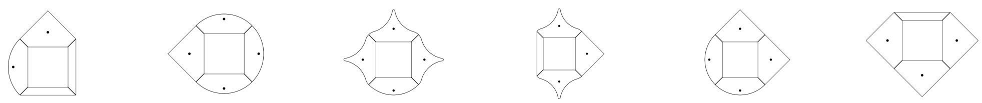
OmniPush: accurate, diverse, real-world dataset of pushing dynamics with RGBD images

Maria Bauza***,Ferran Alet*****, Tomás Lozano-Pérez, Leslie P. Kaelbling and Alberto Rodriguez**
MIT
`{bauza,alet,tlp,albertor,lpk}@mit.edu`



Dataset website: `mit.edu/mcube/omnipush-dataset/`

1 Introduction

Large and diverse datasets have driven recent progress in fields such as computer vision and natural language processing. We currently lack the equivalent richness for physical understanding. There are several large datasets (Yu et al., 2016; Levine et al., 2016; Calli et al., 2015), but they either only contain raw annotated video under highly varying conditions or only involve a small set of objects.

We introduce a new dataset of pushing dynamics, a fundamental robotics problem which has been studied extensively Mason (1986); Lynch et al. (1992); Bauza & Rodriguez (2017), yet still remains poorly characterized for general objects. We have gathered data resulting from a robot pushing planar objects on a surface, tracking their pose with high precision. Compared to Yu et al. (2016), who gather many pushes for 11 shapes, we perform 250 pushes for 250 objects, which vary both in terms of shape and weight distribution. RGBD images of the interactions is also available to complement the accurate recording of the object and pusher's positions.

We hope this dataset can be used both for the robotics and computer vision communities to improve and generalize the manipulation capabilities of robots. In particular, this dataset can serve as a benchmark for a variety of tasks including meta-learning (i.e. learning to learn), video prediction and model learning.

Figure 1: **Data capturing hardware.**

2 Dataset description

The dataset presented in this paper is based on collecting high-fidelity data resulting from the robot executing pushes on a diverse set of objects. We refer to this dataset as the OmniPush dataset inspired by the Omniglot dataset Lake et al. (2015) which contains a large variety of handwritten characters. In our dataset, one of the main novelties is the study of a manipulation task such as pushing for a large set of different objects.

The pushing system used for the data collection, shown in Fig. 1, is based on an industrial robot arm that pushes a given object over a flat surface. We introduce variations on the pushed objects to study how the dynamics of pushing change depending on the object, but constrain the rest of the setup to be the same during the experiments.

* equal contribution
32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

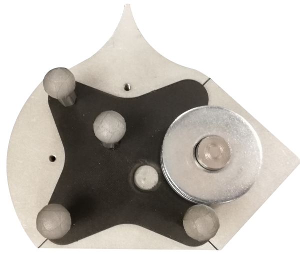


Figure 2: **Object made of the four possible sides.** Weights can be attached in the holes. There is an extra hole in the triangle, covered by the wafers. From right to bottom, the sides of the shape are triangle (94.1g), concave (74.1g), circle (67.5g) and rectangle (31.2g). Following our convention for naming objects, we denote this object as '2C1a3a4a'.

2.1 Shapes

The objects considered are built in a puzzle fashion from a set of aluminum pieces, as shown in Fig. 2. All objects share a square central piece (100.5g) that carries the Vicon trackers and thus they are all tracked equally with sub-millimeter accuracy. To this central piece, we magnetically attach 4 different sides so that they lay flat on the support surface and contribute to the frictional interactions. Each side is selected among four differently shaped pieces, leading to 70 different objects with nearly uniform mass distribution but different shape (Vicon markers and magnets have almost negligible weight). Figure 2 shows an example object made with each possible side. Each of the four shapes for the sides is denote with a different number `1-4`.

To add more diversity, we change the center of friction of some objects by altering their mass distribution. All sides except the rectangular one allow the addition of extra weight. We considered two different extra weights: small (`b`) and large (`c`). No extra weight is denoted as `a`. The triangular side allows the extra weight to be placed in two different positions (exterior weight is denoted as `B` or `C` and interior as `b` or `c`). The biggest weight is around 5 times heavier than the smallest side and we observed that adding extra weight clearly affects the dynamics of pushing.

