Towards a Definition of Disentangled Representations

Irina Higgins*, David Amos*, David Pfau, Sebastien Racaniere,
Loic Matthey, Danilo Rezende, Alexander Lerchner
DeepMind
{irinah,davidamos,pfau,sracaniere,
lmatthey,danilor,lerchner}@google.com

December 7, 2018

Presented by: Hengtao Guo
01/08/2020

Contents

Motivations

1. Disentangled representation helps separate the structure of the learnt object
2. People are not agreeing on the definition of disentanglement
3. This paper uses symmetry transformation and group theory to formalize a definition of the disentangled action and representation

Introduction

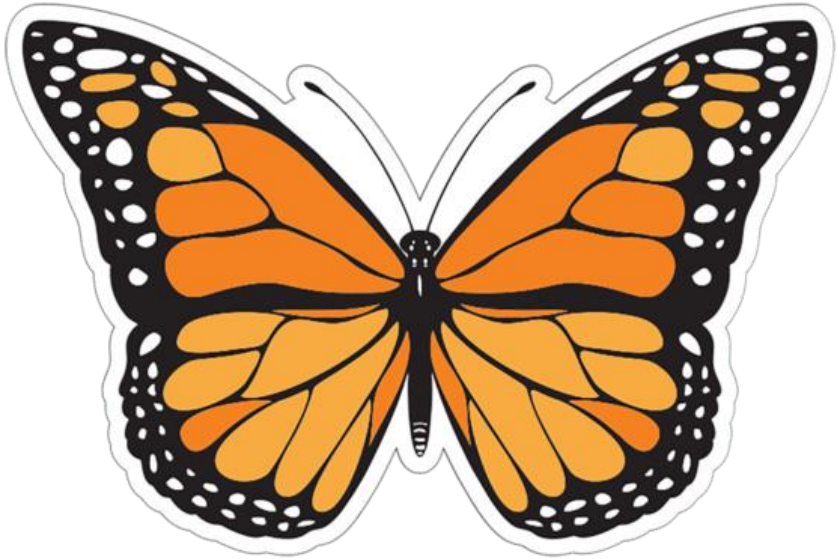
1. In ML, best models' performance often lacks the same level of robustness and generalizability.
2. Introducing certain inductive biases into the model that can reflect the structure of the underlying data. (symmetry of convolutions)
3. **How about learn a representation that is faithful to the underlying data structure?**
4. Intuitively, we define a vector representation as disentangled, if it can be decomposed into a number of subspaces, each one of which is compatible with, and can be transformed independently by a unique symmetry transformation

Introduction

1. This paper only aims to make a theoretical contribution and does not provide a recipe for a general algorithmic solution to disentangled representation learning.
2. our insights can elucidate answers to questions like
   a) what are the “data generative factors”,
   b) which factors should in principle be possible to disentangle (and what form their representations may take),
   c) should each generative factor correspond to a single or multiple latent dimensions, and
   d) should a disentangled representation of a particular dataset have a unique basis (up to a permutation of axes).

Our symmetry world

1. Many natural transformations will **change certain aspects** of the world state while keeping other **aspects unchanged**.
2. The study of symmetries in physics proved that every conservation law is grounded in a corresponding continuous symmetry transformation.

