

RLBench: The Robot Learning Benchmark & Learning Environment

Stephen James¹, Zicong Ma², David Rovick Arrojo², Andrew J. Davison¹

Abstract—We present a challenging new benchmark and learning-environment for robot learning: RLBench. The benchmark features 100 completely unique, hand-designed tasks ranging in difficulty, from simple target reaching and door opening, to longer multi-stage tasks, such as opening an oven and placing a tray in it. We provide an array of both proprioceptive observations and visual observations, which include rgb, depth, and segmentation masks from an over-the-shoulder stereo camera and an eye-in-hand monocular camera. Uniquely, each task comes with an infinite supply of demos through the use of motion planners operating on a series of waypoints given during task creation time; enabling an exciting flurry of demonstration-based learning. RLBench has been designed with scalability in mind; new tasks, along with their motion-planned demos, can be easily created and then verified by a series of tools, allowing users to submit their own tasks to the RLBench task repository. This large-scale benchmark aims to accelerate progress in a number of vision-guided manipulation research areas, including: reinforcement learning, imitation learning, multi-task learning, geometric computer vision, and in particular, few-shot learning. With the benchmark’s breadth of tasks and demonstrations, we propose the first large-scale few-shot challenge in robotics. We hope that the scale and diversity of RLBench offers unparalleled research opportunities in the robot learning community and beyond. Benchmarking code and videos can be found here¹.

I. INTRODUCTION

Robot manipulation systems broadly fall somewhere on a spectrum ranging from traditional, modular methods, that include object recognition, state estimation, and planning, to fully end-to-end approaches that leverage deep learning and large-scale data to learn a mapping from input observations directly to motor actions, with the intuition that the ‘traditional’ modules are embedded in the weights of a deep neural network. Driven by the successful combination of large-scale data [1] and deep learning algorithms in the field of computer vision [2], there is now a large body of work looking at increasing the capabilities of robotic agents through the use of reinforcement learning [3], [4], meta-learning [5], [6], [7], multi-task learning [8], [9], etc. However, there is currently no standard in place for comparing manipulation methods in these respective areas. Although there exist benchmarks such as **OpenAI Gym** [10] and **DeepMind Control Suite** [11] for evaluating continuous-control reinforcement learning algorithms, **their focus is not on real-world problems, and it is often the case that algorithms in these toy-benchmarks do not scale to more complex, real-world tasks.** Few-shot learning methods for robotics also suffer from a lack of well defined tasks; for example, in Finn *et al.* [5] and James *et al.*

[6] there is a very narrow distribution of tasks, where the task of “placing a peach into a red bowl” would be considered a different task to “placing an apple in to a green bowl”. Despite the increase in these data-driven approaches, it is not clear where the ideal location on this ‘learning’ spectrum lies for complex robotics tasks that we may one day want robots performing in our homes. Given all of these problems, there seems to be a need for a benchmark that evaluates not only the diverse range of robot learning fields that are now emerging, but also a range of visually-guided manipulation approaches from both sides of the spectrum.

This motivates the need for a one-size-fits-all benchmark that allows the capability to utilise large-scale data, whilst also allowing classical systems to be compared. To that end, we present RLBench, which is an ambitious large-scale benchmark and learning environment designed to facilitate research in a number of both classical and deep-learning based robot manipulation areas. RLBench is deliberately highly challenging and forward looking. The benchmark includes 100 completely unique, hand-designed tasks ranging in difficulty (shown in Figure 1), which share a common Franka Emika Panda robot arm, featuring a range of sensor modalities, including joint angles, velocities and forces, an eye-in-hand camera and an over-the-shoulder stereo camera setup. Each of the 100 tasks comes with a number of textual descriptions and an infinite set of demonstrations made possible through our task building tools that use waypoint-based motion planning.

In this paper, we discuss a host of research areas that would benefit from this benchmark, including, but not restricted to, reinforcement learning, imitation learning, few-shot learning, multi-task learning, and geometric based methods, such as SLAM. In addition to the benchmark, we also contribute an open-source set of tools that will allow rapid development of new tasks (through the use of PyRep [12]) in order to improve the size and scope of the benchmark over time. To summarise, RLBench has the following 3 key aims:

- Provide a benchmark and learning environment for both ‘robot learning’ and ‘traditional’ methods.
- Provide the a large-scale few-shot challenge, where given M training tasks and N unseen tasks, a system must take K different demonstrations of each of the N unseen tasks, and then be able to perform these tasks in new configurations.
- Provide a set of tools to allow easy task creation.

