Speeding Up Affordance Learning for Tool Use, Using Proprioceptive and Kinesthetic Inputs

Khuong N. Nguyen, Jaewook Yoo, and Yoonsuck Choe
Department of Computer Science and Engineering, Texas A&M University - College Station, TX 77843
{knn1989,jwookyoo,choe}@tamu.edu

***Abstract*—End-to-end learning in deep reinforcement learning based on raw visual input has shown great promise in various tasks involving sensorimotor control. However, complex tasks such as tool use require recognition of affordance and a series of non-trivial subtasks such as reaching the tool, grasping the tool, and wielding the tool. In such tasks, end-to-end approaches with only the raw input (e.g. pixel-wise images) may fail to learn to perform the task or may take too long to converge. In this paper, inspired by the biological sensorimotor system, we explore the use of proprioceptive/kinesthetic inputs (internal inputs for body position and motion) as well as raw visual inputs (exteroception, external perception) for use in affordance learning for tool use tasks. We set up a reaching task in a simulated physics environment (MuJoCo), where the agent has to pick up a T-shaped tool to reach and drag a target object to a designated region in the environment. We used an Actor-Critic-based reinforcement learning algorithm called ACKTR (Actor-Critic using Kronecker-Factored Trust Region) and trained it using various input conditions to assess the utility of proprioceptive/kinesthetic inputs. Our results show that the inclusion of proprioceptive/kinesthetic inputs (position and velocity of the limb) greatly enhances the performance of the agent: higher success rate, and faster convergence to the solution. The lesson we learned is the important factor of the intertwined relationship of exteroceptive and proprioceptive in sensorimotor learning and that although end-to-end learning based on raw input may be appealing, separating the exteroceptive and proprioceptive/kinesthetic factors in the input to the learner, and providing the necessary internal inputs can lead to faster, more effective learning.**

***Index Terms*—affordance, tool use, proprioception, kinesthesia, sensorimotor system, reinforcement learning.**

I. INTRODUCTION

End-to-end deep reinforcement learning algorithms such as Deep Q-Network [1] have become a powerful tool for reinforcement learning in complex perceptual environments [2]–[5]. In these algorithms, the agent learns directly from raw visual inputs, e.g., a series of frames from video games or 3D environment simulations, bypassing any feature extraction stage.

However, complex tasks such as tool use [6], requires that the intelligent agent possesses high levels of sensorimotor skills which facilitate a variety of perception capabilities [7] enabling it to complete a series of non-trivial subtasks such as recognizing the affordance of the tool, reaching the tool, grasping the tool, and wielding the tool. While the advancement of deep reinforcement learning research in recent years has been giving rise to many powerful methods, end-to-end deep reinforcement learning approaches that utilize only raw pixel-wise images may still fail in this type of complicated task or may take too long to learn.

Inspired by the biological sensorimotor system and the cognitive psychological concept of affordance, we investigate the relationship between proprioception/kinesthesia (internal perception) and exteroception (external perception such as vision) for use in affordance learning within a tool use domain. We set up this task in a simulated physics environment (created using MuJoCo physics engine [8]), where the agent has to pick up a T-shaped tool to reach and drag a target object to a designated region in the environment. We used a synchronous version of the ACKTR (Actor-Critic using Kronecker-Factored Trust Region) [9] algorithm which is an Actor-Critic-based reinforcement learning algorithm and trained it employing various input conditions to test the utility of proprioceptive/kinesthetic feedbacks. Our results show that the inclusion of proprioceptive/kinesthetic inputs (position and velocity of the limb) greatly enhances the performance of the agent: higher success rate, and faster convergence. Furthermore, the results confirmed that exteroceptive and internal somatic (proprioceptive/kinesthetic) inputs together facilitate the affordance learning process (similarly with the biological sensorimotor system) and thus lead to better and more effective learning.

The rest of the paper is organized as follows. We first provide some background on the concept of affordance, discuss the role of proprioception/kinesthesia in the sensorimotor loop, and describe our tool use environment. Next, we present our method to examine the effectiveness of proprioceptive/kinesthetic inputs and visual inputs in affordance learning. We then present the experiments and results, followed by discussion and conclusion.

II. BACKGROUND

A. Affordance and Affordance Learning

The theory of affordance was first introduced by the psychologist James J. Gibson. In his seminal work, Gibson defined the term as "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up." [10]. In other words, affordance refers to the link between the agent and the environment plus the possibility of interactions. Perceiving the affordance is perceiving the interactive properties of the

environment and the agent. An agent situated in an environment perceives the properties of the environment including what is inside of that environment and observe the relationship between those and the properties of the agent itself, then infer what actions can be taken.

According to the affordance theory, the intelligent agent observes the environment through the affordances. This ability, in turn, allows the agent to deal with more complex and dynamic situations. The theory has been receiving much attention in the artificial intelligence and robotics community [11]–[14]. Most of the focus in these studies is on affordance learning which is the process of learning to perceive the possibilities of action in the environment. Although Gibson invented the concept on affordance, he did not provide concrete procedures for learning to perceive affordance. See E. J. Gibson's work on perceptual learning based on affordance, which provides more ideas in this respect [15].

While there have been many studies on the learning of affordances [14], [16]–[18], only a few tried to investigate the underlying fundamentals of affordance to utilize this concept better [19], [20]. It is generally agreed that affordance learning is the prerequisite of affordance perception. To perceive an affordance, one must learn to become familiar with the affordance. However, to identify the affordance, besides understanding the environment and knowing the properties of the environment, one must also have a good sense of its intrinsic properties in order to observe the relationship between the environment and itself. Our work in this study focuses on these fundamentals.

