DrQv2, we develop and evaluate a pixel-based variant

TABLE I

RESULTS ON 5 TASKS REPORTED AS THE MEAN AND STANDARD DEVIATION (MEAN±STD) IN EVALUATION ON EXPERIMENTS DESCRIBED IN EACH COLUMN. **GRAY** ROWS INDICATE PERFORMANCE IN 5 SEEDS FOR 5 EVALUATION EPISODES AFTER 500000 TRAINING STEPS. WE ALSO EVALUATE THE BEST SEED FROM EACH EXPERIMENT OVER 10 EPISODES IN THE FOLLOWING ENVIRONMENTS RELATIVE TO THE TRAIN ENVIRONMENT: 1) SAME SHOWN IN WHITE 2) DR DYNAMICS IS SHOWN IN BLUSH, AND 3) DR VISUAL IN YELLOW.

	Visual-Proprio			Visual-Tactile-Proprio								Tactile-Proprio
	MP DrQv2	MP DrQv2 DR Dyn	MP DrQv2 DR Vis	MP DrQv2	MP DrQv2 DR Dyn	MP DrQv2 DR Vis	SP DrQv2	MP DrQv2 Tac Gate	MP DrQv2 w/o Tac Aug	MP DrQv2 Cam Drop	DrTD3	MP DrQv2
TactileReach	107 ∓ 9	-	59 ∓ 14	160 ∓ 79	-	53 ∓ 11	109 ∓ 4	$\mathbf{236 \mp 126}$	109 ∓ 3	67 ∓ 10	115 ∓ 45	74 ∓ 45
Same	110 ∓ 19	-	57 ∓ 25	$\mathbf{361 \mp 8}$	-	75 ∓ 13	119 ∓ 62	351 ∓ 20	162 ∓ 103	100 ∓ 34	-	87 ∓ 99
DR Visual	45 ∓ 37	-	53 ∓ 23	58 ∓ 75	-	52 ∓ 22	44 ∓ 37	91 ∓ 134	43 ∓ 33	$\mathbf{102 \mp 79}$	-	-
Door	340 ∓ 51	339 ∓ 26	152 ∓ 80	$\mathbf{369 \mp 3}$	268 ∓ 127	123 ∓ 45	272 ∓ 152	277 ∓ 127	227 ∓ 155	245 ∓ 157	27 ∓ 24	210 ∓ 140
Same	370 ∓ 7	328 ∓ 106	54 ∓ 45	$\mathbf{372 \mp 1}$	358 ∓ 3	26 ∓ 11	372 ∓ 1	369 ∓ 4	260 ∓ 139	366 ∓ 14	-	369 ∓ 1
DR Dynamics	279 ∓ 126	240 ∓ 131	37 ∓ 50	$\mathbf{350 \mp 21}$	317 ∓ 97	25 ∓ 15	309 ∓ 116	218 ∓ 159	277 ∓ 110	292 ∓ 131	-	273 ∓ 151
DR Visual	88 ∓ 141	78 ∓ 143	111 ∓ 142	82 ∓ 148	16 ∓ 39	168 ∓ 150	48 ∓ 104	110 ∓ 20	108 ∓ 29	363 ∓ 24	-	$\mathbf{366 \mp 2}$
TactileLift	146 ∓ 38	-	-	167 ∓ 72	190 ∓ 46	49 ∓ 9	-	$\mathbf{194 \mp 70}$	167 ∓ 54	68 ∓ 35	40 ∓ 10	44 ∓ 10
Same	$\mathbf{328 \mp 54}$	-	-	315 ∓ 103	251 ∓ 125	36 ∓ 15	-	305 ∓ 92	235 ∓ 131	136 ∓ 154	-	51 ∓ 47
DR Dynamics	235 ∓ 119	-	-	242 ∓ 124	$\mathbf{257 \mp 103}$	40 ∓ 13	-	246 ∓ 89	242 ∓ 124	98 ∓ 96	-	74 ∓ 107
DR Visual	51 ∓ 67	-	-	47 ∓ 84	68 ∓ 111	61 ∓ 30	-	40 ∓ 31	51 ∓ 108	$\mathbf{86 \mp 110}$	-	57 ∓ 106

of TD3 [36], dubbed DrTD3, which utilizes the same encoder architecture and data augmentation strategy as MP-DrQv2, but without the n-step TD error estimates and scheduled exploration noise.

