Symmetry-Based Disentangled Representation Learning requires Interaction with Environments

Hugo Caselles-Dupré
caselles@ensta.fr

Michael Garcia-Ortiz
mgarciaortiz@aldebaran.com

David Filliat
david.filliat@ensta.fr

Flowers Laboratory (ENSTA ParisTech - INRIA), AI Lab (Softbank Robotics Europe)

Introduction

Context: Finding a generally accepted formal definition of a disentangled representation for an agent behaving in an environment.

Higgins et al. proposed Symmetry-Based Disentangled Representation Learning, a definition based on a characterization of symmetries in the environment using group theory.

Intuition: Focus on transformations that change some properties of the underlying world state, while leaving all other properties invariant.

Problem: How to learn Symmetry-Based disentangled representations in practice?

Contributions:

- Theoretical and empirical arguments that proves SBDRL cannot only be based on fixed data samples.
- Guidelines on how to learn SB-disentangled representations in practice.

Definition of SBD-representations

Transformations that change properties of the underlying world state, while leaving all other properties invariant.

Formal definition of a SBD-representation $f : W \to Z$ w.r.t to a group decomposition $G = G_1 \times .. \times G_n$ of the world's symmetries:

- There is a group action $\cdot_Z : G \times Z \to Z$.
- f is equivariant between the group actions on W and Z.
- There is a decomposition $Z = Z_1 \times .. \times Z_n$ such that each Z_i is fixed by the action of all $G_j, j \neq i$ and affected only by G_i.

$$g \cdot_Z f(w) = f(g \cdot_W w) \quad \leftrightarrow \quad \begin{array}{ccc} G \times W & \xrightarrow{\cdot_W} & W \\ id_G \times f \downarrow & & \downarrow f \\ G \times Z & \xrightarrow{\cdot_Z} & Z \end{array}$$

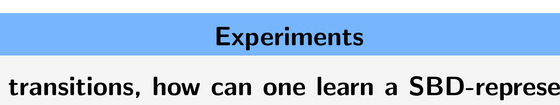
Figure 1: Illustration of SBDRL definition.

Main result

The main result proves that interaction with environments is necessary to learn SBD-representations (see paper for formal theorems).

Theorem 1. *Worlds with different physics can produce the same training set of still observations.*
It is thus impossible to reliably learn the symmetries effect on the world using only still observations.

The world's transition function is not learnable using only still observations.



Experiments

Using transitions, how can one learn a SBD-representation in practice?

We provide practical guidelines on a simple environment. We show how to learn linear and non-linear SBD-representations.

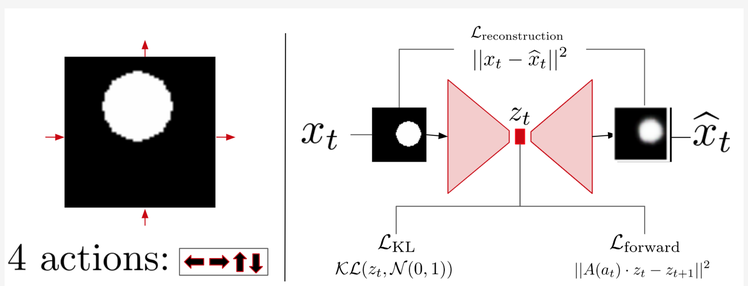


Figure 2: **Left:** Environment studied in this paper. **Right:** Proposed architecture for joint learning of the group action and state representation.

There are two options: **decoupled** or **joint** learning of the group action and state representation.
Both options uses transitions instead of still images (Theorem 1).

Experimental results

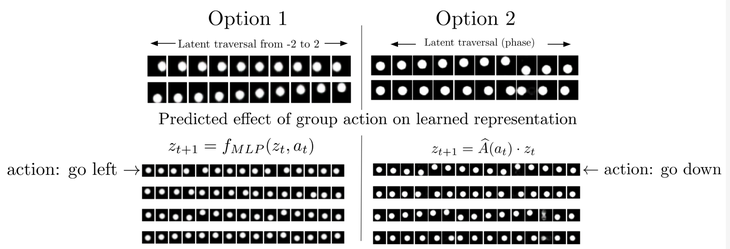


Figure 3: **Left:** First option: decoupled learning of SB-disentanglement. Latent traversal spanning from -2 to 2 over each of the representation's dimensions, followed by the (here, non-linear) predicted effect of the group action associated each action (left, right, down, up). **Right:** Second option: joint learning of LSB-disentanglement. The representation is complex: latent traversal over the phase of each of the representation's dimensions, followed by the predicted linear effect of the group action associated each action (down, left, up, right).

Both approaches are viable for learning (L)SBD-representations.

Usefulness for subsequent tasks

Are SBD-representations useful in practice?

Is it increasingly better to use non-disentangled/non-linear SBD/LSBD representations for downstream tasks?

Example on inverse model task →

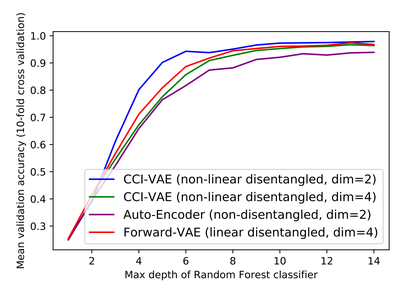


Figure 4: Downstream task evaluation of representation models: inverse model prediction. Mean 10-fold cross validation accuracy as functions of dataset size and classifier capacity (max depth parameter of Random Forest).

Paper and code

arXiv.org 1904.00243

Caselles/NeurIPS19-SBDRL