B. The Sensorimotor Loop and Proprioception

The sensorimotor loop integrates the sensory system (in charge of sampling sensory information) and the motor system (in charge of producing motor actions) in an agent. It is referred to as a loop because it samples sensory input and generates motor actions, which in turn lead to new sensory input. It has been shown that the sensorimotor loop is essential in intelligent systems [21]–[23]. Motor skills are formed and reinforced via this loop.

To produce effective and sophisticated motor actions, the sensorimotor loop requires high-quality sensory input signals including exteroceptive (sight, taste, smell, touch, hearing), proprioceptive (body position and orientation), kinesthetic (body motion) input and the efficient processing of these input signals. In this study, our focus is on the role of proprioception/kinesthesia which is the perception of the body relative spatial position and its movements. This perceptual ability enables one to be aware of oneself and is a vital part of sensorimotor learning [24], [25]. Since affordance learning is the process of gaining knowledge and identifying the relationships between the environment and the agent, it is thought that such internal senses play an essential role.

C. Tool Use in Animals

It was previously thought that human is the only species that can use tools. In fact, we used to think that the ability to use tool separate us from animals. However, this idea is challenged as more evidence was uncovered since the last century. A small number of animals were found to exhibit the ability to use tools. For instance, chimpanzees use simple wood and stone tools to obtain food and water [26], [27]. Crows and parrots use sticks or strings to reach objects beyond their reach [28], [29]. Dolphins use sponges as a tool to assist in digging [30]. Recently, it has also been discovered that Australian blackspot tuskfish can use rocks as a tool to bash open the clamshell and get the flesh inside [31]. While some tool use behaviors were a genetic feature in the species itself [32], [33] which are instinctive, others were developed through social learning by watching others use tools, or by using explorative behavior. These tool use behaviors involve learning and cognitive development and may be considered to be examples of intelligent tool use.

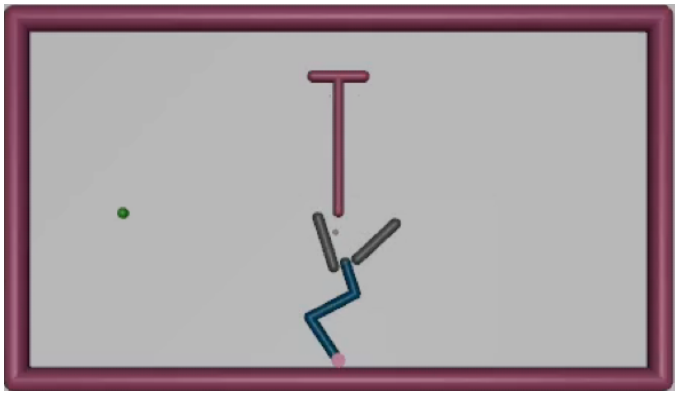


Fig. 1: A snapshot of the Tool-use environment. This environment is enclosed in a rectangular arena (dark pink border bars). The agent has three controlled joints (one joint shown in pink at the bottom and two green joints) and a gripper (dark gray). The T-shaped tool is located above the agent (dark pink t-shaped object). The job of the agent is to learn the affordance of the tool in order to grasp and use the tool to drag the object (green round object to the left of the agent) down to the bottom of the arena.

Intelligent tool use behavior is one of the most notable signs of intelligence since it requires high levels of sensorimotor skills and problem-solving capabilities. To use tools, an intelligent agent has to develop sophisticated skills from learning the affordance of the tool by determining the features and functions of objects and using its explorative behavior and problem-solving skills. In the field of Artificial Intelligence, the number of studies of tool use is still limited [34], [35]. This paper investigates the affordance learning aspect in tool use behavior.

III. METHODS

A. Tool-use Environment

The tool-use environment that we used was developed using the MuJoCo physics simulator (friction, force, etc. modeled: [8]) and OpenAI Gym [36]. The environment introduces a

robot with a three-joint arm with one gripper (with one more joint), a T-shaped tool, and a small target object, all enclosed in a rectangular arena (fig. 1). The task of the agent is to try to grasp the tool and use it to drag the object down to the target area which is the bottom of the arena.

The Tool-use environment was designed as a continuous action space control benchmark. The action includes four continuous values where three of them are the torques that apply to the three lower joints on the agent. The fourth action value is applied to the gripper joint and considered like a discrete value where 1.0 is applied to close the gripper if the action input for the joint is greater than 0, otherwise -1.0 is applied to open the gripper. Since the actions are in a continuous space, their values are produced by sampling a multidimensional normal distribution (4 in this case) with a probability density function of

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where μ and σ are the mean and standard deviation generated by the reinforcement learning algorithm and π is the pi constant which is ≈ 3.14159265359.

There are four subtasks to be completed in a specific order to reach the goal. (1) $ReachTool$: First of all, the gripper of the agent needs to reach the T-shaped tool handle. (2) $GraspTool$: Secondly, it has to grasp the tool by the handle. (3) $ToolReachObj$: Thirdly, the agent must guide the tool to the object. (4) $ObjReachTar$: Finally, the agent must use the tool to drag the object down to the bottom of the arena. These subtasks can be combined to construct curricula when training the agent (some form a prerequisite of the other).