The set of 250 objects explored consists of 3 groups, depending on the extra weights added:

- **70 objects without extra weight.** We explore *all* objects that can be assembled using 4 different sides and no extra weights, up to rotational symmetries.
- **90 objects with one extra weight.** We added only one extra weight (small or large) to one of the sides. To improve the structure of the dataset, we collected 45 objects with a small weight and the same 45 shapes with a large weight. For example, if we randomly selected the object `2C1a3a4a` shown in Fig. 2, then we also included the object `2c1a3a4a`.
- **90 objects with two small extra weights.** We randomly select objects made of the possible sides: `1a, 1b, 2a, 2b, 2B , 3a, 3b` and `4a` (without large weights) and take only those with two small weights. We get objects of the form: `1b1b3a2a` and `1a2B2a3b`.

2.2 Data collection

The data is collected with an ABB IRB 120 robot arm with a cylindrical pusher made of steel attached at its end effector. The surface where the frictional interaction happens is an ABS (Acrylonitrile Butadiene Styrene) sheet. The pusher's position is accurately recorded using the robot kinematics while the objects are tracked using Vicon cameras. The pushes are also recorded using a RealSense camera that provide RGBD images of the interaction. The data collection is autonomous and independent of the object. Given an object to explore, we collect data of its dynamics under pushing by following this scheme:

1. Put the pusher at a random position between 9 to 10cm from the object's center. This prevents the pusher from starting in a position that collides with the object.
2. Select a random direction and make a 5cm straight push for 1s (the interaction is close to quasistatic, meaning that the object stops moving as soon as the robot stops pushing it (Bauza & Rodriguez, 2017). Repeat 5 times without changing the pusher position between the end of one push and the start of the next.
3. Go back to step 1.

This scheme applies to all pushes unless the object ends outside a predefined region of the workspace. In that case, the robot pulls the object back to the center of the workspace and data collection continues at step 1. To increase the frequency of pushes with contact, we set a condition on how to randomly select the direction of each push. If there was contact in the previous push and the pusher starts from where it ended, then with probability 0.8 we select an angle that only deviates $\pm 90°$ from the previous pushing direction. Otherwise, the direction of pushing is uniformly sampled across all angles. If the pusher would end outside the workspace we resample the direction of pushing.

In total, we collected 250 pushes for each of the 250 objects, making more than 60K accurately recorded pushes. In addition, we also collected 2500 pushes for 5 of these objects and 1000 for the set of objects used in Yu et al. (2016) following the same process. We believe this dataset contains a diverse and complex set of examples that are sufficient to capture some of the most fundamental characteristics of pushing such as the effect of different pressure distributions and the effect of the shape of an object. The dataset is available at `mit.edu/mcube/omnipush-dataset/`.

2.3 Out-of-distribution objects and surface

The dataset also includes 1000 pushes collected with the proposed protocol for each of the 11 objects used in Yu et al. (2016). Similarly, we also collected 1000 pushes for 5 of the 250 our new objects and 5 of the 11 old objects from Yu et al. (2016). These sets are intended to give us an estimate of how much a given algorithm can generalize to related tasks that are outside the original distribution of tasks. Therefore in addition to the transfer between tasks inside the same distribution as previous meta-learning datasets, we can check for dataset distribution bias.

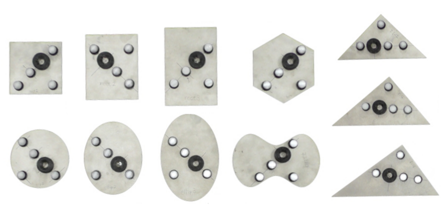


Figure 3: **Out-of-distribution objects.** These objects, made of a different material, stainless steel, are much heavier and have different shapes compared to the 250 objects used to build our meta-learning dataset. Their mass distribution is uniform. More details on these objects can be found at Yu et al. (2016).

2.4 Data description

The data from each push contains both RGBD images and information of the object and pusher states.

Poses: for every push we track x, y, θ of the object, x, y position of the pusher and v_x, v_y velocity of the pusher at a rate of 250Hz. A particularly interesting subset of this data includes the initial pose of both the object and the pusher along with the velocity of the latter and the final position and orientation of the object. Making the assumption that the result of the push does not depend on the absolute pose of the object, we can remove 3 dimensions by changing to the frame of the object (which sets $x = y = \theta = 0$ for the object); we can further remove another dimension by using the fact that the pusher moves at constant speed and therefore only represent its angle. Therefore, the target is always 3 dimensions $(\Delta x, \Delta y, \Delta\theta)$ and 3,5 and 7 inputs depending on the assumptions.

