Learning Robotic Manipulation through Visual Planning and Acting

Angelina Wang
UC Berkeley
angelina.wang@berkeley.edu

Thanard Kurutach
UC Berkeley
thanard.kurutach@gmail.com

Aviv Tamar
UC Berkeley
aviv.tamar.mail@gmail.com

Pieter Abbeel
UC Berkeley
pabbeel@gmail.com

Abstract

Planning for robotic manipulation requires reasoning about the changes a robot can affect on objects. When such interactions can be modelled analytically, as in domains with rigid objects, efficient planning algorithms exist. However, in both domestic and industrial domains, the objects of interest can be soft, or deformable, and hard to model analytically. For such cases, we posit that a data-driven modelling approach is more suitable. Building on the recent Causal InfoGAN generative model, in this work we learn to imagine goal-directed object manipulation directly from raw image data of self-supervised interaction of the robot with the object. After learning, given a goal observation of the system, our model can generate an imagined plan – a sequence of images that transition the object into the desired goal. To execute the plan, we use it as a reference trajectory to track with a visual servoing controller, which we also learn from the data as an inverse dynamics model. In a simulated manipulation task, we show that separating the problem into visual planning and visual tracking control is more sample efficient and more interpretable than alternative data-driven approaches. We further demonstrate preliminary results on learning to imagine and execute deformable rope manipulation. A video of our work can be found at tinyurl.com/visualplanningacting

1 INTRODUCTION

The main difficulty in planning the manipulation of deformable objects is that, in contrast with rigid objects, there is no obvious mapping from an observation of the object to a compact representation in which planning can be performed. Thus, traditional task and motion planning approaches, which require manual design of the predicates, preconditions, and effects in the problem, are difficult to apply (McConachie et al. [2017], Srivastava et al. [2014]). In recent years, several studies have proposed a data-driven, self-supervised paradigm for robotic manipulation (Agrawal et al. [2016], Nair et al. [2017], Finn and Levine [2017]). In this approach, the robot 'plays' with the object using some random manipulation policy (e.g., randomly grasping or poking an object), and collects perceptual data about the interactions with the object. Later, machine learning is used to train a policy that performs the task directly from the perceptual inputs. By relying directly on data, these approaches overcome the modelling challenges of classical planning approaches, and scale to handle high-dimensional perceptual inputs such as raw images.

In this work we ask – can we learn to automatically *generate the visual plan and follow it* in a data-driven way? Concretely, given the current image of the system and some desired goal observation, we would like to generate a sequence of images that manipulate the object to the desired configuration,

Preprint. Work in progress.

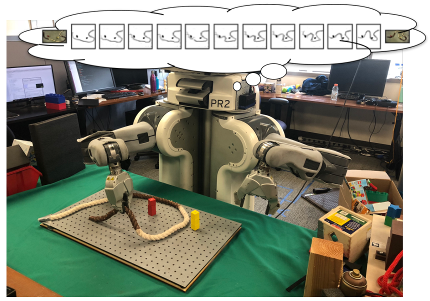

Figure 1: The PR2 robot first collects data through self-supervised random rope manipulation. Then, given a goal observation for the rope, we *plan* a visual trajectory of a possible manipulation sequence that reaches the goal (shown on top). Finally, visual servoing is used to execute the imagined plan.

without any human guidance, and then use this plan in conjunction with an inverse model for actually manipulating the object.

However, learning visual planning directly from raw image data has so far been limited to very simple planning tasks, such as reaching or pushing rigid objects Finn and Levine [2017], Ebert et al. [2017]. In this work, we take a step towards learning complex visual planning for robotic manipulation, by learning features that *are compatible* with a strong planning algorithm. At the basis of our approach is the recent NIPS 2018 work, the Causal InfoGAN (CIGAN) model from Kurutach et al. [2018]. In CIGAN, a deep generative model is trained to predict the possible next states of the object, with a constraint that linear trajectories in the latent state of the model produce feasible observation sequences. Kurutach et al. [2018] used a CIGAN model for planning goal-directed trajectories simply by linearly interpolating in the latent space, and then mapping the latent trajectory to observations for generating the visual plan.