The reward functions utilizing the subtasks are as follows. Note that none of the body, tool, or object position. (except for proprioceptive and/or kinesthetic inputs, to be discussed later) are provided directly to the agent: These values are used only to compute the reward values.

$$r_\rho = \begin{cases} 300 + k_s * (s_{max} - s_k), & \text{if (4)} \\ 1.5 + 1.5 * r_{c3}, & \text{if } \neg(4) \wedge (3) \\ 0.25 + 1.25 * r_{c2}, & \text{if } \neg(4) \wedge \neg(3) \wedge (2) \\ 0.125, & \text{if } \neg(4) \wedge \neg(3) \wedge \neg(2) \wedge (1) \\ 0.125 * r_{c1}, & \text{otherwise} \end{cases}$$

where:

- $\rho = \{p^{\overrightarrow{\mathrm{G}}}, p^{\overrightarrow{\mathrm{L}}}, p^{\overrightarrow{\mathrm{R}}}, p^{\overrightarrow{\mathrm{H}}}, p^{\overrightarrow{\mathrm{H}}}, p^{\overrightarrow{\mathrm{O}}}, p^{\overrightarrow{\mathrm{T}}}\}$
- k_s is a constant used to weight the speed of the task completion. It is set to 3.0 in our experiment.
- s_{max} is the maximum time step for each episode (500 steps).
- s_k is the time steps so far.
- $r_{c1} = 1 - tanh^2(\frac{\|p^{\overrightarrow{\mathrm{G}}} - p^{\overrightarrow{\mathrm{H}}}\|_2}{k_w})$
- $r_{c2} = 1 - tanh^2(\frac{\|p^{\overrightarrow{\mathrm{E}}} - p^{\overrightarrow{\mathrm{O}}}\|_2}{k_w})$
- $r_{c3} = 1 - tanh^2(\frac{\|p^{\overrightarrow{\mathrm{O}}} - p^{\overrightarrow{\mathrm{T}}}\|_2}{k_w})$
- k_w is a constant that is set to the width of the arena.
- $p^{\overrightarrow{\mathrm{G}}}$ is the pinch position of the gripper.
- $p^{\overrightarrow{\mathrm{L}}}$ is the position of the left claw of the gripper.
- $p^{\overrightarrow{\mathrm{R}}}$ is the position of the right claw of the gripper.
- $p^{\overrightarrow{\mathrm{H}}}$ is the position of the tool handle.
- $p^{\overrightarrow{\mathrm{E}}}$ is the position of the tool end-effector.
- $p^{\overrightarrow{\mathrm{O}}}$ is the position of the object.
- $p^{\overrightarrow{\mathrm{T}}}$ is the position of the target (arena's bottom).

and:

$$ReachTool = \|p^{\overrightarrow{\mathrm{G}}} - p^{\overrightarrow{\mathrm{H}}}\|_2 < k_g \quad (1)$$

where $k_g = 1/2 *$ gripper's length

$$GraspTool = ReachTool \wedge (\theta_c < \|p^{\overrightarrow{\mathrm{L}}} - p^{\overrightarrow{\mathrm{R}}}\|_2 < \theta_o) \quad (2)$$

where θ_c & θ_o are the gripper's close and open thresholds

$$ToolReachObj = GraspTool \wedge (\|p^{\overrightarrow{\mathrm{E}}} - p^{\overrightarrow{\mathrm{O}}}\|_2 < k_o) \quad (3)$$

where $k_o = 1/2 *$ tool tip's length

$$ObjReachTar = GraspTool \wedge (\|p^{\overrightarrow{\mathrm{O}}} - p^{\overrightarrow{\mathrm{T}}}\|_2 < k_a) \quad (4)$$

where $k_a = 1/2 *$ arena boundary's thickness

B. Synchronous ACKTR

The Actor-Critic using Kronecker-Factored Trust-Region (ACKTR) algorithm was introduced by [9] where the method showed higher performance than other state-of-the-art reinforcement learning approaches such as A2C [5], PPO [37], and TRPO [4]. In this study, we used a modified version of ACKTR which turns the original ACKTR into a synchronous ACKTR version [38]. This method gives us the benefit of A2C where it creates multiple versions of the agent to interact with multiple versions of the environment to learn more efficiently. At the same time, it still preserves all the advantages that the original ACKTR has.

Our actor-critic neural network architecture (fig. 2) is divided into two main components. The first component is shared between the actor-network and the critic network which includes 3 convolutional layers (32@8x8, 64@4x4, and 32@3x3) and one dense layer (512 units) with a relu unit comes with each layer. The dense layer is concatenated with the proprioceptive/kinesthetic inputs (if the internal inputs are used) to feed into the second component. The second component consists of the individual parts of the actor-network with three dense layers (64 units, 64 units, and 4 output units) and the individual parts of the critic network with another three dense layers (64 units, 64 units, and 1 output unit) with a tanh unit comes with each layer. The first individual layer of each network in the second component receives input from the concatenated dense layer of the first component.

Each state S (the input) at time t provided by the environment consist of 4 consecutive images of the environment and 4 consecutive proprioceptive and/or kinesthetic feedbacks (generated from 4 consecutive time steps). The pixel inputs are acquired using the image frames of the scenes, and the proprioceptive/kinesthetic feedbacks (joint positions, or joint velocities, or both) are computed from the information

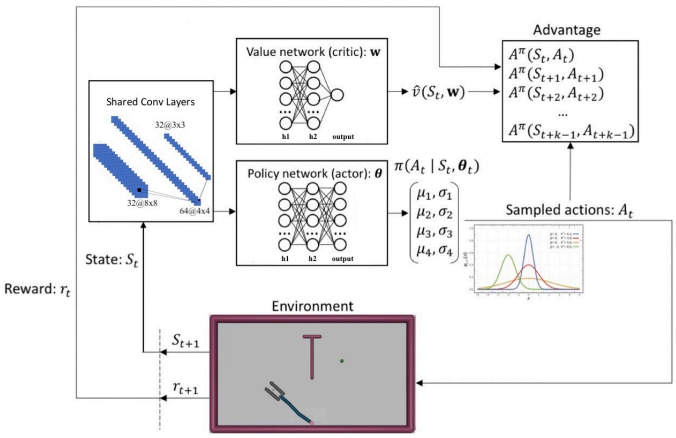


Fig. 2: The actor-critic network architecture used in this study. Note that the state information S_t includes both the pixel-based visual input and the optional proprioceptive/kinesthetic inputs. See text for details.

provided by OpenAI's Mujoco-py API. Pixel inputs are normalized by dividing by 255, and proprioceptive feedbacks are normalized by using running estimates of the means and standard deviations.