RGBD: We also extend these data by including RGBD recording of the pushes using a RealSense camera. This results in a complete dataset that can be studied directly from a vision perspective or by using the accurate positions recovered by the tracking system.

2.5 Preliminary evaluation

We focus on studying the poses part of the dataset mapping from 3 inputs (x, y position of the pusher with respect to the object and θ the angle of the relative velocity of the pusher to the object) to 3 outputs: change in x, y, θ of the object. Results can be seen in table 1. First, we estimated the Bayes error rate: we use 5 shapes with 2500 points each and train a separate neural network per shape, getting an error of 4% of the original variance. Bauza & Rodriguez (2017) determined 2500 pushes to be a point beyond which extra data helped little; therefore, it is likely the Bayes error rate (irreducible noise) is around that order of magnitude (we cannot estimate it exactly without infinite data). This noise can be due to small errors in actuation or sensing or non-uniformities of the surface which makes our dimension reduction incorrect.

We try two main approaches: an object-independent neural network that pools the data from all objects into a single dataset and a meta-learning approach that builds object-specific models. Meta-learning, or learning to learn, aims to leverage similarities between tasks to learn a learning algorithm that is very data-efficient in an unseen task. First, we can see that objects have roughly the same

Evaluation	MSE	distance equivalent
Predicting no movement	1.00	21.6 mm
Out of distribution obj.&surface (no meta-learning)	.51	15.4 mm
Out of distribution obj.&surface (Alet et al. (2018))	.34	12.5 mm
Out of distribution objects (no meta-learning)	.34	12.7 mm
Out of distribution objects (Alet et al. (2018))	.30	11.8 mm
Out of distribution surface (no meta-learning)	.21	9.8 mm
Out of distribution surface (Alet et al. (2018))	.09	6.6 mm
Original distribution (no meta-learning)	.14	8.1 mm
Original distribution (Alet et al. (2018))	.08	6.1 mm
Estimated Bayes error rate	.04	4.3 mm

Table 1: Summary of results. Different datasets are separated by horizontal lines.

movement, since an object-independent baseline (no meta-learning) captures 86% of the variance. For the meta-learning algorithm we use BounceGrad (Alet et al., 2018), which outperformed MAML on the similar pushing dataset of Yu et al. (2016), which almost halves the mean squared error.

There is still the challenge of generalizing with fewer than 50 points and generalizing to objects outside the distribution; for which performance was significantly better than chance but much worse compared to the within-distribution objects.

3 Conclusion

Our goal is that this dataset can serve as a benchmark for several applications both in computer vision and robotics. Some of the applications that we envision are:

- **Meta-learning.** By providing a large dataset with many different objects, we believe it is possible to learn the dynamics of a new object from few examples by using the data and learned model from previous objects.
- **Video prediction.** RGBD is recorded for all the interactions. This opens the possibility to predict the outcome of a given sequence of images using video prediction.
- **General dynamics learning.** Aside from learning each shape independently, we believe this dataset can help to build pushing models that generalize to different object shapes and mass distributions.

References

Ferran Alet, Tomás Lozano-Pérez, and Leslie P Kaelbling. Modular meta-learning. *Conference on Robot Learning*, 2018.

Maria Bauza and Alberto Rodriguez. A probabilistic data-driven model for planar pushing. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 3008–3015. IEEE, 2017.

Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Benchmarking in manipulation research: The YCB object and model set and benchmarking protocols. *arXiv preprint arXiv:1502.03143*, 2015.

Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

Sergey Levine, Peter Pastor, Alex Krizhevsky, and Deirdre Quillen. Learning hand-eye coordination for robotic grasping with large-scale data collection. In *International Symposium on Experimental Robotics*, pp. 173–184. Springer, 2016.

Kevin M Lynch, Hitoshi Maekawa, and Kazuo Tanie. Manipulation and active sensing by pushing using tactile feedback. In *Intelligent Robots and Systems (IROS)*, 1992.

Matthew T Mason. Mechanics and planning of manipulator pushing operations. *IJRR*, 5(3), 1986.

Kuan-Ting Yu, Maria Bauza, Nima Fazeli, and Alberto Rodriguez. More than a million ways to be pushed. a high-fidelity experimental dataset of planar pushing. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pp. 30–37. IEEE, 2016.