Building on CIGAN, we propose a method for *visual planning and acting*, where sensory data obtained from self-supervised interaction is used to learn both a CIGAN model for visual planning and an inverse model for tracking a visual plan. After learning, given a goal observation for the system, we first use CIGAN to imagine a sequence of images that transition the system from its current configuration towards the goal. Then, we use the imagined trajectory as a reference for tracking using the inverse model.

We show that separating the control task into visual planning and visual tracking leads to an interpretable decision making paradigm, which is also more sample efficient than data-driven methods which learn actions directly from images. In addition, our contributions include (1) a context-conditional CIGAN model, (2) improved CIGAN planning using A* in latent state space, and most importantly, (3) showing that this approach is powerful enough to leave simulation and work on a PR2 rope manipulation domain.

2 Visual Planning and Acting

Our approach is model-based, where we first use the data $\mathcal{D}$ to learn both a CIGAN model M_{CIGAN} and an inverse dynamics model M_{IM}. For any two start and goal observations o_{start}, o_{goal}, the CIGAN model M_{CIGAN} can generate a visual plan that transitions the system from start to goal, $o_{start}, o_1, \ldots, o_k, o_{goal}$. Since the CIGAN model is trained to generate feasible pairs of observations, we are guaranteed that the plan generated by a well-trained CIGAN model will be feasible, in the sense that the robot can actually execute it.

Our Visual Planning and Acting (VPA) method for solving the goal directed planning problem is a combination of planning and replanning using the CIGAN model M_{CIGAN}, and trajectory tracking using the inverse model M_{IM}. The VPA algorithm is given as follows:

1. **Plan:** given a pair, o_{start}, o_{goal}, use the the CIGAN model M_{CIGAN} to generate a planned sequence of observations $o_{start}, o_1, ..., o_m, o_{goal}$.
2. **Act:** If the length of the plan m is zero, take an action u to reach the goal $u = M_{\text{IM}}(o_{start}, o_{goal})$, then stop. Else:

3. Take an action u to reach the first observation in the plan $u = M_{\text{IM}}(o_{start}, o_1)$ and take a new observation of the current system state o_{new}.
4. **Replan:** update o_{start} to be the current observation o_{new}, and go back to step 1.

The only data required is images taken from self-supervised manipulation of an object. Nevertheless, our method enjoys the **interpretability** of model based methods – at every step of our algorithm we have a visual plan of the proposed manipulation. We found that this allows us to reliably evaluate the performance of VPA before performing any robot experiment, significantly reducing robot-time and effort.

3 EXPERIMENTS

We demonstrate our method on two different domains. The first is a blocks world simulation in Mujoco from Todorov et al. [2012]. In this domain, we perform a comparison with batch off-policy RL – an alternative method for learning a control policy from data. The second domain is a real world rope manipulation problem with a PR2 robot that also includes several configurable obstacles.

3.1 Simulated Block Domain

Quantitatively, we evaluated VPA on 50 random initial and goal configurations that were not in the data, as shown in Table 1.

We compare VPA with an alternative data-driven approach based on model-free batch RL, namely, fitted Q-iteration Riedmiller [2005]. This is a strong baseline, that makes use of both the action-labeled and unlabled data, and incorporates several recent techniques for image-based RL. However, as stated earlier, RL is known to have difficulties with large state spaces (image), reward specification, and sample efficiency. To demonstrate this, we also run RL with several *artificial benefits*: (1) simple state space – true positions of the blocks, (2) true reward – based on real distance to target, and (3) more data – 30k action-labeled samples. Our results, reported in Table 1 show that, surprisingly, VPA significantly outperforms RL even with the artificial benefits.

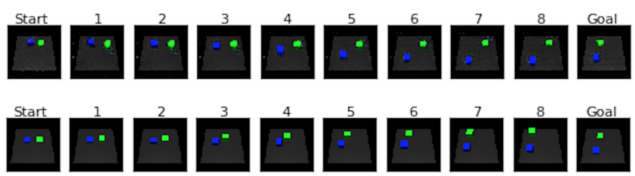

Figure 2: Top: imagined plan by Causal InfoGAN. Start and Goal image are both $o_{closest}$ to the actual o_{start} and o_{goal}, which are shown right below them. Bottom: actual successful results of running entire VPA pipeline on Mujoco

Table 1: The average final L2 distance to goal and the success rate to move two blocks to be within 0.5 radius to the goal when executed on 50 new tasks.

