APPLE: Toward General Active Perception via Reinforcement Learning

Tim Schneider[1,2], Cristiana de Farias[1], Roberto Calandra[3], Liming Chen[2], and Jan Peters[4]

***Abstract*—Active perception is a fundamental skill that enables us humans to deal with uncertainty in our inherently partially observable environment. For senses such as touch, where the information is sparse and local, active perception becomes crucial. Hence, in recent years, it has emerged as an important research domain in robotics. However, current methods are often bound to specific tasks or make strong assumptions, which limit their generality. To address this gap, this work introduces `APPLE` (Active Perception Policy Learning) – a novel framework that leverages reinforcement learning (RL) to address a range of different active perception problems. `APPLE` jointly trains a transformer-based perception module and decision-making policy with a unified optimization objective, learning how to actively gather information. By design, `APPLE` is not limited to a specific task and can, in principle, be applied to a wide range of active perception problems. We evaluate two variants of `APPLE` across different tasks, including tactile exploration problems. Experiments demonstrate the efficacy of `APPLE`, achieving high accuracies on both regression and classification tasks. These findings underscore the potential of `APPLE` as a versatile and general framework for advancing active perception in robotics. Project page: https://timschneider42.github.io/apple**

I. Introduction

Imagine searching for a set of tools inside a cluttered toolbox. Rather than waiting passively for the information to reveal itself, humans place their hands inside the box and begin exploring, probing and adjusting their motions based on the feedback received. This process illustrates the concept of *active perception*: the deliberate selection of actions to acquire information in the face of uncertainty [1]. Active perception does not aim to exhaustively explore every aspect of the world; it focuses on reducing uncertainty about specific properties of the environment. Equipping robots with this capability is essential for autonomy in unstructured, noisy environments.

Tactile sensing is a natural fit for studying active perception because it is inherently local; purposeful interaction is often the only way to gather meaningful data [2]. At the same time, existing active tactile sensing methods often rely on task-specific heuristics, such as maximizing force closure or reconstructing shape, and frequently assume objects remain stationary [3, 4, 5]. Reinforcement learning (RL), on the other hand, provides a more general framework by learning sequential strategies directly from interaction. While some RL approaches for active tactile perception exist [6, 7], they are often sample-inefficient or bound to specific tasks.

[1]Department of Computer Science, TU Darmstadt, Germany.
[2]LIRIS, CNRS UMR5205, École Centrale de Lyon, France.
[3]LASR Lab & CeTI, TU Dresden, Germany.
[4]DFKI, Hessian.AI, RIG, and Centre for Cognitive Science, TU Darmstadt, Germany.
Corresponding author: `tim@robot-learning.de`

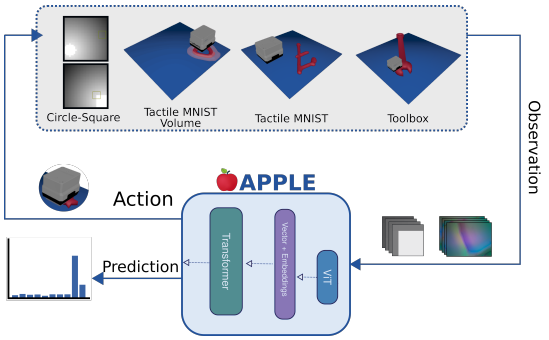

Figure 1: Our method ***A**ctive **P**erception **P**olicy **Le**arning* (`APPLE`) jointly optimizes a policy for information gathering and a prediction model for inference using a shared transformer-based backbone. Shown at the top are benchmark tasks used to evaluate `APPLE`.

In this work, we ask whether a principled RL algorithm can discover active perception policies using only ground-truth labels and a differentiable loss. We frame active perception within partially observable Markov decision processes (POMDPs) and introduce `APPLE` (**A**ctive **P**erception **P**olicy **Le**arning). `APPLE` jointly trains a decision policy and a perception module on a shared transformer backbone, allowing it to function on diverse tasks without specialized heuristics.

Our main contributions are:

- A unified formulation for active perception as an interactive supervised learning problem.
- `APPLE`, a framework that jointly optimizes RL policies and perception modules on a shared transformer backbone, requiring minimal task-specific assumptions.
- Empirical evaluation of two variants of `APPLE` across multiple tasks, demonstrating that `APPLE` discovers effective exploration policies across varied objectives.

II. Active Perception Policy Learning

Our objective in this work is to develop an active perception method that, unlike prior approaches, is not tied to a particular task or environment. Just as RL requires only a reward function, we want to specify a *perception objective* and let the agent learn an appropriate perception policy on its own, without imposing strong task-specific assumptions. On a high level, we frame active perception as a supervised learning problem. The agent's goal is to minimize a loss $\ell(y_t, \overset{*}{y}_t)$ between its current prediction y_t and the ground-truth label $\overset{*}{y}_t$. However, unlike in classical supervised learning, we assume that the agent is not simply presented with a static data point as input, but rather with an interactive environment that it can actively gather data from. E.g., the agent could be presented with an object and has

to decide actively how to examine it to extract the information it needs. This perspective defines active perception fundamentally as a sequential decision-making problem embedded within a supervised learning problem.

