Learning What You Can Do Before Doing Anything

Oleh Rybkin*, Karl Pertsch*, Konstantinos G. Derpanis, Kostas Daniilidis, Andrew Jaegle

Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab

Ryerson University

USC University of Southern California

Idea

Task: We want to learn a semantic latent space z of actions u that an agent can execute, as well as what will be the consequences of these actions in our observation space.

Setup: We do not have access to actions, but can observe trajectories executed by the agent. Specifically, we use a dataset of videos of the agent.

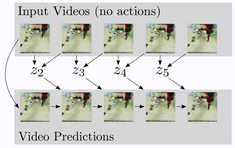
Assumptions (in this work): The environment dynamics in our observation space are fully determined by the executed actions and history of observations.

Idea: Use variational autoencoders to model the distribution of possible actions.

Application: We use the model for Model Predictive Control, similarly to [6]. However, we need less active observations to train, and can potentially leverage e.g. Internet videos for training.

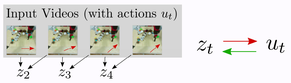
Approach

Passive (Unsupervised) Learning



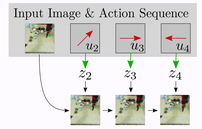
We learn a representation of actions z with stochastic video prediction and the proposed composability objective.

Active Learning (Calibration)



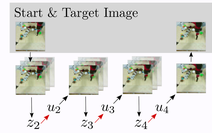
With few action annotations, we calibrate our representation to the robot at hand to perform prediction or control.

Action-Conditioned Video Prediction



We use the action annotations to learn a bijection between z and u, which we can use for action-conditioned prediction

Planning in learned action space



We can perform planning in the learned action space z, and decode the plan into grounded actions u to execute them.

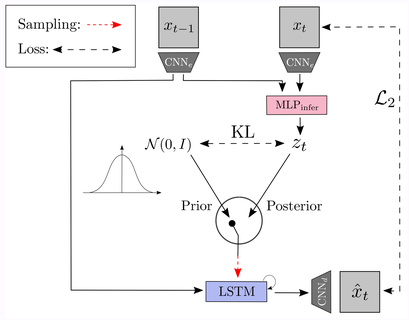
Background

Stochastic video prediction

We use a recurrent latent variable z_t to represent the distribution of possible future frames given a history of frames.

z_t is learned in a recurrent conditional variational autoencoder framework [1,2] with a reconstruction and a Kullback-Leibler divergence loss.

Balancing the two losses with a β term allows to recover a minimal representation [4].



Variational Information Bottleneck (VIB)

The Information Bottleneck [3] objective for a representation $\boldsymbol{Z}$ with input $\boldsymbol{X}$ and output $\boldsymbol{Y}$:

$$\max I(Z, Y) \text{ s.t. } I(X, Z) \leq I_c.$$

Variational Information Bottleneck [4] optimizes the above using the Lagrangian:

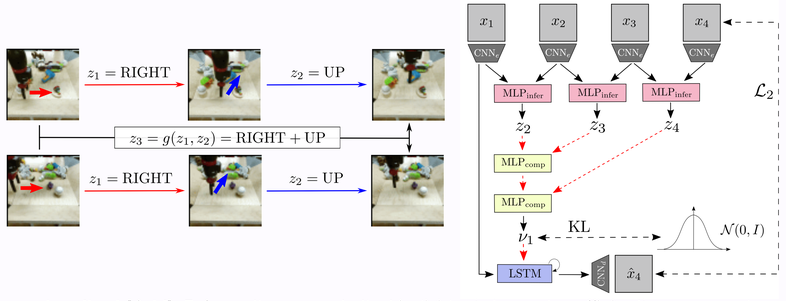
$$\mathbb{E}_{p(z|x)} \log\, q(y|Z) - \beta \mathrm{KL}[p(Z|x)||p(Z)].$$