II. RELATED WORK

We review existing datasets, benchmarks, and learning environments that could be considered similar to ours in

¹Dyson Robotics Lab, Imperial College London

²UROP, Imperial College London

¹<https://sites.google.com/view/rlbench>

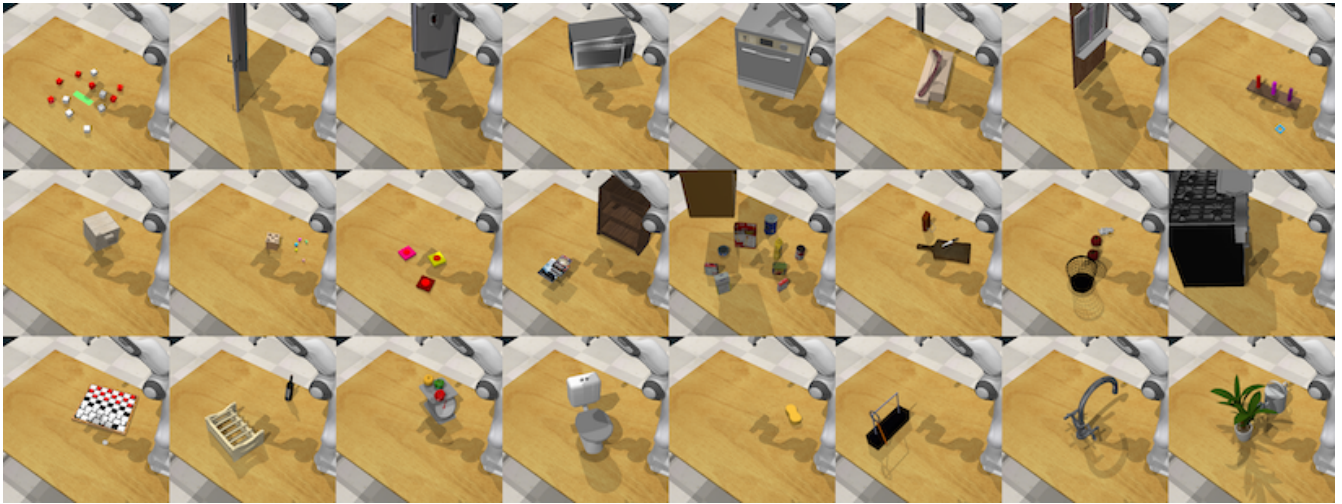


Fig. 1: RLBench is a large-scale benchmark consisting of 100 completely unique, hand-designed tasks. In this figure we show a sample of 24 tasks that feature in the benchmark. Example tasks include stacking a set of 6 colored blocks in a pyramid (top left), inserting a shape onto a peg (top right), finish setting up a checkers board (bottom left), and watering a plant (bottom right). To get a better understanding of the variety of tasks, please watch the video.

an effort to further motivate RLBench. Firstly we cover reinforcement learning benchmarks, followed by benchmarks designed specifically for manipulation.

a) Reinforcement Learning: Largely as a consequence of the seminal work that saw an algorithm learn to play a range of Atari 2600 video games to superhuman level directly from image pixels [13], deep reinforcement learning (DRL) has increasingly become prevalent in the literature, leading to a number of recent further success in the games of Go [14], Chess [15], StarCraft [16], and Dota [17]. With the success of these approaches, there has been a surge in developing DRL algorithms to solve continuous control environments [18], [19], [20], [21], [22]. These learned (continuous control) agents are usually tested on benchmarks such as OpenAI Gym [10] or the DeepMind Control Suite [11]. However, apart from a small number of robotic tasks in OpenAI Gym, these benchmarks feature only toy tasks that often do not resemble real-world problems that robots will need to overcome. To combat this, many projects create their own manipulation tasks to evaluate their approach, making comparisons difficult. As a direct consequence of this, these created tasks can often succumb to unintentionally introducing another hyperparameter into the method in the form of the task design itself. For example, a method could fail on a more challenging task, and so results would only be presented for a simpler set of tasks. This is something a standard benchmark of tasks could alleviate. (We should mention the very recently announced Meta-World project [23], a multi-task benchmark for meta-learning research in manipulation, though full documentation describing the aims of that project is not available at the time of writing.)

b) Manipulation: Most related work in benchmarking robot manipulation algorithms often concentrates on solving only one of the manipulation sub-problems, focusing on either perception, grasping, or planning. But first, we look

at benchmarks that evaluate the system as a whole. The Amazon Robotics Challenge (ARC) [24] was an attempt to create a benchmark for robotic picking and stowing. Although it was a successful challenge that drew many conclusions, such as the usefulness of a dual gripper and suction cup end-effector [25], it was difficult to reproduce in a lab setup. The ACRV Picking Benchmark [26] aimed to solve this by creating a similar, but reproducible setup to the ARC. The issue with picking and stowing is that it is but one of many possible tasks; RLBench on the other hand comes with 100 unique tasks, many of which involve some aspect of picking and placing. Similarly to ARC, the RoboCup@Home competition [27] is run annually, but has a greater range of tasks that must be completed. However, given that no large-scale data is given beforehand, this makes reinforcement learning and other end-to-end approaches difficult to apply in the competition. RLBench is a platform that can unify both old and new methods and compare them on an even playing field.

