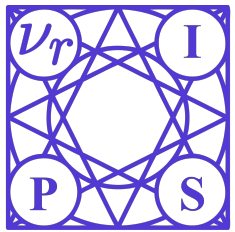
“Symmetry-Based Disentangled Representation Learning requires Interaction with Environments”

Hugo Caselles-Dupré, Michael Garcia-Ortiz, David Filliat

Flowers Laboratory, INRIA & ENSTA ParisTech
AI Lab, Softbank Robotics Europe



Neural Information Processing Systems 2019

Plan

1) Recap on Symmetry-Based Disentangled Representation Learning (SBDRL)

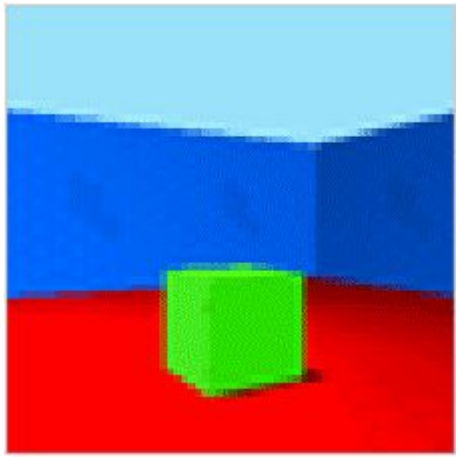
   I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

2) Our contributions: mainly SBDRL requires transitions and not still images

3) Discussion & future work

Motivation

- Representation learning: vectorial representations of data
- Disentanglement: isolate factors of variation (latent variable <-> high-level features)

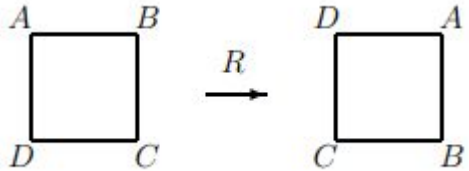


- Problem: disentanglement needs a proper definition, otherwise confusing

I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

Solution: Symmetries

- Which transformations change some properties of the underlying world state, while leaving all other properties invariant?



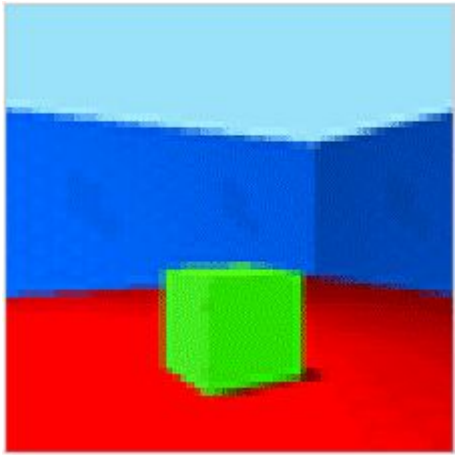
- Successful approach in physics: symmetries have revolutionized the understanding of the world structure.

I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

Why group theory?

- Group theory: the symmetry group of an object (image, signal, etc.) is the group of all transformations under which the object is invariant.

- Ex, scene understanding: transformations include translations, rotations and changes in object colour.



3Dshapes dataset

I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

Group theory

Group definition: A group is a set and operation (G, .) s.t. For all a,b,c in G:

- Closure: ab is in G.
- Associativity: (ab)c = a(bc)
- Identity element: e s.t. ea = ae = e
- Inverse element: $aa^{-1} = e$

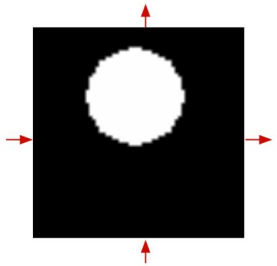
Group action: G acts on a set X through a group action $\cdot$ s.t.:

$$\forall x \in E,\ e \cdot x = x$$
$$\forall (g, g') \in G^2,\ \forall x \in E,\ g' \cdot \underbrace{(g \cdot x)}_{\in E} = \underbrace{(g' g)}_{\in G} \cdot x.$$

Disentangled representation definition I

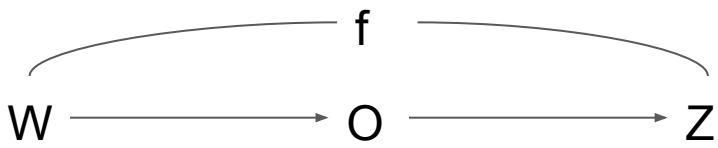
- World: observations O of world states W.
- Symmetries: translations.