1) Perception Architecture: We investigate two image encoder architectures: **MultiPath (MP)** and **SinglePath (SP)**. In the MultiPath paradigm, a unique encoding network is used for vision and tactile, while in SP the same encoder is shared by both modalities. Our agent utilizes the past 3 frames as an input state, resulting in an observation size of $(9 \times 84 \times 84)$ for the RGB image and an observation size of $(3 \times 84 \times 84)$ for the tactile depth image when using the palm tactile sensor and $(6 \times 84 \times 84)$ when utilizing the gripper. In the SinglePath setting, images from both sensors are combined on the channels axis, producing an input of $(12 \times 84 \times 84)$ to the convolutional encoder with the palm sensor. We employ the same convolutional encoder architecture described in DrQv2 [2] in all pixel-based encoders in this paper. We consider **MP vs SP** in Table I and find that overwhelmingly, the MP architecture offers higher performance, justifying the increase in the number of model parameters and wall clock training time necessary for separate encoders.

2) Tactile Gating: The inclusion of **tactile gating** in the model architecture is shown to improve the learning speed in contact-rich tasks. We present the learning curve for TactileReach in Fig. 5 where the use of a tactile gate provides an improvement in performance of 25%. We also note that this agent needs fewer environment interactions to learn to solve the task, and is capable of exploiting the tactile sensor more effectively than baseline methods. We note that for the Door task, the inclusions of tactile gating did not significantly alter the learning performance, but did improve robustness to visual perturbations. We hypothesize that this is due to the fact that tactile reasoning is less critical for the successful executions of this tasks in the simulated environment.

3) Visual Degradation: We test the several methods of **visual degradation** where we reduce the quality of the visual observations to encourage the agent to utilize tactile information. We tested visual degradation by training agents with **visual dropout** and **DR Visual** and find that visual dropout produces more performant agents in both the standard environment and DR Visual evaluation. Note that visual degradation techniques have a negative impact on overall system performance, but do seem to improve agent reliance on tactile information. Given the promising performance of image dropout to improve multimodal sensing, this technique should be explored further, for example by exploiting the simulation environment to implement conditional camera degradation that depends on the presence of a tactile signal.

4) Tactile Augmentation: We find that Tactile Augmentation improves performance on most tasks as shown in a head-to-head comparison of columns MP-DrQv2 and DrQv2 w/o Tac Aug in Table I. This suggests that the augmentation approaches that are successful in pixel-based RL may translate well to the tactile signal.

VII. CONCLUSION

This paper explores the ability of deep reinforcement learning methods to fuse and exploit visual and tactile feedback to learn manipulation policies. We focus on the problem setting where tactile feedback is provided by optical-based sensors that render high resolution pixel-level information. We find that the fusion of both modalities results in optimal performance on a set of manipulation tasks as well as an improved robustness to system perturbations in the dynamics and lighting conditions.

Key to these results is the inclusion of a tactile gate that controls the flow of tactile feedback through the agent's network. We show that tactile gating results in an agent that can exploit tactile sensing earlier and achieve higher performance than benchmarks that employ common encoder-based perceptual modules. Additionally, we show that the use of data augmentation techniques adapted from DrQv2 on both tactile and visual streams is beneficial to robust learning. While this is well known for visual feedback, we show that this technique also leads to significant improvements with tactile sensing.

REFERENCES

[1] C. Bazzacchi, R. Volcic, and F. Domini, "Effect of visual and haptic feedback on grasping movements," *J. Neurophysiol*, vol. 112, pp. 3189–3196, 2104.

[2] D. Yarats, R. Fergus, A. Lazaric, and L. Pinto, "Mastering visual continuous control: Improved data-augmented reinforcement learning," *arXiv preprint arXiv:2107.09645*, 2021.

[3] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, Canada, 2019, pp. 8943–8950.

[4] 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.

[5] I. Taylor, S. Dong, and A. Rodriguez, "Gelslim3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," 2021.

[6] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "OmniTact: A multi-dimensional high resolution touch sensor," *arXiv preprint arXiv:2003.06965*, 2020.