For evaluating imitation learning systems in particular, RoboTurk [28] was a recent attempt to leverage crowd sourcing to obtain data for tasks, but because of this the system has only three tasks. Whilst RoboTurk is entirely in simulation, Simitate [29] on the other hand is a hybrid approach, where real world observations (RGB-D camera calibrated against a motion capturing system) are combined with a simulated environment for benchmarking. In contrast to RoboTurk, we do not crowd source our demonstrations, but instead rely on an infinite supply of generated demonstrations collected via motion planners. Although Simitate offers the benefit of being partially a real-world dataset, the addition of new tasks requires time-consuming calibration and motion capturing; our system on the other hand sacrifices the real-world aspect, but in exchange we receive the ability generate a diverse range of tasks in a scalable way.

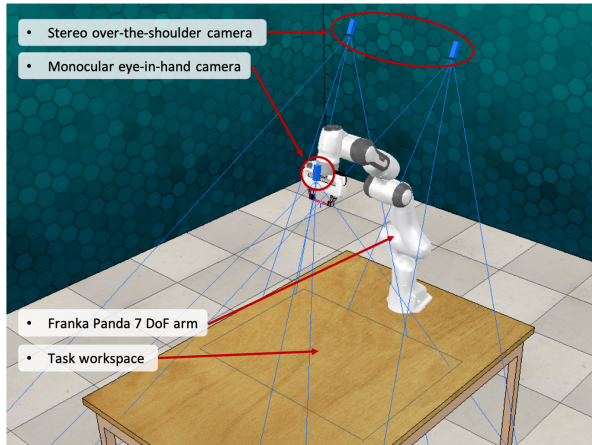


Fig. 2: The V-REP scene consists of a Franka Panda affixed to a wooden table, surrounded by 3 directional lights. Observations include rgb, depth, and segmentation masks from an over-the-shoulder stereo camera and a eye-in-hand monocular camera, along with robot proprioceptive data, which includes joint angles, velocities, and torques, and the gripper pose. The arm can be easily swapped out for another arm if required.

Moving on from whole-system benchmarks, there are a host of benchmarks that focus on sub-problems, for example perception datasets, from both the computer vision community (such as ILSVRC [1], COCO [30], Pascal-VOC [31], etc), and the robotics community (such as BigBIRD [32], YCB-Video [33], etc). For grasping, both OpenGrasp [34] and VisGrAB [35] are popular simulation-based benchmarks, whilst the YCB dataset [36] focuses on real-world objects. In comparison to these, RLbench allows robotic systems to be evaluated on the complete robotic pipeline, rather than limited to sub-problems such as object detection, state estimation, grasp selection, and planning.

III. BENCHMARK PROPERTIES

When designing RLbench, we have prioritised several key properties:

a) Diversity: Algorithms we develop should be general. In order to effectively learn inter-task relationships, a truly diverse range of tasks is needed to help avoid over-fitting.

b) Reproducibility: Reproducibility is challenging in robotics as each lab has their own robotic setup. Moving to simulation solves this, but at the risk of developing solutions that may not run as well in the real-world. However, with the rise of deep-learning methods becoming more prominent in robotics, we believe it is important to find the potential and limits of these methods in a controlled, reproducible environment.

c) Scale: Given the amount of data modern machine learning methods need, it is important to not only have a large collection of tasks, but also the ability to produce a large number of demonstrations from these tasks.

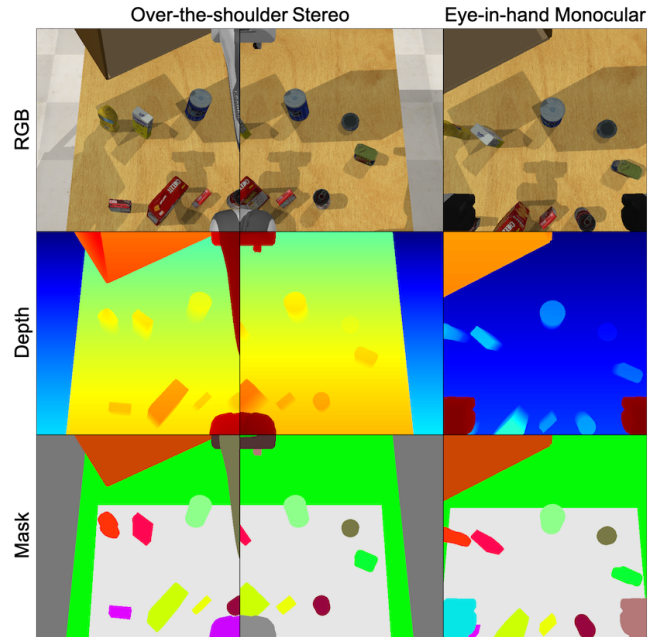


Fig. 3: A sample of the visual observations given from both the over-the-shoulder stereo and eye-in-hand monocular cameras, which supply rgb, depth, and mask images.

d) Extensibility: Following on from the previous point, we hope to continue to grow this repository of tasks. Therefore it is crucial that the task building system is as easy as possible to use. By leveraging the recently released robotics toolkit, PyRep [12], we are able to make a broad range of tasks in a short amount of time.

e) Tiered Difficulty: Attempting to get robots to do a single task can be challenging let alone expecting them to do numerous tasks. We therefore wanted to have a range of tasks, including both easy tasks, such as reaching, which would be well suited to new and emerging methods, to more challenging, long-time-horizon tasks that can stress-test well known state-of-the-art algorithms in use today.

f) Realism: Although we cannot claim full photorealism in our rendering system, or general realistic physics, we have put substantial effort into high quality components such as using a realistic robot model, graphics with lighting and shadows and a domain randomisation rendering option in order to maximise the potential for research on sim-to-real transfer.