At time t, reward r_t and state

$$S_t = 4 * \{pixel\_input, p^{\vec{j_1}}, p^{\vec{j_2}}, p^{\vec{j_3}}, p^{\vec{g}}\ v^{\vec{j_1}}, v^{\vec{j_2}}, v^{\vec{j_3}}, v^{\vec{g}}\}$$

(pixel and optional proprioceptive/kinesthetic inputs) are generated, then fed into the policy and the value network. Next, the policy network produces the stochastic policy $\pi(A_t|S_t, \theta)$. The actions A_t are sampled from the multidimensional normal distribution. The sampled action vector A_t are fed into the advantage function and used to perform the next action. The value network estimates the value of the state $v^{(S_t,w)}$ which is fed into the advantage function. The advantage function $A^\pi(S_t, A_t)$ accumulates the values of A_t, $v^{(S_t,w)}$, and the reward r_t for k time steps, then use them to update the policy and value network using the natural gradient for the trust region. The environment receives the sampled action vector A_t, takes the next step, and produces the new state and reward (S_{t+1} and r_{t+1}). Since we used a synchronous version of the ACKTR, multiple copies of the environment and the agents (12 threads) are created. The agents will explore the state space simultaneously and update the networks by averaging the gradients over the 12 threads when all threads are finished. Compared to ACKTR, synchronous ACKTR greatly speeds up the learning process since it provides a more diverse experience for the agents. We also tried several other algorithms including DDPG, A3C, A2C, TRPO, and PPO, but ACKTR worked best for this particular task. Note that Henderson et al. in their comprehensive comparison of different RL algorithms [39] suggest that the choice of the best RL algorithm for a task can be tricky and nuanced since a best RL algorithm for one task may not perform well on another task.

IV. EXPERIMENTS & RESULTS

We conducted a series of experiments to answer three main questions:

1) Does proprioceptive/kinesthetic feedbacks contribute to vision-based affordance learning?
2) What are the impacts of using different types of internal inputs (proprioceptive and/or kinesthetic)?
3) What is the underlying mechanism of affordance perception in relation to exteroception and proprioception/kinesthesia?

The experiments in this study include training the agent in the tool-use environment (1) using pure pixel input only, (2) using pixel input with joint positions (proprioception), (3) using pixel input with joint velocities (kinesthesia), and (4) using pixel input with both joint positions and velocities (both internal inputs). We evaluated each experiment by the performance of the network (how much reward the model gains), and how fast each trial converges.

A. Pure Pixel-wise Input (Exteroception only)

The first experiment involves the use of only pixel input as the surrogate for exteroceptive feedback to train the actor-critic network. The agent had a hard time trying to learn the affordances, and its performance could not get past a reward of 55 (see fig. 3a and fig. 3e). For comparison, the max reward achieved in other conditions is 1,654.

B. Pixel Input with Agent's Joint Velocities (Kinesthesis)

For the second experiment, we began to provide kinesthetic feedbacks along with pixel inputs. Kinesthetic inputs, in this case, are the velocities of the joints. This time, the network was able to converge after 11,000 time steps and achieved a max reward of 1,489. (see fig. 3b and fig. 3f).

C. Pixel Input with Agent's Joint Positions (Proprioception)

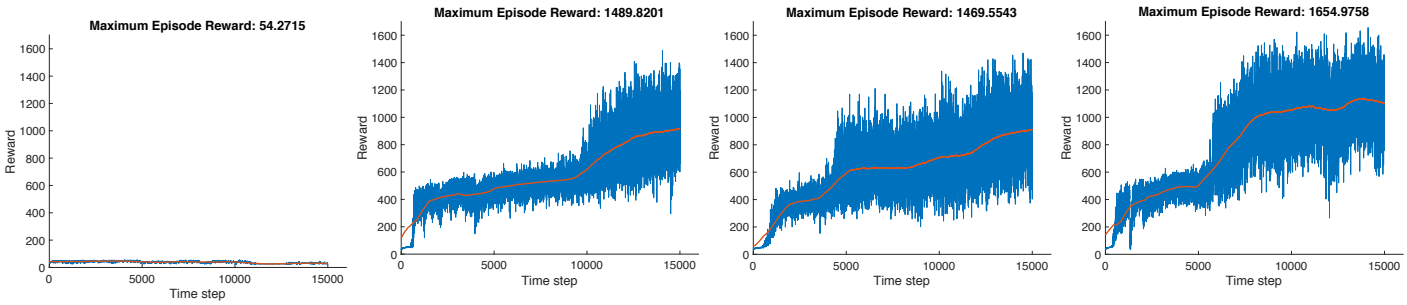
In the third experiment, instead of providing the joint velocities to the network, we provided the joint positions. This time, the network converged faster than the second experiment (around 5,000 time steps) but achieved a slightly lower max reward of 1,469 (see fig. 3c and fig. 3g).