Applied to the recurrent latent variable defined above, we recover the β-VAE [1,2,5] formulation:

$$\sum_t \left[\mathbb{E}_{p(z_t|x_{t-1})} \log\, q(x_t|Z_t, x_{t-1}) - \beta \mathrm{KL}[p(Z_t|x_{t-1}), p(Z)]\right]$$

Composability Training

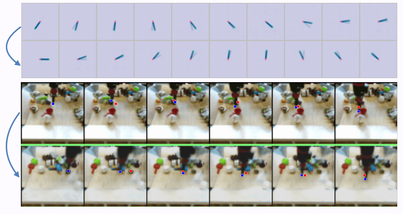
Problem (left): The learned representation z is *entangled* with the immediate state x, meaning that z does not represent the action consistently across trajectories.

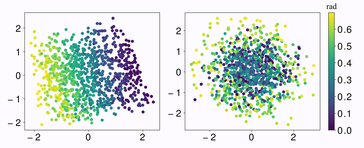


Proposed method (right): Enforce the property that individual z's can be efficiently composed.

Learned Disentaglement

Trajectory transplantation





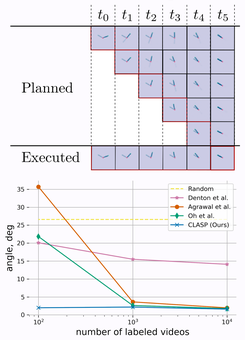
PCA of the latent samples z colored by the value of the true action u.

Action-conditioned prediction

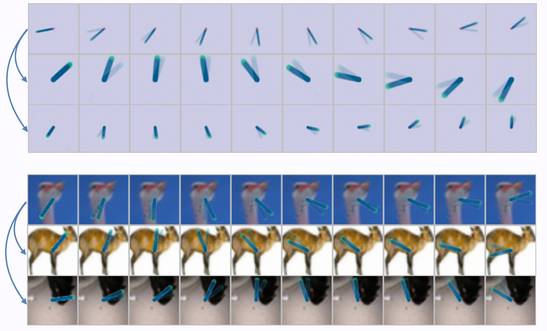
	Reacher	BAIR
Method	Error [deg]	Error [px]
Random	26.6 ± 21.5	-
Denton & Fergus	22.6 ± 17.7	3.6 ± 4.0
CLASP (Ours)	$\mathbf{2.9 \pm 2.1}$	$\mathbf{3.0 \pm 2.1}$
Supervised	2.6 ± 1.8	2.0 ± 1.3

Learning from Different Agents

Planning



Trajectory transplantation



Results summary

Even though the problem is underconstrained, an agent's action space can be learned from visual observations without explicit supervision by using suitable *inductive biases*.

Minimality and *composability* of action representations provide strong inductive biases for this task.

The key challenge is to learn a representation that is *disentangled* from the static scene content, such as the robot's immediate state, visual characteristics, and background.

The system can be trained with *passive observations*, e.g., from videos collected online.

The disentanglement allows the representation to be used for *action-conditioned prediction* and *planning* after a calibration phase with a small number of action-labeled observations.

References

[1] Denton, E. and Fergus, R., Stochastic Video Generation with a Learned Prior, in *ICML*, 2018.
[2] Lee, A., Zhang, R., Ebert, F., Abbeel, P., Finn, C. and Levine, S., Stochastic Adversarial Video Prediction, *arXiv*:1804.01523, 2018.
[3] Shwartz-Ziv, R. and Tishby, N., Opening the Black Box of Deep Neural Networks via Information, *arXiv*:1703.00810, 2017.
[4] Alemi, A., Fischer, I., Dillon, J. and Murphy, K., Deep Variational Information Bottleneck, in *ICLR*, 2017.
[5] Higgins, I, Matthey, L, Pal, A, Burgess, C, Glorot, X, Botvinick, M, Mohamed, S and Lerchner, A, β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework, in *ICLR*, 2017.
[6] Finn, C and Levine, S, Deep Visual Foresight for Planning Robot Motion in *ICRA*, 2017.