IV. RLbench

RLbench is an ambitious project which we hope to grow over many years. The benchmark and learning environment is built around a V-REP [37] and PyRep [12] interface. PyRep is a toolkit for robot learning research, built on top of V-REP that features a number of improvements, including speed, rendering, and flexible a API for robot control and scene manipulation. Using the combination of these two libraries, we have been able to build this ambitious benchmark, which we now describe in greater detail.

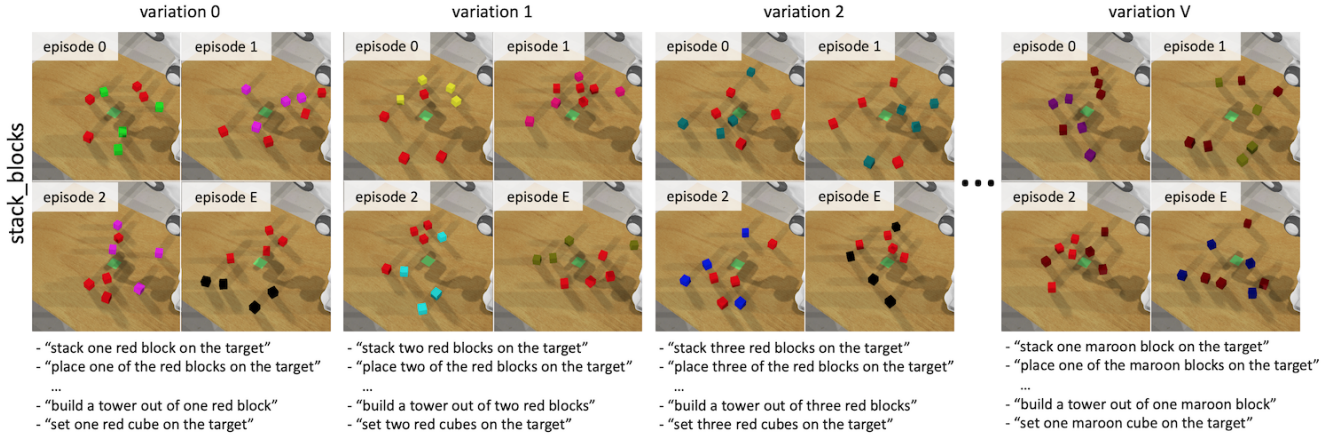


Fig. 4: An example showing the distinction between task, variation, and episode. In this case, the ‘*stack.blocks*’ task has V variations, each with E episodes. Each variation comes with a list of textual descriptions that describes the objective. Across variations, usually target objects or colours are changed, whereas across episodes positions are changed.

A. Scene

The V-REP scene, shown in Figure 2, remains constant across all tasks and contains the Franka Emika Panda 7 DoF arm affixed to a wooden table, surrounded by 3 directional lights. As shown in Figure 3, visual observations can be perceived from a stereo camera, and a monocular wrist camera, which supply rgb, depth, and segmentation mask data on each frame. In addition to visual observations, robot proprioceptive data can be retrieved, which includes joint angles, velocities, and torques, along with the end-effector pose. Tasks are loaded into the scene and placed at the centre of the workspace. Every task starts with the same assumption that no objects are held, therefore, unlike many works in the literature, tasks that involve tools will first need to grasp the object appropriately in order to accomplish the task. Although this makes the environments considerably harder to complete, we believe it is an important assumption to make given that household robots will one day work under such conditions.

B. Tasks, Variations & Episodes

RLBench employs 3 key terms: *Task*, *Variation*, and *Episode*. Each task consists of one or more variations, and from each variation, an infinite number of episodes can be drawn. Each variation of a task comes with a list of textual descriptions that verbally summarise this variation of the task, which could prove useful for human robot interaction (HRI) and natural language processing (NLP) research. A summary of this can be seen in Figure 4. Formally, we define an episode trajectory τ to consist of a series of observations \mathbf{o} and actions \mathbf{a} : $\tau = [(\mathbf{o}_1, \mathbf{a}_1), \dots, (\mathbf{o}_T, \mathbf{a}_T)]$. These episodes are sampled from a variation $\tau \sim \nu$. Finally, we define each task to be a set of variations, $\mathcal{T} = \{\nu_1, \dots, \nu_N\}$.

We now motivate the need for the concept of a ‘variation’ with an example. It is naturally difficult to come up with a precise way to differentiate between tasks given their subjective nature. For example, one could argue that “pick up the apple” and “pick up the banana” are different tasks,

whilst one could also equally argue that they are the same “pick up the X” task. We therefore introduce the variation concept, which allows cases like the above to be grouped as very similar tasks. Moreover, given the way the task building tools are designed (discussed in Section IV-E), the variation concept allows a convenient way of getting as much from a task definition as possible, given that there is usually only a small amount of additional work needed to generate a large number of variations for a given task.