O = images, W = (x,y), G = translations



4 actions: ⬅➡⬆⬇

- Goal: disentangled encoder f (projects W on Z)



I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

Disentangled representation definition II

- G applies an action on W: $g_x \cdot_W w = ((x+1) \mod N, y)$

  → This action should be the same on Z. $g \cdot_Z f(w) = f(g \cdot_W w)$

Defines Symmetry-Based representations

- Hypothesis: G can be decomposed into sub-groups that do not affect each other.

  → A sub-group only acts on a subspace of the representation.

Defines disentanglement

I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

Disentangled representation definition III

Definition:

Defines Symmetry-Based representations

1. There is an action $\cdot : G \times Z \to Z$,
2. The map $f : W \to Z$ is equivariant between the actions on W and Z, and

3. There is a decomposition $Z = Z_1 \times ... \times Z_n$ or $Z = Z_1 \oplus ... \oplus Z_n$ such that each Z_i is fixed by the action of all $G_j, j \neq i$ and affected only by G_i.

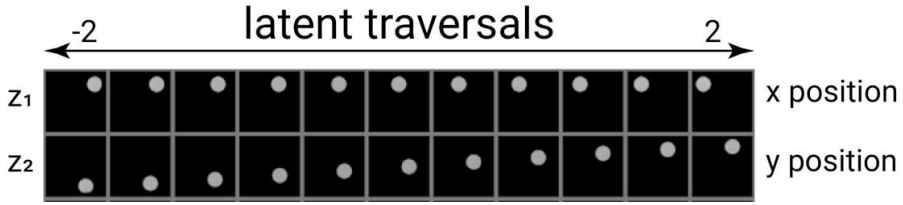
Defines disentanglement

I. Higgins et al., "Towards a Definition of Disentangled Representations", 2018.

In practice?

- I. Higgins et al. use CCI-VAE with still samples.

- Manage to learn: f(w)=f((x,y))=(λ1*x, λ2*y).

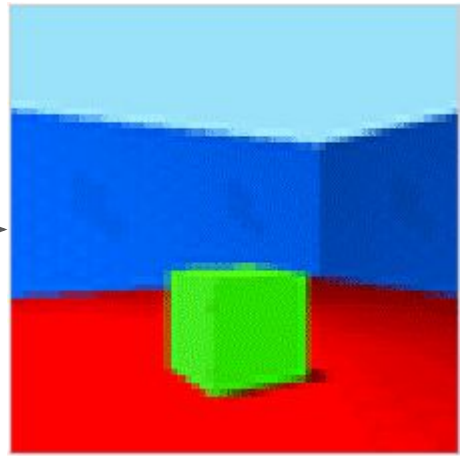


- But what about the group action? $g \cdot_Z f(w) = f(g \cdot_W w)$

Main result: intuition

SBDRL requires transitions instead of still samples.

The specific order in which the transitions happen, i.e. the transition function of the world or “physics” is not learnable using only still samples.



Main result: formulation

SBDRL requires transitions instead of still samples.

Theorem 1. *Suppose we have a SB representation* $(f, \cdot_Z)$ *of a world* $\mathcal{W}_0 = (W = (w_1, .., w_m) \in \mathbb{R}^{m \times d}, \cdot_{\mathcal{W}_0})$ *w.r.t to* $G = G_1 \times ... \times G_n$ *using a training set* $\mathcal{T}$ *of unordered observations of* $\mathcal{W}_0$*. Let* W_k *be the set of possible values for the* k^{th} *dimension of* $w \in W$*.*
Then:

1. *There exists at least* $k_{W,G} = n[(\min_k(card(W_k))!] - 1$ *worlds* $(\mathcal{W}_1, .., \mathcal{W}_{k_{W,G}})$ *equipped with the same world states* $\mathcal{W}_i = (w_1, .., w_m)$ *and symmetries* G*, but different group actions* $\cdot_{\mathcal{W}_i}$*.*
2. *For these worlds,* $(f, \cdot_Z)$ *is not a SB representation.*
3. *These worlds can produce exactly the same training set* $\mathcal{T}$ *of still images.*

Practical options

How to learn SB-disentangled representations in practice?