D. Pixel Input with Agent's Joint Positions & Velocities (Proprioception + Kinesthesia)

Finally, when we used both the joint velocities and positions as proprioceptive inputs, the network converged a little bit slower than when we used joint positions only (5,900 time steps) but achieved an overall higher performance and had a peak reward of 1,655 (see fig. 3d and fig. 3h).

V. OBSERVATION AND ANALYSIS

In this section, we will analyze the experiments and form the answers to the three questions posed in the previous section.

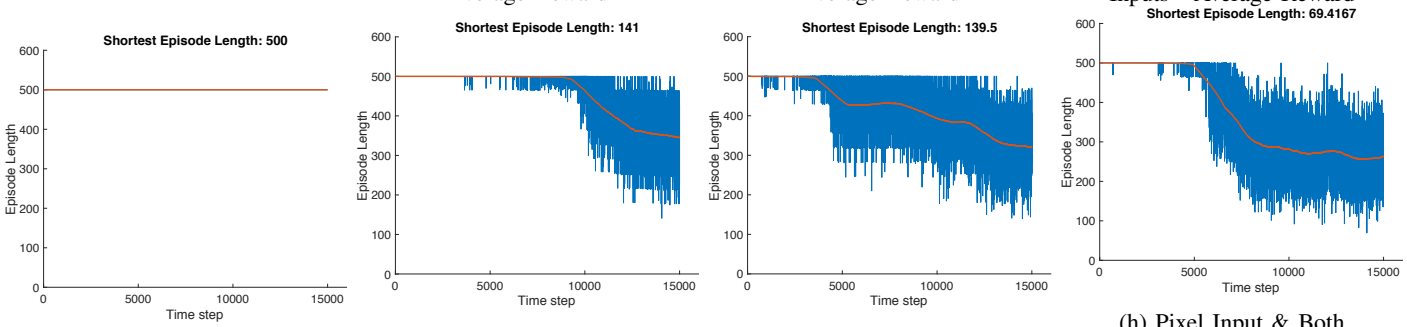


(a) Pixel Input - Average Reward

(b) Pixel Input & Joint Velocities - Average Reward

(c) Pixel Input & Joint Positions - Average Reward

(d) Pixel & Both Proprioceptive Inputs - Average Reward



(e) Pixel Input - Episode Length

(f) Pixel Input & Joint Velocities - Episode Length

(g) Pixel Input & Joint Positions - Episode Length

(h) Pixel Input & Both Proprioceptive Inputs - Episode Length

Fig. 3: The performance plots of the four experiments (the higher the reward, the better, and the shorter the episode length, the better). The red line in each sub-figure is the running average. (a) & (e) shows the performance of the network without the use of any proprioceptive/kinesthetic input. The network could not learn to accomplish the task, and the reward could not get past 55 as well as the episode length remains at 500 (worst). (b) & (f) show the performance of the network with the use of pixel input and joint velocities (kinesthesia). The training was able to converge. (c) & (g) show the performance of the network with the use of pixel input and joint positions (proprioception). This time the network converge faster than using joint velocities. (d) & (h) show the performance of the network with the use of pixel input and both types of internal bodily inputs (joint velocities and joint positions), which shows faster convergence and higher reward.

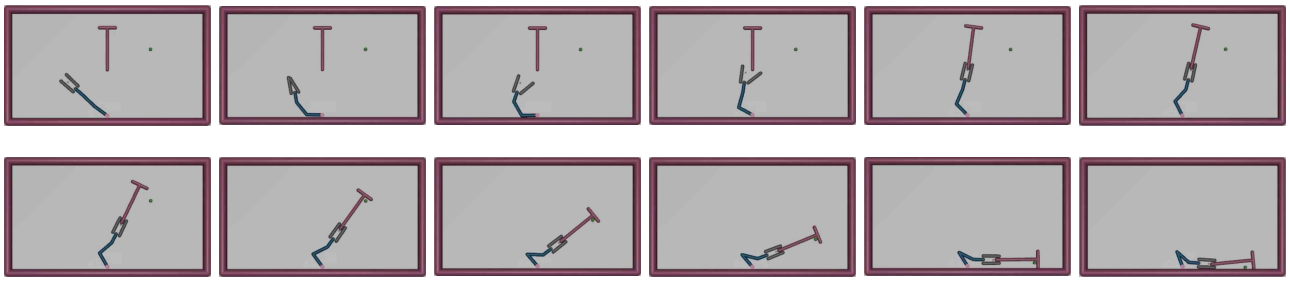


Fig. 4: A successful episode of the Tool-use task. The object and the agent are randomly positioned relative to the tool (frames are ordered left to right, top to bottom). The agent starts by bending down, open its gripper, and begins to reach the tool. The agent manages to get into the right position (aligning the tool to be inside of the gripper) and closes the gripper to firmly grasp the tool. It then wields the tool toward the object. After reaching the object it drags the object down to the bottom border of the arena, successfully finishing the task and ending the episode. This configuration uses pixel input and joint velocities (kinesthesia)

A. Analysis of The Tool-use Task

Fig. 4 shows a successfully completed episode of the tool-use task. In each episode, except for the tool, the initial posture of the agent, and the object are varied. The object location is randomly selected so that it is out of the agent's reach and the agent can either be on the right or the left side of the tool. To successfully complete the task, at first, the agent has to open its gripper and bend the arm while approaching the tool handle. Before approaching the tool handle, the agent should have correctly perceived the affordance of the tool to approach