Method	L2 distance	Success Rate
VPA (2k)	**0.335 ±0.121**	**90%**
Batch RL (positions, real r, 2k)	0.657 ±0.701	76%
Batch RL (positions, real r, 30k)	0.675 ±0.739	74%
Batch RL (image, real r, 2k)	1.172 ±0.991	16%
Batch RL (image, real r, 30k)	1.186 ±0.940	42%
Batch RL (image, embedded r, 2k)	1.346 ±0.891	14%
Batch RL (image, embedded r, 30k)	1.445 ±1.096	18%

3.2 Real Robot Rope Manipulation Domain

We conduct experiments with a PR2 robot manipulating a flexible rope that is fixed on one end, and can move between two obstacles. This domain is inspired by wire threading – an important industrial task that is extremely challenging for autonomous robots. For data collection, we followed the approach in Nair et al. [2017], for generating random pokes of the rope, and collected 2k samples for observations and actions and an additional 10k for just observations.

In terms of success rate, we qualitatively inspected the plans and found that approximately 15% were visually accurate representations of rope manipulation. The most common failure cases are inaccurate encoding, leading to a misspecified goal image, or the rope breaking during the trajectory. We believe that more data and further improvements to C^3IGAN would significantly improve these results. From the visually correct plans, the inverse model was able to successfully execute 20%. This is somewhat worse than the results of Nair et al. [2017], which we attribute to the order of magnitude smaller data set we used, and our additional obstacles. We emphasize that even though our success rates are not high, our method is interpretable, and most failure cases can be caught by visual inspection, without running the robot. We see these results as a proof of concept for a new paradigm for robot manipulation.

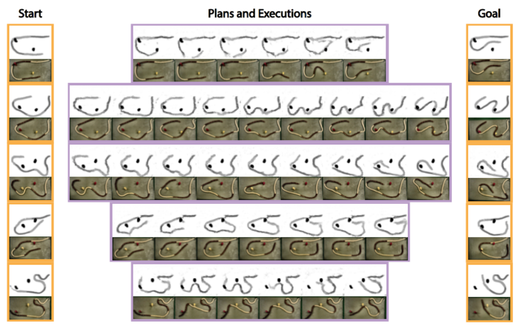

Figure 3: 5 examples of VPA on the rope domain. The top 4 are successful runs, and the bottom 1 is where a plan is generated to reach the goal, but the action policy is not strong enough to carry it out. Looking at one row at a time, the left image is the start state and the right is the goal state. In the middle, the grayscale images are the visualized plan, and the colored images are the actual results of the rope when we run the inverse model to have the PR2 take actions.

References

P. Agrawal, A. Nair, P. Abbeel, J. Malik, and S. Levine. Learning to poke by poking: Experiential learning of intuitive physics. 2016.

F. Ebert, C. Finn, A. X. Lee, and S. Levine. Self-supervised visual planning with temporal skip connections. pages 344–356, 2017.

C. Finn and S. Levine. Deep visual foresight for planning robot motion. pages 2786–2793, 2017.

T. Kurutach, A. Tamar, G. Yang, S. Russell, and P. Abbeel. Learning plannable representations with causal infogan. *arXiv preprint arXiv:1807.09341*, 2018.

Dale McConachie, Mengyao Ruan, and Dmitry Berenson. Interleaving planning and control for deformable object manipulation. In *International Symposium on Robotics Research (ISRR)*, 2017.

Ashvin Nair, Dian Chen, Pulkit Agrawal, Phillip Isola, Pieter Abbeel, Jitendra Malik, and Sergey Levine. Combining self-supervised learning and imitation for vision-based rope manipulation. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 2146–2153. IEEE, 2017.

Martin Riedmiller. Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method. In *European Conference on Machine Learning*, pages 317–328. Springer, 2005.

Siddharth Srivastava, Eugene Fang, Lorenzo Riano, Rohan Chitnis, Stuart Russell, and Pieter Abbeel. Combined task and motion planning through an extensible planner-independent interface layer. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 639–646. IEEE, 2014.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033. IEEE, 2012.