C. Environment

Users will interface with the benchmark and learning environment through the *Environment* class. The *Environment* is the entry point and can spawn child environments, called *TaskEnvironment*, for the tasks you are interested in solving. The environment API, which Figure 5 demonstrates, is modelled after a typical agent-environment reinforcement learning setup. Each task has a completely sparse reward of +1 which is given only on task completion. Users have a wide variety of action spaces at their disposal, which include absolute or delta joint velocities, absolute or delta joint positions, absolute or delta joint torque, absolute or delta end-effector velocities, and finally absolute or delta end-effector poses.

D. Demonstrations

RLBench, through the task building tool mentioned in Section IV-E, provides expert algorithm π^* for each different task and their corresponding variations, allowing for demonstration episodes to be generated. The episodes produced via π^* come from using the Open Motion Planning Library [38].

E. Task Builder

Two common simulation environments in the literature today are Bullet [39] and MuJoCo [40]. However, given that these are physics engines rather than robotics frameworks, it can often be cumbersome to build rich environments and integrate standard robotics tooling such as inverse and

```

1 from rlbench.environment import Environment
2 from rlbench.action_modes import ActionMode
3 from rlbench.tasks import ReachTarget
4
5 DATASET = 'path/to/demo/dataset'
6
7 env = Environment(
8     DATASET, ActionMode.ABS_JOINT_VELOCITY)
9 env.launch()
10
11 task = env.sample_task()
12 demos = task.get_demos(2)
13
14 agent = Agent()
15 agent.ingest(demos)
16
17 training_steps = 100
18 episode_length = 100
19 obs = None
20 for i in range(training_steps):
21     if i % episode_length == 0:
22         descriptions, obs = task.reset()
23         action = agent.act(obs)
24         obs, reward, terminate = task.step(action)
25 env.shutdown()

```

Fig. 5: Example usage of the RLBench Environment for training a reinforcement learning agent. When using demonstrations, users can either point to a set of saved demonstrations (as shown here), or alternatively generate demonstrations on the fly.

forward kinematics, user interfaces, motion libraries, and path planners. Given the scale of RLBench, we needed a tool for designing tasks as easily as possible.

The task building tool is the interface for users who wish to create new tasks to be added to the RLBench task repository. Each task has 2 associated files: a V-REP model file (*.tm*), which holds all of the scene information and demo waypoints, and a python (*.py*) file, which is responsible for wiring the scene objects to the RLBench backend, applying variations, defining success criteria, and adding other more complex task behaviours. Figure 6 shows an example of how simple many tasks files can be.

In order to use the task creator, users must understand how tasks are initialised and placed in the scene. When a user asks for a new task from RLBench, the task is initialised by calling *init_task()*, and is only called once. Following that, *init_variation(int i)* is called at the beginning of each variation, and gets passed the variation number, which should be less than or equal to the number of variations for that task (which can be obtained by calling *variation_count()*). This function returns a list of strings which provide descriptions that could be associated with this variation of the task; an analysis of the frequency of words in these descriptions can be seen in top of Figure 7. Finally, *init_episode()* is called each time a new episode (of the same variation) is requested.

Once a task has been created, we provide a task validation tool, that attempts to collect a number of demonstrations of the designed task in order to ensure that the path planning aspect of the task only fails a small number of times. Once the validator passes, the user will be free to perform a

```

1 from rlbench.backend.task import Task
2 from rlbench.backend.conditions import
   DetectedCondition, GraspedCondition
3 from pyrep.objects.shape import Shape
4 from pyrep.objects.proximity_sensor import
   ProximitySensor
5
6 class TakeLidOffSaucepan(Task):
7
8     def init_task(self):
9         lid = Shape('saucepan_lid')
10        success_detector = ProximitySensor('success')
11        self.register_graspable_objects([lid])
12        cond_set = [
13            GraspedCondition(self.robot.gripper, lid),
14            DetectedCondition(lid, success_detector)
15        ]
16        self.register_success_conditions([cond_set])
17
18     def init_episode(self, index):
19         return ['take lid off the saucepan']
20
21     def variation_count(self):
22         return 1

```

Fig. 6: An example of a task python file. When using the task building tool, users are able to simultaneously edit the V-REP scene whilst also changing the various behaviour of a task. In this example, the task is to take a lid off of a saucepan. By interfacing with the scene using PyRep, we register that the episode should terminate and be considered a success only if the saucepan lid is detected by a proximity sensor and that the lid is being held. The backend handles the randomisation of the position of the task at the beginning of each episode.

GitHub pull request in order to contribute to the growing task repository.

V. THE RLBENCH FEW-SHOT CHALLENGE (v 1.0)

A big gap in the literature today is a means to evaluate and compare few-shot learning methods for robotics. We place particular emphasis on the few-shot regime, because much like humans, robots should have the ability to leverage knowledge from previously learned tasks in order to learn new ones quickly in new and unfamiliar environments. Despite this, most approaches in manipulation have focused on learning a single task, with a limited notion of generalisation, and no way of leveraging the knowledge to learn other tasks more efficiently.