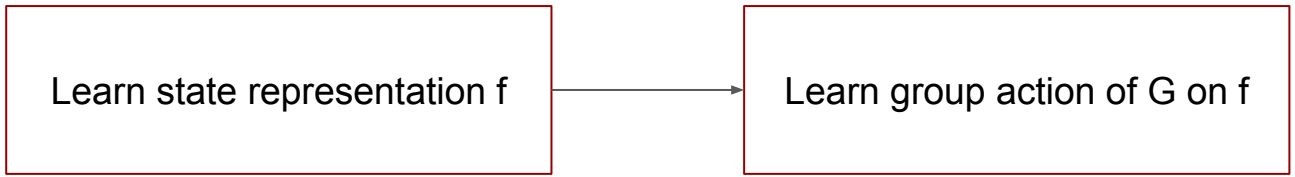
2 options arise

Option 1: a la World Models

Definition requires:

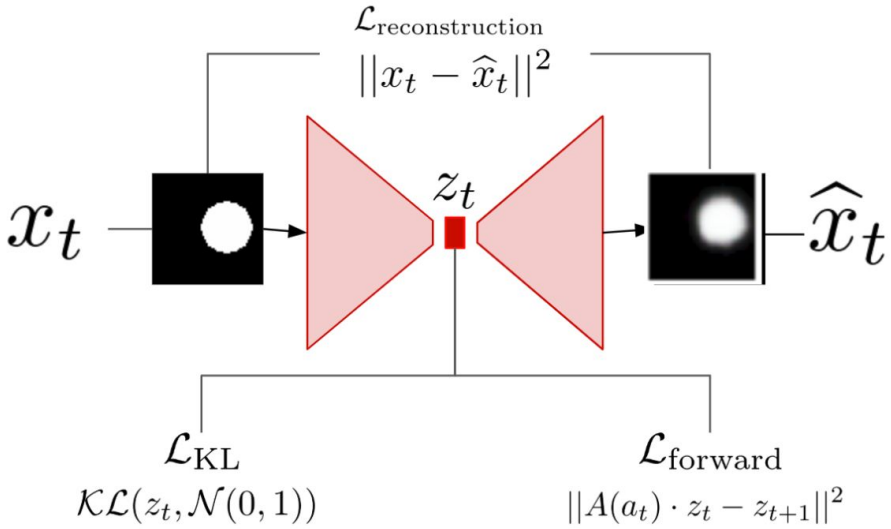
$$g \cdot_Z f(w) = f(g \cdot_W w)$$

Two-steps approach:



Option 2: End-to-end learning

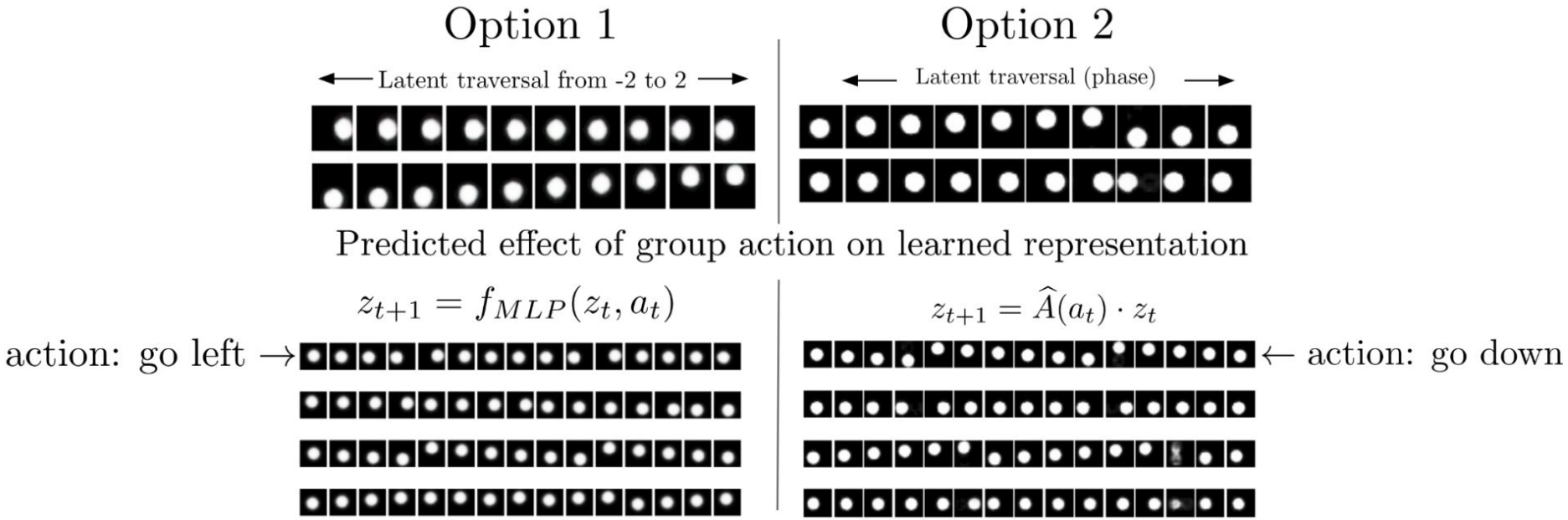
Learn state representation f and group action at the same time