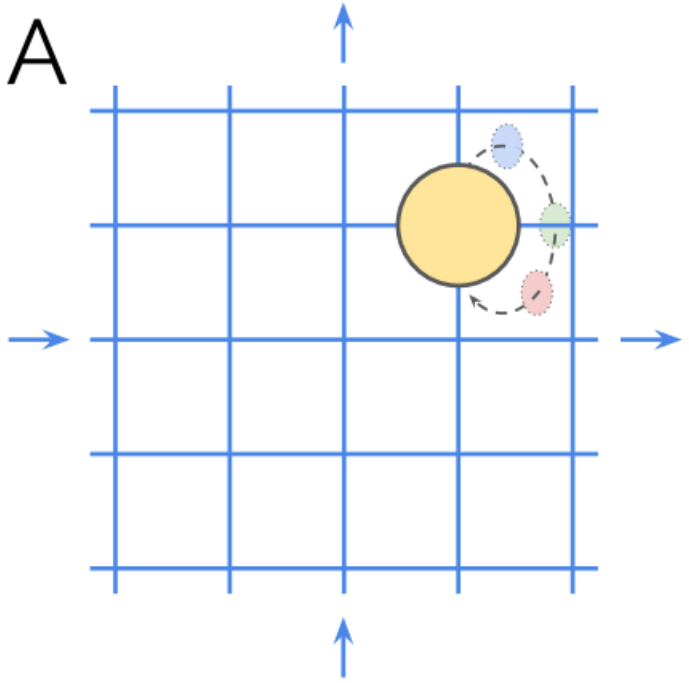


Our symmetry world

1. The more precise and exact these symmetries are, the more powerful the generalisation is to new domains.
2. In machine perception, the most powerful generalisation we can hope for is by understanding what properties of the world remain the same when transformed in certain ways.
3. In scene understanding, these transformations include translations, rotations and changes in object colour.
4. Symmetry group is a group of actions that do not change the identity of the object

A roadmap to defining disentangled representations

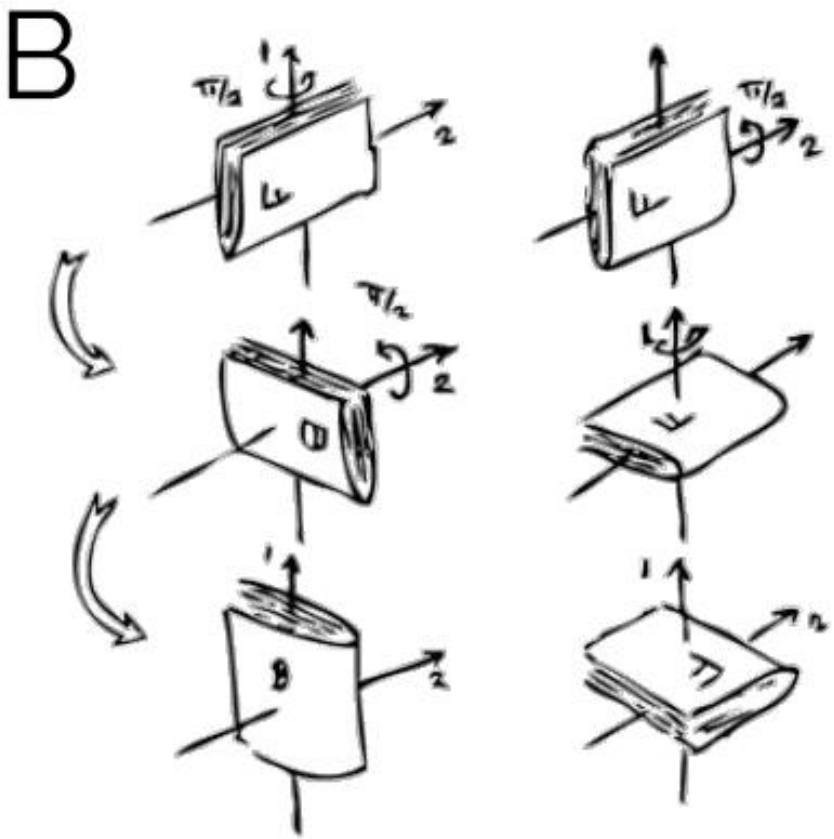
1. Disentangled group actions: change a certain aspect of the world state, while keeping others fixed.



$$G = G_x \times G_y \times G_c$$

A roadmap to defining disentangled representations

1. Intuitively, we might think that a representation of such a group could be disentangled into rotations about the x, y and z axes.
2. But is that the case?
3. It is not a disentangled representation