The few pieces of work that perform few-shot learning in robotics [5], [6], [7] focused on a very narrow definition of task and often treat a variation of the same task as another task; for example, placing a peach into a red bowl would be considered a different task to placing an apple into a green bowl. In order to develop truly general algorithms, we feel that it is important to have a diverse range of tasks to train and test on. To that end, we propose the following challenge:

Given N unseen tasks, provide the system with K different demonstrations of each of the N tasks, and then evaluate the systems ability to perform these tasks in new configurations. Specifically, we suggest the following procedure:

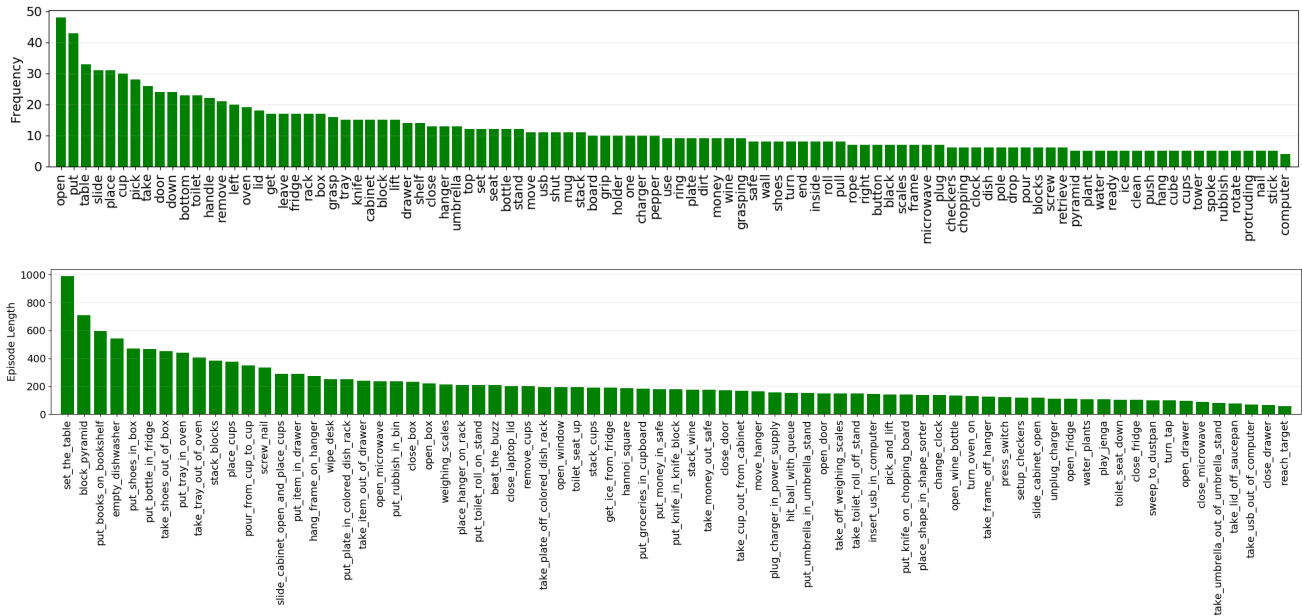


Fig. 7: **Top** shows the frequency of words in the variation descriptions with function words removed, leaving only content words. **Bottom** shows the average length of 5 demonstrations from a sample of 75 tasks (taken from the first variation). The tasks lengths vary from 100 to 1000 timesteps. Longer tasks usually involve many composed sets of actions, for example, the ‘empty_dishwasher’ task involves opening the washer door, sliding out the tray, grasping a plate, and then lifting the plate out of the tray. These long-horizon tasks can facilitate interesting research in reinforcement learning in robotic tasks.

- Of the 100 unique tasks, 10% of the tasks have been selected for the test set (meta-test) which span a range of difficulties, while the rest are chosen for training (meta-train). These train-test splits will be made available on the benchmark’s webpage.
- The training tasks can be used in any way desired by the user. RL Bench supplies a large number of pre-generated demos for each task that can be downloaded, although there is also the option to generate demos on the fly (or for users to create their own).
- During test time, the system is given K demonstrations of the unseen task (K -shot), and then success should be reported on new episodes of that same task. The only information available to the system should be the number of demos N and their corresponding observations. There must be no prior knowledge of the unseen tasks given to the system that are not included in the training tasks. Users report 1-shot, 5-shot, and 20-shot results for their method.

We purposefully call this challenge *v* 1.0 as we expect the number of tasks to grow considerably over the years; as this happens, we will create newer versions that span a broader range of tasks; therefore, we hope this versioning will ensure results remain meaningful and reproducible as the benchmark grows. State-of-the-art few-shot learning methods such as recurrent methods [41], [42], [43], metric learning methods [44], [45], and gradient based methods [46], [47] have not been tested on such a grand scale, and we look forward to seeing how they perform on this benchmark.