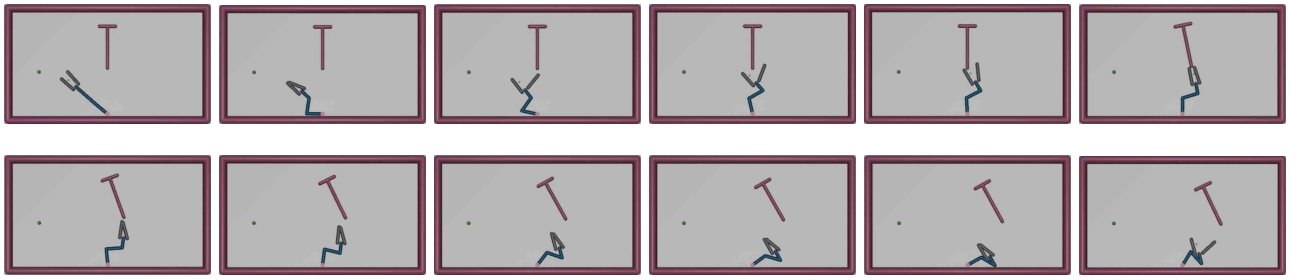


Fig. 5: A failed episode of the Tool-use task (frames are ordered left to right, top to bottom). The agent was not able to identify the affordance of the tool correctly, resulting in an unsuccessful attempt to grasp the tool: It temporarily held the tool but lost the grip, and never recovered. This configuration uses only pixel input and no proprioceptive or kinesthetic input.

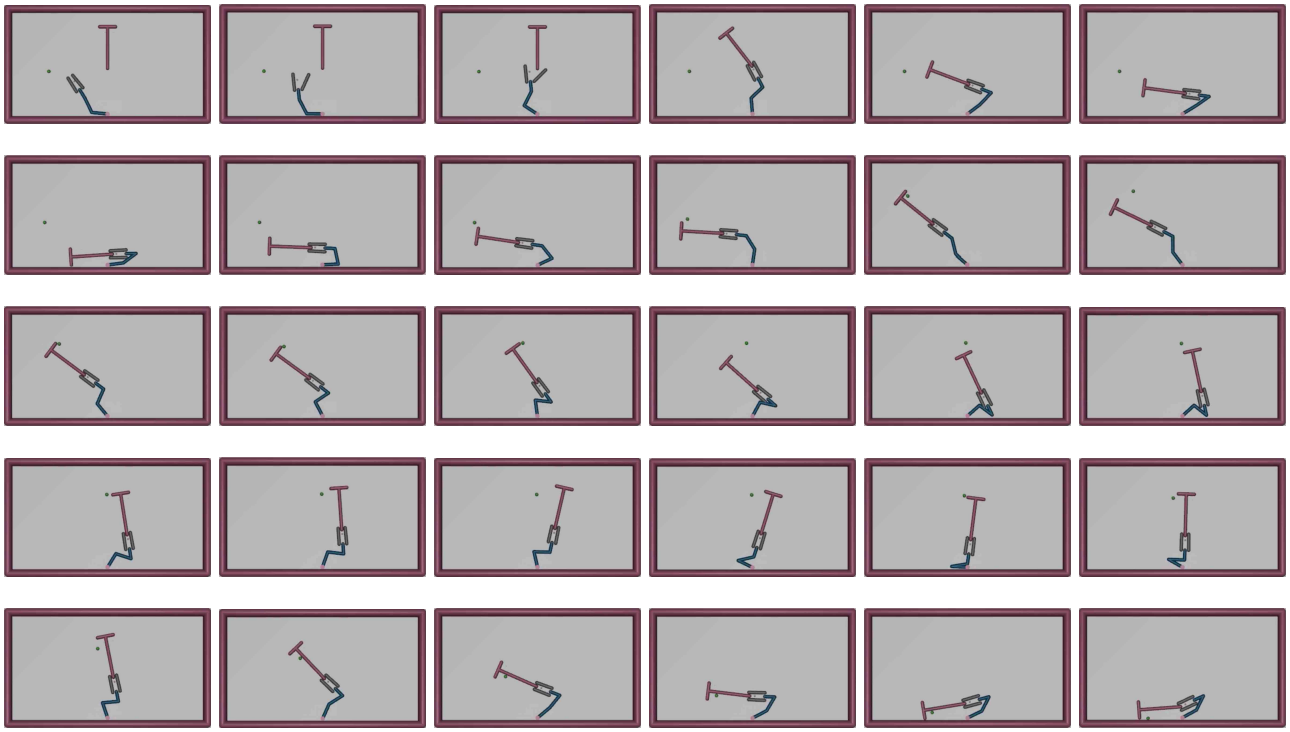


Fig. 6: A non-trivial successful episode where the agent learned to correct itself after performing some actions that might have led to failure but did not (frames are ordered left to right, top to bottom). The agent at first was able to grasp the tool but somehow managed to push the object upward. The agent then maneuvered the tool around the other side of the object and successfully dragged it down on the opposite side. This configuration uses pixel input and both types of internal bodily inputs(joint velocities and joint positions).

it properly (bend its arm and approach the tool to place the tool handle between the grippers) so that its gripper does not hit the tool and push the tool away (friction, mass, and force are appropriately modeled in MuJoCo). When the gripper reaches the tool handle, the agent closes the gripper to grasp the tool and then guides the tool toward the object while keeping the gripper closed. Note that the tool has to be well inside the grippers, or the agent will grasp the tool just by the tip of the handle and risk losing grip of the tool. Finally, it uses the tool to drag the object to the bottom. Most of the time, if the agent cannot identify the affordance of the tool, it will lead to a failed episode due to the unsuccessful grasping of the tool (see fig. 5).