A roadmap to defining disentangled representations

A vector representation is called a ***disentangled representation*** *with respect to a particular decomposition of a symmetry group into subgroups, if it decomposes into independent subspaces, where each subspace is affected by the action of a single subgroup, and the actions of all other subgroups leave the subspace unaffected.*

A formal definition of disentangled representations

----Disentangled Group Actions

Suppose that we have a group action $\cdot : G \times X \to X$, and that the group G decomposes as a direct product $G = G_1 \times G_2$. We are going to refer to the action of the full group as $\cdot$, and the actions of each subgroup as $\cdot_i$. Then we propose the following definition: the action is *disentangled* (with respect to the decomposition of G) if there is a decomposition $X = X_1 \times X_2$, and actions $\cdot_i : G_i \times X_i \to X_i, i \in \{1,2\}$ such that

$$(g_1, g_2) \cdot (v_1, v_2) = (g_1 \cdot_1 v_1, g_2 \cdot_2 v_2) \tag{1}$$

In particular this says that an element of G_1 acts on X_1 but leaves X_2 fixed, and vice versa.

Growing long hair does not affect my eye color

A formal definition of disentangled representations

----Group Theory

A *group* $(G, \circ)$ is a set G together with a binary operation $\circ : G \times G \to G$ satisfying the following axioms:

1. Associativity $\forall x,y,z \in G : x \circ (y \circ z) = (x \circ y) \circ z$
2. Identity $\exists e \in G, \forall x \in G : e \circ x = x \circ e = x$
3. Inverse $\forall x \in G, \exists x^{-1} \in G : x \circ x^{-1} = x^{-1} \circ x = e$

Note that the binary operation is not required to be commutative. That is, we need not have $x \circ y = y \circ x, \forall x, y \in G$. A group that is commutative is called Abelian.

A formal definition of disentangled representations

----Disentangled Representations

1. A generative process $b : W \to O$
2. An inference process $h : O \to Z$
3. The composition $f : W \to Z,\ f = h \circ b.$
4. We want the action on *Z* to correspond to the action on *W*

$$g \cdot f(w) = f(g \cdot w) \quad \forall g \in G, w \in W$$

(Change the first value of the encodings can generate different hair length)

A formal definition of disentangled representations

----Disentangled Representations

Hence, f can be called a G-morphism or an equivariant map.

$$\begin{array}{ccc} G \times W & \xrightarrow{\cdot_W} & W \\ {\scriptstyle id_G \times f}\downarrow & & \downarrow{\scriptstyle f} \\ G \times Z & \xrightarrow{\cdot_Z} & Z \end{array}$$

A formal definition of disentangled representations

----A worked example

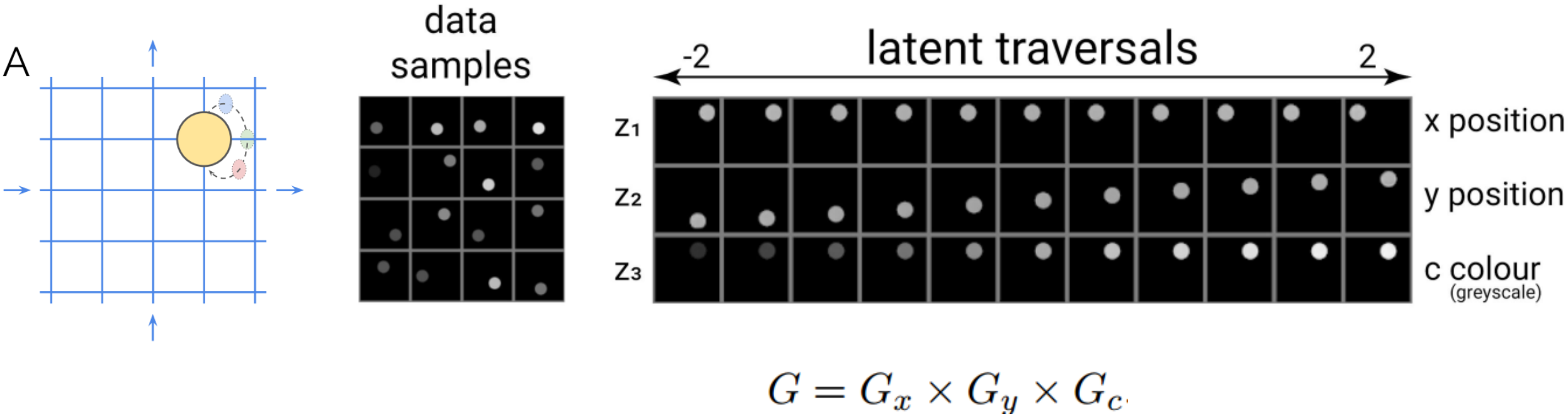
In other words, an agent's representation Z is disentangled with respect to the decomposition $G = G_1 \times ... \times G_n$ if