VI. OTHER APPLICATIONS & CHALLENGES

Further to the few-shot learning challenge highlighted in Section V, we briefly overview other areas of research that could benefit from RL Bench.

a) Reinforcement Learning: There is a large body of work in continuous control reinforcement learning that evaluate their algorithms on benchmarks such as OpenAI Gym [10] or DeepMind Control Suite [11]. Unlike these benchmarks, RL Bench has been tailored for visually-guided manipulation, which makes this an ideal platform for evaluating current and future reinforcement learning algorithms on real-world based tasks. Moreover, given the large number of demonstrations provided, it opens up the space to accelerate and facilitate research in bootstrapping reinforcement learning policies with demonstrations in order to reduce sample complexity. In addition, with the provided eye-in-hand camera observations, we open research in partial observability or incremental estimation for continuous control tasks.

b) Imitation Learning: Almost all imitation learning work design their own tasks for evaluating their method, making reproducibility difficult. A set number of demonstrations are shipped with RL Bench, but there is also the option in the framework to generate demonstrations on-the-fly, meaning that you can generate an infinite amount for your imitation learning algorithm.

c) Sim-to-Real Transfer: Recently there has been a large amount of work in learning control policies in simulation and then transferring these to the real world [48], [49], [50], [51], [52], [53]. The simulated Franka Panda within

RLBench can be easily swapped out, with one line of code, for another arm that researchers may have in their lab; this means that sim-to-real methods could be compared more easily on a standard set of tasks. Moreover, given the task-building tool and demonstration generation that RLBench has to offer, new tasks can easily be designed to demonstrate particular features in novel sim-to-real methods.

d) Multi-task Learning: In contrast to few-shot learning, multi-task learning concerns itself with learning several tasks simultaneously without particularly being expected to generalise to radically different tasks at test time. In this setup, all tasks from both meta-training and meta-testing can be used during training, and then during testing, the system must be able to generalise to unseen examples of those tasks. Given the difficulty of the challenge laid out in Section V, tackling the multi-task problem could provide valuable insights to increasing performance in the few-shot domain.

e) SLAM: Simultaneous Localisation and Mapping (SLAM) is concerned with constructing a map of an unknown environment while simultaneously keeping track of an agent’s location within it. Traditionally SLAM has been limited to navigation, virtual reality and augmented reality domains; but ultimately we can envision SLAM systems playing a key role in robots interacting with the world, i.e. a focus on more task-based SLAM. However, if we would like a manipulation system to make use of a SLAM map, it is not currently clear what the best way to represent this map is: whether it be sparse [54], [55], dense [56], [57], or semi-dense [58]. Moreover, it is not clear what level accuracy the map would need in order to achieve a desired task. RLBench could facilitate research in unifying SLAM and manipulation more tightly.

VII. SUMMARY AND FUTURE WORK

We have presented RLBench, an attempt to accelerate research in robotic manipulation that can be used in a broad range of robotic related research. We have posed the few-shot learning challenge for manipulation, and have highlighted a number of research areas that could benefit from this large scale benchmark and learning environment.

Given the scale of this project, we envision that there may be teething problems as people begin using the platform, and so we aim to maintain and continuously improve the benchmark during launch. Further to that, we hope, along with the help of the community, to continuously expand the tasks available for both training and evaluation. We hope RLBench will become a key resource for a broad range of robot manipulation related research, and look forward to seeing what the community achieves with this diverse range of tasks.

ACKNOWLEDGMENTS

We thank Juxi Leitner, Ankur Handa and Eugene Valasakis for insightful feedback on an early draft of this paper. Research presented here has been supported by Dyson Technology Ltd.

REFERENCES

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [3] S. James and E. Johns, “3d simulation for robot arm control with deep q-learning,” *arXiv preprint arXiv:1609.03759*, 2016.
- [4] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke *et al.*, “Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation,” *arXiv preprint arXiv:1806.10293*, 2018.
- [5] C. Finn, T. Yu, T. Zhang, P. Abbeel, and S. Levine, “One-shot visual imitation learning via meta-learning,” *Conference on Robot Learning*, 2017.
- [6] S. James, M. Bloesch, and A. J. Davison, “Task-embedded control networks for few-shot imitation learning,” *Conference on Robot Learning*, 2018.
- [7] T. Yu, C. Finn, A. Xie, S. Dasari, T. Zhang, P. Abbeel, and S. Levine, “One-shot imitation from observing humans via domain-adaptive meta-learning,” *Robotics: Science and Systems*, 2018.
- [8] C. Devin, A. Gupta, T. Darrell, P. Abbeel, and S. Levine, “Learning modular neural network policies for multi-task and multi-robot transfer,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 2169–2176.
- [9] K. Hausman, J. T. Springenberg, Z. Wang, N. Heess, and M. Riedmiller, “Learning an embedding space for transferable robot skills,” *International Conference on Learning Representations*, 2018.
- [10] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, “Openai gym,” *arXiv preprint arXiv:1606.01540*, 2016.
- [11] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. d. L. Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq *et al.*, “Deepmind control suite,” *arXiv preprint arXiv:1801.00690*, 2018.
- [12] S. James, M. Freese, and A. J. Davison, “Pyrep: Bringing v-rep to deep robot learning,” *arXiv preprint arXiv:1906.11176*, 2019.
- [13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, p. 529, 2015.
- [14] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, “Mastering the game of go with deep neural networks and tree search,” *nature*, vol. 529, p. 484, 2016.
- [15] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, no. 7676, p. 354, 2017.
- [16] DeepMind, “Alphastar: Mastering the real-time strategy game starcraft ii,” <https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii>, 2019.
- [17] OpenAI, “Openai five,” <https://blog.openai.com/openai-five/>, 2018.
- [18] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.
- [19] 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.
- [20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
- [21] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” *arXiv preprint arXiv:1801.01290*, 2018.
- [22] S. Fujimoto, H. van Hoof, and D. Meger, “Addressing function approximation error in actor-critic methods,” *arXiv preprint arXiv:1802.09477*, 2018.
- [23] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, S. Levine, and C. Finn, “Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning,” <https://meta-world.github.io>, 2019.