B. Does proprioceptive/kinesthetic feedback contribute to vision-based affordance learning?

Based on our experimental results, proprioceptive/kinesthetic feedback helped speed up the learning process. Fig. 3 shows that the training only converges when either one or both of the two types of internal bodily (somatic) inputs are provided. One more observation is that more such

bodily inputs lead to faster convergence and higher reward. Additionally, while observing the trained agent perform its task, we found that in some instances, the agent even learned to correct itself after executing some movements that we thought might result in an unsuccessful episode (see fig. 6). The above suggests that proprioceptive/kinesthetic feedback plays an essential role in affordance learning for tool use, and in sensorimotor skill development in general.

C. What are the impacts of using different types of internal bodily inputs?

Internal bodily senses relating to motion consist of two main components which are the sense of relative position (proprioception) and the sense of movement (kinesthesis: see [40]). Both types of feedbacks were used in this studies. Even though training converges with the use of either joint positions or joint velocities alone, using joint positions converges faster than using joint velocities. This might suggest that spatial information is of more value than motion information in the process of affordance learning for tool use. A more dynamically challenging task than the one used in this paper may demand the use of both information.

D. What is the underlying mechanism of affordance perception in relation to proprioception and kinesthesia?

Affordance perception is the perception of the possibilities for interactions with the environment. However, it seems that current research on affordance perception focuses more on the exteroceptive side, not the internal bodily senses of the agent itself. Affordance perception is a two-way relationship between the intelligent agent and the environment involving the agent's exteroception heavily and proprioception/kinesthesis. The experiments in our study made clear the complementary relationship between exteroception and proprioception/kinesthesia and that without such internal bodily feedbacks it is harder or even impossible to perceive the affordance of the tool.

VI. DISCUSSION

Gibson stated that "An affordance, as I said, points two ways, to the environment and the observer. So does the information to specify an affordance." [10]. This suggests that affordance perception involves the complementary relationship between self-perception (understanding of your own body) and environmental perception (knowledge of the environment). Learning affordances is learning the connections between the environment and the learner. Additionally, it is important to note that affordance learning is considered as an active learning process in which the learner constantly develops and refines its sensorimotor skills by organizing the information from its learned experiences. Therefore, to learn the affordances, the learner continuously picks up the information from the environment through exteroception, retrieves internal information from proprioception and kinesthesia, and uses both sources of information to develop its ability to identify the affordances in the environment.

Through our experiments, we were able to confirm that proprioception and/or kinesthesia is an essential factor in tool-use affordance learning. Our results demonstrated that using proprioceptive/kinesthetic inputs significantly speeds up the learning process as models that involve both exteroception and proprioception/kinesthesia performed substantially better than models that utilize exteroception alone. Even with impaired internal bodily senses (joint positions or joint velocities alone), the learning is still much better than when such senses were absent. The above leads us to the conclusion that it is critical to consider both the exteroceptive factor and the internal factors in the inputs provided to the learner, and that utilizing the necessary internal feedbacks can lead to much more effective sensorimotor learning.

One interesting future direction is to make the task more challenging by making it into a combined tool construction and tool use task. Tool use is found in a select few animal species, but complex (multi-part) tool construction is almost absent in the animal kingdom. [41] showed that a neuroevolution-based agent can learn to construct a simple tool (an extended stick) in a reaching task similar to the one presented in this paper. However, in that work, all the inputs were hand-coded features (polar coordinate values between the end-effector and the objects), and the affordances were not recognized from the visual scene of the environment (no gripping: tool automatically attached to the end-effector when reached). It would be interesting to apply our affordance learning approach to such a tool-construction task.

Finally, here is a brief thought on end-to-end reinforcement learning. End-to-end learning really highlights the strength of deep neural network models, a large part of which is bypassing the tedious feature engineering step. So, in the application of end-to-end deep neural networks to reinforcement learning, often only raw visual (or other exteroceptive) inputs are supplied to the learning agent. Does providing internal bodily senses such as proprioception and kinesthesia violate this end-to-end property (i.e. are they just another hand-coded, manually engineered features)? We do not think so since in the animal, these internal signals directly come from the afferents embedded in the muscle fiber of the animal. So, in this sense, in our view, the inclusion of proprioceptive/kinesthetic signals in the input to the model does not violate the end-to-endness of end-to-end reinforcement learning.

VII. CONCLUSION

In this paper, we have shown how to accelerate the affordance learning process in the tool use task by using proprioceptive and kinesthetic feedbacks (spatial position and velocity of the limb) along with exteroceptive input (visual input). We conducted multiple experiments in which an agent learned to recognize the affordance of a tool placed in the environment and utilize this tool to accomplish its goal. These experiments with the use of different types of proprioceptive and kinesthetic feedbacks enabled us to analyze their impact on learning which led to higher success rate and faster convergence. The lesson we learned from this study is that although end-to-end

learning with raw exteroceptive input alone may be appealing, separately providing the exteroceptive and internal sensory factors in the input to the learner can lead to faster and more effective learning. In future work, we will extend the ideas in this paper to tackle with tool construction, and investigate how affordance learning can contribute.

REFERENCES

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[2] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning." in *AAAI*, vol. 2. Phoenix, AZ, 2016, p. 5.

[3] Z. Wang, V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. de Freitas, "Sample efficient actor-critic with experience replay," *arXiv preprint arXiv:1611.01224*, 2016.

[4] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International Conference on Machine Learning*, 2015, pp. 1889–1897.

[5] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*, 2016, pp. 1928–1937.

[6] K. R. L. Hall, "Tool-using performances as indicators of behavioral adaptability," *Current Anthropology*, vol. 4, no. 5, pp. 479–494, 1963.

[7] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behavioral and brain sciences*, vol. 24, no. 5, pp. 939–973, 2001.