[7] F. R. Hogan, M. Jenkin, S. Rezaei-Shoshtari, Y. Girdhar, D. Meger, and G. Dudek, "Seeing through your skin: recognizing objects with a novel visuotactile sensor," in *WACV*, Held Online, 2021.

[8] O. B. Kroemer, R. Detry, J. Piater, and J. Peters, "Combining active learning and reactive control for robot grasping," *Robotics and Autonomous systems*, vol. 58, no. 9, pp. 1105–1116, 2010.

[9] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.

[10] F. R. Hogan and A. Rodriguez, "Feedback control of the pusher-slider system: A story of hybrid and underactuated contact dynamics," *arXiv preprint arXiv:1611.08268*, 2016.

[11] M. A. Toussaint, K. R. Allen, K. A. Smith, and J. B. Tenenbaum, "Differentiable physics and stable modes for tool-use and manipulation planning," 2018.

[12] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-RL for insertion: Generalization to objects of unknown geometry," in *International Conference on Robotics and Automation (ICRA)*, X'ian, China, 2021.

[13] Y. Zhu, J. Wong, A. Mandlekar, and R. Martín-Martín, "Robosuite: A modular simulation framework and benchmark for robot learning," in *arXiv preprint arXiv:2009.12293*, 2020.

[14] M. I. Tiwana, S. J. Redmond, and N. H. Lovell, "A review of tactile sensing technologies with applications to biomedical engineering," *Sensors and Actuators A: Physical*, vol. 179, pp. 17–31, 2012.

[15] C. Chi, X. Sun, N. Xue, T. Li, and C. Liu, "Recent progress in technologies for tactile sensors," *Sensors(Basel)*, p. 948, 2018.

[16] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Madrid, Spain: IEEE, 2018, pp. 1927–1934.

[17] S. Tian, F. Ebert, D. Jayaraman, M. Mudigonda, C. Finn, R. Calandra, and S. Levine, "Manipulation by feel: Touch-based control with deep predictive models," in *International Conference on Robotics and Automation (ICRA)*, 2019, pp. 818–824.

[18] F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez, "Tactile dexterity: Manipulation primitives with tactile feedback," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 8863–8869.

[19] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, "Swingbot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5633–5640.

[20] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," *arXiv preprint arXiv:2104.01167*, 2021.

[21] H. van Hoof, N. Chen, M. Karl, P. van der Smagt, and J. Peters, "Stable reinforcement learning with autoencoders for tactile and visual data," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejon, Korea, 2016, pp. 3928–3934.

[22] A. Church, J. Lloyd, R. Hadsell, and N. F. Lepora, "Optical tactile sim-to-real policy transfer via real-to-sim tactile image translation," 2021.

[23] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[24] A. X. Lee, A. Nagabandi, P. Abbeel, and S. Levine, "Stochastic latent actor-critic: Deep reinforcement learning with a latent variable model," *arXiv preprint arXiv:1907.00953*, 2019.

[25] D. Yarats, A. Zhang, I. Kostrikov, B. Amos, J. Pineau, and R. Fergus, "Improving sample efficiency in model-free reinforcement learning from images," *arXiv preprint: arXiv 1910.01741*, 2019.

[26] D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson, "Learning latent dynamics for planning from pixels," *arXiv preprint arXiv:1811.04551*, 2018.

[27] D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, "Dream to control: Learning behaviors by latent imagination," *arXiv preprint arXiv:1912.01603*, 2019.

[28] D. Yarats, I. Kostrikov, and R. Fergus, "Image augmentation is all you need: Regularizing deep reinforcement learning from pixels," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=GY6-6sTvGaf

[29] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, 2017, pp. 23–30.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.

[32] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 5026–5033.

[33] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE J. Robotics Autom.*, vol. 3, pp. 43–53, 1987.

[34] S. Wang, M. Lambeta, L. Chou, and R. Calandra, "Tacto: A fast, flexible and open-source simulator for high-resolution vision-based tactile sensors," *Arxiv*, 2020. [Online]. Available: https://arxiv.org/abs/2012.08456

[35] M. Bauza, E. Valls, B. Lim, T. Sechopoulos, and A. Rodriguez, "Tactile object pose estimation from the first touch with geometric contact rendering," *arXiv preprint arXiv:2012.05205*, 2020.

[36] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International Conference on Machine Learning*, 2018, pp. 1582–1591.