Hence, formally, we aim to solve the following optimization problem:

$$\min_{\theta} J(\pi_\theta) \coloneqq \mathbb{E}_{p\left(\mathbf{h},\overset{*}{\mathbf{y}},\mathbf{o},\mathbf{a},\mathbf{y}\,\middle|\,\pi_\theta\right)}\left[\sum_{t=0}^{\infty}\gamma^t \ell(\overset{*}{y}_t, y_t)\right]$$

where π_θ is the agent's policy parameterized by θ, h_t is the environment hidden state at time t, o_t is the observation, the agent makes, and a_t is the action the agent takes. Computing the gradient of $J(\pi_\theta)$ yields

$$\frac{\partial}{\partial\theta} J(\pi_\theta) = \underbrace{\mathbb{E}_{p\left(\mathbf{h},\overset{*}{\mathbf{y}},\mathbf{o},\mathbf{a}\,\middle|\,\pi_\theta\right)}\left[\frac{\partial}{\partial\theta}\ln\pi_\theta(\mathbf{a}\,|\,\mathbf{o})\sum_{t=0}^{\infty}\gamma^t\tilde{r}(h_t,\overset{*}{y}_t,a_t,y_t)\right]}_{\text{policy gradient}} + \underbrace{\mathbb{E}_{p\left(\overset{*}{\mathbf{y}},\mathbf{o}\right)}\left[\sum_{t=0}^{\infty}\gamma^t\frac{\partial}{\partial\theta}\ell_{\pi_\theta}(\overset{*}{y}_t, o_{0:t})\right]}_{\text{prediction loss gradient}}. \quad (1)$$

Hence, the gradient of the objective function $J(\pi_\theta)$ decomposes into a policy gradient and a supervised prediction loss gradient. Fortunately, there are many existing approaches to estimate policy gradients in the RL literature. With Eq. (1) we have a recipe for transforming these RL methods into active perception methods by following these steps:

1) The active perception setting is partially observed. Hence, instead of a state s_t, the policy and (if applicable) the critic networks receive a trajectory of past observations $o_{0:t}$.

2) If applicable, during the training of the critic, the presence of the prediction loss ℓ_{π_θ} requires us to dynamically recompute the total reward when evaluating the Bellman residual.

3) During the policy update, the policy gradient is augmented by the prediction loss gradient.

We call this method **A**ctive **P**erception **P**olicy **Le**arning (APPLE) and propose two variants, based on SAC [8] and CrossQ [9], which we call APPLE-SAC and APPLE-CrossQ. For input processing, we assume that the sequence of past observations $o_{0:t}$ consists of both images and scalar data and use an architecture similar to the Video-Vision-Transformer (ViViT) [10], learning a shared embedding for actor, critic, and label predictor on-the-fly. An overview of our method can be found in Fig. 1.

III. Evaluation

With APPLE, our goal is to answer the following question: Can we design a general and principled RL-based algorithm that successfully discovers active-perception policies for various input modalities using only a task label and a differentiable loss during training without the need to design task-specific exploration heuristics? To answer this question, we show results of three simulated experiments[1]: (1) MHBS, a taxel-grid based classification task introduced in [6], (2) CircleSquare, a vision-based classification task in which the agent has to move a glimpse and perform binary classification, and (3) TactileMNIST, a task from the Tactile MNIST Benchmark Suite [11], in which the agent has to classify 3D MNIST digits using a simulated GelSight Mini sensor. Visualizations of these tasks can be seen in Fig. 1.

[1]More experiments can be found in the main paper linked in the project page: https://timschneider42.github.io/apple.

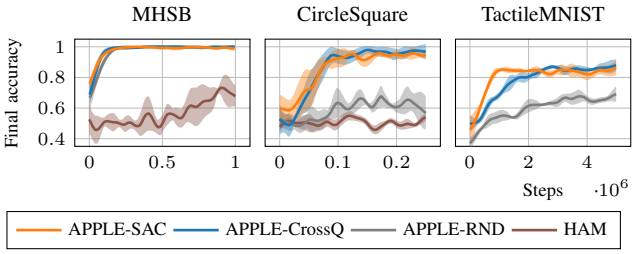

Figure 2: Final prediction accuracies for our methods APPLE-SAC and APPLE-CrossQ, HAM [6], and APPLE-RND across various tasks. *MHSB* refers to the tactile classification task used in [6]. All methods were trained with 5 seeds. Shaded areas represent one standard deviation. Metrics on TactileMNIST are computed on evaluation tasks with unseen objects.