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.

A formal definition of disentangled representations

----A worked example

CCI-VAE [13]

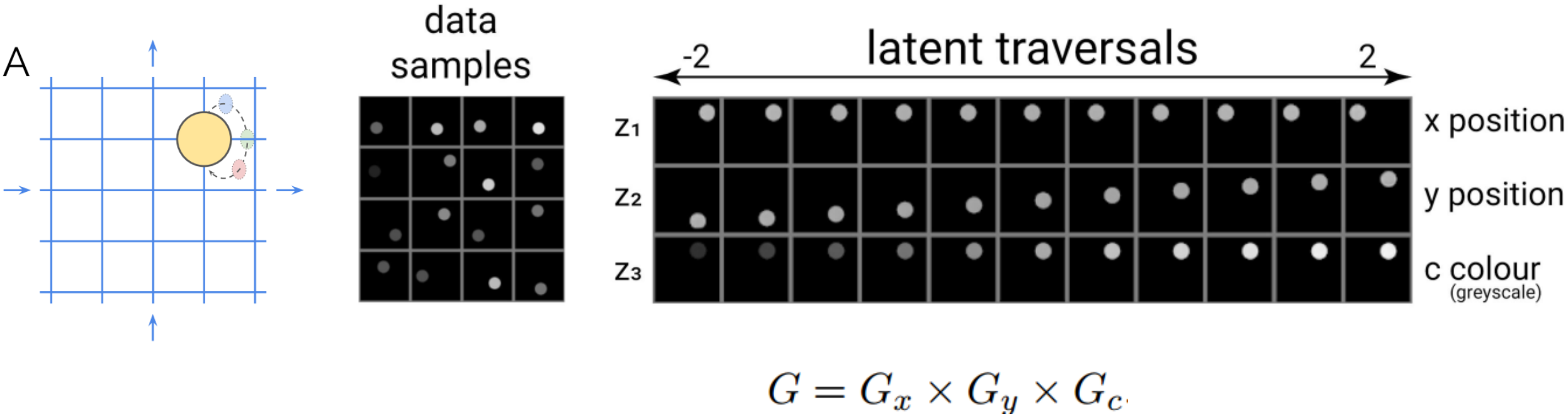


$$G = G_x \times G_y \times G_c$$

A formal definition of disentangled representations

----A worked example

CCI-VAE [13]



$$G = G_x \times G_y \times G_c$$

Backward compatibility of the new definition

1. **Modularity:** Modularity measures whether a single latent dimension encodes no more than a single data generative factor.
2. **Compactness:** Compactness measures whether each data generative factor is encoded by a single latent dimension.
3. **Explicitness:** Explicitness measures whether the values of *all* of the data generative factors can be decoded from the representation using a *linear* transformation

Conclusions

1. The structure of the world that disentangled representations should capture are the symmetry transformations of the world state.
2. Then used group and representation theory to show how the structure of the symmetry transformations can be reflected in the representation vector space.
3. Assumed that the symmetry groups can be decomposed as direct products of subgroups in a natural way and that their interesting decompositions into subgroups were known.