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

[9] Y. Wu, E. Mansimov, R. B. Grosse, S. Liao, and J. Ba, "Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation," in *Advances in neural information processing systems*, 2017, pp. 5279–5288.

[10] J. J. Gibson, *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[11] M. Lopes, F. S. Melo, and L. Montesano, "Affordance-based imitation learning in robots," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 2007, pp. 1015–1021.

[12] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: from sensory–motor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[13] T. Hermans, J. M. Rehg, and A. Bobick, "Affordance prediction via learned object attributes," in *IEEE International Conference on Robotics and Automation (ICRA): Workshop on Semantic Perception, Mapping, and Exploration*. Citeseer, 2011, pp. 181–184.

[14] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 951–970, 2013.

[15] E. J. Gibson, *An odyssey in learning and perception*. Mit Press, 1994.

[16] T.-T. Do, A. Nguyen, I. Reid, D. G. Caldwell, and N. G. Tsagarakis, "Affordancenet: An end-to-end deep learning approach for object affordance detection," *arXiv preprint arXiv:1709.07326*, 2017.

[17] C. Wang, K. V. Hindriks, and R. Babuska, "Active learning of affordances for robot use of household objects," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 2014, pp. 566–572.

[18] A. Gonçalves, G. Saponaro, L. Jamone, and A. Bernardino, "Learning visual affordances of objects and tools through autonomous robot exploration," in *Autonomous Robot Systems and Competitions (ICARSC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 128–133.

[19] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, "Affordances in psychology, neuroscience and robotics: a survey," *IEEE Transactions on Cognitive and Developmental Systems*, 2016.

[20] E. Ugur, Y. Nagai, E. Sahin, and E. Öztop, "Staged development of robot skills: behavior formation, affordance learning and imitation," *IEEE Transactions on Autonomous Mental Development*, 2015.

[21] C. Press, H. Gillmeister, and C. Heyes, "Sensorimotor experience enhances automatic imitation of robotic action," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 274, no. 1625, pp. 2509–2514, 2007.

[22] M. Hülse, S. Wischmann, P. Manoonpong, A. von Twickel, and F. Pasemann, "Dynamical systems in the sensorimotor loop: On the interrelation between internal and external mechanisms of evolved robot behavior," in *50 years of artificial intelligence*. Springer, 2007, pp. 186–195.

[23] R. D. Beer, "A dynamical systems perspective on agent-environment interaction," *Artificial intelligence*, vol. 72, no. 1-2, pp. 173–215, 1995.

[24] Z. Hasan and D. Stuart, "Animal solutions to problems of movement control: The role of proprioceptors," *Annual review of neuroscience*, vol. 11, no. 1, pp. 199–223, 1988.

[25] B. L. Riemann and S. M. Lephart, "The sensorimotor system, part ii: the role of proprioception in motor control and functional joint stability," *Journal of athletic training*, vol. 37, no. 1, p. 80, 2002.

[26] C. E. Tutin, R. Ham, and D. Wrogemann, "Tool-use by chimpanzees (pan t. troglodytes) in the lopé reserve, gabon," *Primates*, vol. 36, no. 2, pp. 181–192, 1995.

[27] C. Boesch and H. Boesch, "Tool use and tool making in wild chimpanzees," *Folia primatologica*, vol. 54, no. 1-2, pp. 86–99, 1990.

[28] A. H. Taylor, B. Knaebe, and R. D. Gray, "An end to insight? new caledonian crows can spontaneously solve problems without planning their actions," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 279, no. 1749, pp. 4977–4981, 2012.

[29] A. Auersperg, A. M. von Bayern, S. Weber, A. Szabadvari, T. Bugnyar, and A. Kacelnik, "Social transmission of tool use and tool manufacture in goffin cockatoos (cacatua goffini)," *Proc. R. Soc. B*, vol. 281, no. 1793, p. 20140972, 2014.

[30] R. Smolker, A. Richards, R. Connor, J. Mann, and P. Berggren, "Sponge carrying by dolphins (delphinidae, tursiops sp.): a foraging specialization involving tool use?" *Ethology*, vol. 103, no. 6, pp. 454–465, 1997.

[31] W. Staff, "Fish photographed using tools to eat," Jun 2017. [Online]. Available: https://www.wired.com/2011/07/fish-tool-use/

[32] V. Banschbach, A. Brunelle, K. Bartlett, J. Grivetti, and R. Yeamans, "Tool use by the forest ant aphaenogaster rudis: ecology and task allocation," *Insectes sociaux*, vol. 53, no. 4, pp. 463–471, 2006.

[33] J. K. Finn, T. Tregenza, and M. D. Norman, "Defensive tool use in a coconut-carrying octopus," *Current Biology*, vol. 19, no. 23, pp. R1069–R1070, 2009.

[34] R. S. Amant and A. B. Wood, "Tool use for autonomous agents." in *AAAI*, 2005, pp. 184–189.

[35] A. M. Arsenio, "Learning task sequences from scratch: applications to the control of tools and toys by a humanoid robot," in *Proceedings of the 2004 IEEE International Conference on Control Applications, 2004.*, vol. 1. IEEE, 2004, pp. 400–405.

[36] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[37] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[38] I. Kostrikov, "Pytorch implementations of reinforcement learning algorithms," https://github.com/ikostrikov/pytorch-a2c-ppo-acktr, 2018.

[39] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," *arXiv preprint arXiv:1709.06560*, 2017.

[40] S. Gilman, "Joint position sense and vibration sense: anatomical organisation and assessment," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 5, pp. 473–477, 2002.

[41] R. Reams and Y. Choe, "Emergence of tool construction in an articulated limb controlled by evolved neural circuits," in *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE, 2017, pp. 642–649.