- [24] C. Eppner, S. Höfer, R. Jonschkowski, R. Martín-Martín, A. Sieverling, V. Wall, and O. Brock, "Lessons from the amazon picking challenge: Four aspects of building robotic systems," in *Robotics: Science and Systems*, 2016.
- [25] D. Morrison, A. W. Tow, M. McTaggart, R. Smith, N. Kelly-Boxall, S. Wade-McCue, J. Erskine, R. Grinover, A. Gurman, T. Hunn *et al.*, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7757–7764.
- [26] J. Leitner, A. W. Tow, N. Sünderhauf, J. E. Dean, J. W. Durham, M. Cooper, M. Eich, C. Lehnert, R. Mangels, C. McCool *et al.*, "The acrv picking benchmark: A robotic shelf picking benchmark to foster reproducible research," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4705–4712.
- [27] T. Wisspeintner, T. Van Der Zant, L. Iocchi, and S. Schiffer, "Robocup@ home: Scientific competition and benchmarking for domestic service robots," *Interaction Studies*, vol. 10, no. 3, pp. 392–426, 2009.
- [28] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," *arXiv preprint arXiv:1811.02790*, 2018.
- [29] R. Memmesheimer, I. Mykhalchyshyna, V. Seib, and D. Paulus, "Simitate: A hybrid imitation learning benchmark," *arXiv preprint arXiv:1905.06002*, 2019.
- [30] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [31] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.
- [32] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "Bigbird: A large-scale 3d database of object instances," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 509–516.
- [33] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," *Robotics: Science and Systems*, 2018.
- [34] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann, "The.opengrasp benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 1761–1767.
- [35] M. Popović, G. Kootstra, J. A. Jørgensen, D. Kragic, and N. Krüger, "Grasping unknown objects using an early cognitive vision system for general scene understanding," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 987–994.
- [36] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set," *IEEE Robotics & Automation Magazine*, vol. 22, no. 3, pp. 36–52, 2015.
- [37] E. Rohmer, S. P. Singh, and M. Freese, "V-rep: A versatile and scalable robot simulation framework," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1321–1326.
- [38] I. A. Sucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, 2012.
- [39] E. Coumans, "Bullet real-time physics simulation," <https://pybullet.org/wordpress/>, 2013.
- [40] 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*. IEEE, 2012, pp. 5026–5033.
- [41] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *International conference on machine learning*, 2016, pp. 1842–1850.
- [42] Y. Duan, J. Schulman, X. Chen, P. L. Bartlett, I. Sutskever, and P. Abbeel, "RI²: Fast reinforcement learning via slow reinforcement learning," *arXiv preprint arXiv:1611.02779*, 2016.
- [43] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," *arXiv preprint arXiv:1707.03141*, 2017.
- [44] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," in *Advances in neural information processing systems*, 2016, pp. 3630–3638.
- [45] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, 2017, pp. 4077–4087.
- [46] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1126–1135.
- [47] A. A. Rusu, D. Rao, J. Sygnowski, O. Vinyals, R. Pascanu, S. Osindero, and R. Hadsell, "Meta-learning with latent embedding optimization," *arXiv preprint arXiv:1807.05960*, 2018.
- [48] S. James, A. J. Davison, and E. Johns, "Transferring end-to-end visuomotor control from simulation to real world for a multi-stage task," *Conference on Robot Learning*, 2018.
- [49] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.
- [50] J. Matas, S. James, and A. J. Davison, "Sim-to-real reinforcement learning for deformable object manipulation," *arXiv preprint arXiv:1806.07851*, 2018.
- [51] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Science Robotics*, vol. 4, no. 26, p. eaau5872, 2019.
- [52] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige *et al.*, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4243–4250.
- [53] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, and K. Bousmalis, "Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 627–12 637.
- [54] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [55] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 15–22.
- [56] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *2011 international conference on computer vision*. IEEE, 2011, pp. 2320–2327.
- [57] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-Time Dense Surface Mapping and Tracking," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, 2011.
- [58] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European conference on computer vision*. Springer, 2014, pp. 834–849.