In each experiment, we compare both APPLE approaches against a random baseline APPLE-RND, which shares the same configuration as APPLE-SAC, but does not optimize an action policy and instead, chooses actions randomly. Importantly, the perception module is still trained, enabling the model to learn how to interpret observations even without control over the movement of its (haptic) glances. Additionally, we compare our approach against HAM [6], which employs a recurrent neural network (LSTM) to integrate tactile observations over time and jointly learns to classify objects while optimizing its exploratory actions through REINFORCE.

The results in Section III show that APPLE successfully learns exploration strategies across all tested tasks. At the same time, the poor performance of the APPLE-RND baseline across our tasks confirms the necessity of structured exploration and confirms that our methods learned policies that go beyond random exploration. HAM [6], performs worse than APPLE-RND in all experiments, which we attribute to the fact that it is using on-policy RL, which discards samples after a single update, limiting sample efficiency. In contrast, our off-policy methods reuse samples – a critical factor in active perception, where supervised learning benefits from multiple passes over the same data. The performance of APPLE on these tasks demonstrates its potential as a robust, general framework for active perception.

IV. Conclusion and Future Work

We present APPLE, a framework that integrates reinforcement learning with transformer-based architectures to enable active tactile perception. Empirical results demonstrate that APPLE develops efficient exploration strategies across various active perception tasks, consistently outperforming baselines. While these experiments span diverse scenarios, the current benchmarks remain relatively fundamental; thus, deploying APPLE in complex, unstructured environments remains an open challenge. To facilitate this transition to the real world, future research will focus on enhancing the framework's sample efficiency.

Acknowledgments: This work was supported by the German Federal Ministry of Education and Research (BMBF) and the French Research Agency, l'Agence Nationale de Recherche (ANR), through the projects *Aristotle* (Grant no.: ANR-21-FAI1-0009-01), *Chiron* (Grant no.: ANR-20-IADJ-0001-01), and Astérix (Grant no.: ANR-23-EDIA-0002) through the EU's Horizon Europe project *ARISE* (Grant no.: 101135959), the BMBF's projects Robotics Institute Germany (RIG) (Grant no.: 16ME1001), *DEMETER* (Grant no.: 01DR25003), the French national investment priority program's PSPC FAIR WASTE project, and *Genius Robot* (Grant no.: 01IS24083). Furthermore, this work is also supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy EXC 2050/1 – Project ID 390696704 – Cluster of Excellence *Centre for Tactile Internet with Human-in-the-Loop* (CeTI) of Technische Universität Dresden by the BMBF, and by the German Academic Exchange Service (DAAD) in project 57616814 (SECAI School of Embedded and Composite AI). The computations were conducted on the IAS Compute Cluster.

References

[1] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

[2] Tony J. Prescott, Mathew E. Diamond, and Alan M. Wing. Active touch sensing. *Phil. Trans. R. Soc. B*, 366(1581):2989–2995, Nov 2011.

[3] Marten Björkman, Yasemin Bekiroglu, Virgile Högman, and Danica Kragic. Enhancing visual perception of shape through tactile glances. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3180–3186. IEEE, 2013.

[4] Stanimir Dragiev, Marc Toussaint, and Michael Gienger. Uncertainty aware grasping and tactile exploration. In *2013 IEEE International conference on robotics and automation*, pages 113–119. IEEE, 2013.

[5] Cristiana De Farias, Naresh Marturi, Rustam Stolkin, and Yasemin Bekiroglu. Simultaneous tactile exploration and grasp refinement for unknown objects. *IEEE Robotics and Automation Letters*, 6(2):3349–3356, 2021.

[6] Sascha Fleer, Alexandra Moringen, Roberta L Klatzky, and Helge Ritter. Learning efficient haptic shape exploration with a rigid tactile sensor array. *PloS one*, 15(1):e0226880, 2020.

[7] Jingxi Xu, Shuran Song, and Matei Ciocarlie. Tandem: Learning joint exploration and decision making with tactile sensors. *IEEE Robotics and Automation Letters*, 7(4):10391–10398, 2022.

[8] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

[9] Aditya Bhatt, Daniel Palenicek, Boris Belousov, Max Argus, Artemij Amiranashvili, Thomas Brox, and Jan Peters. CrossQ: Batch Normalization in Deep Reinforcement Learning for Greater Sample Efficiency and Simplicity. In *The Twelfth International Conference on Learning Representations*, 2019.

[10] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6836–6846, 2021.

[11] Tim Schneider, Guillaume Duret, Cristiana de Farias, Roberto Calandra, Liming Chen, and Jan Peters. Tactile mnist: Benchmarking active tactile perception. *arXiv preprint arXiv:2506.06361*, 2025.