Results

Both approaches are successful empirically.

Option 2 makes more sense as latent space has to organize specifically.

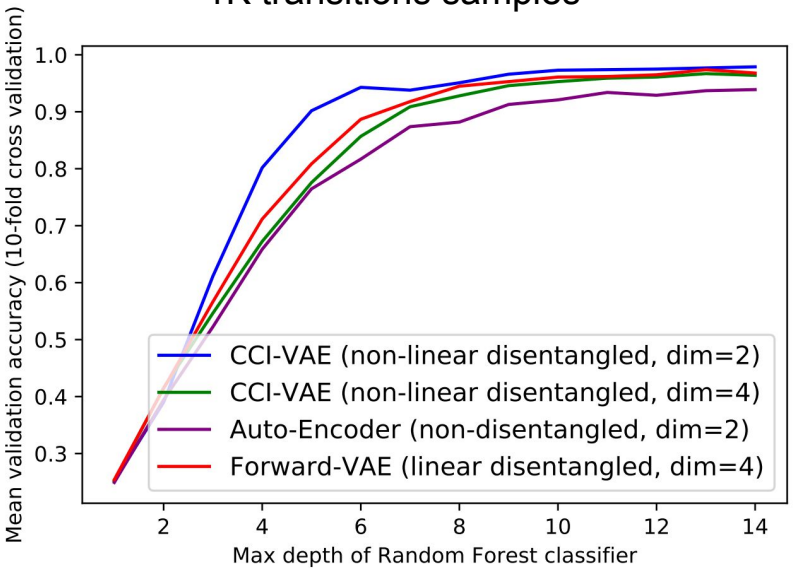


Usefulness of learned representations

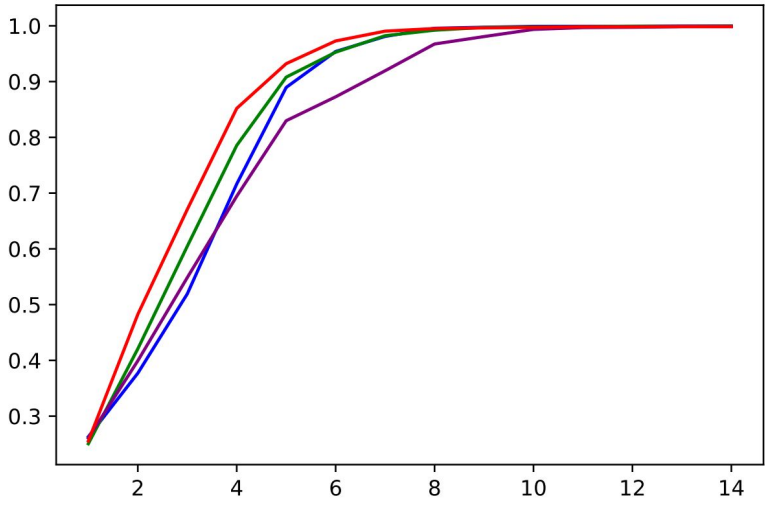
Learning downstream tasks should be easier.

Preliminary experiments: Inverse model $a_t = f_{inverse}(s_t, s_{t+1})$

1K transitions samples



10K transitions samples



Open question

How to learn SBD representations in more complex environments?

- High-level actions might be associated to a symmetry.

- Local symmetries.

Conclusion

- Formal definition of disentanglement is needed.

- SBDRL: Disentanglement is defined w.r.t a decomposition of the symmetry group of the world.

Learning a SB-disentangled representation requires transitions instead of still samples.

Thanks!

Code (Colab)

Arxiv

Contact :

casellesdupre.hugo@gmail.com

@Caselles

Site