

Towards better understanding of disentangled and concept based representation learning

BricksND - Benchmark for disentangled concept learning

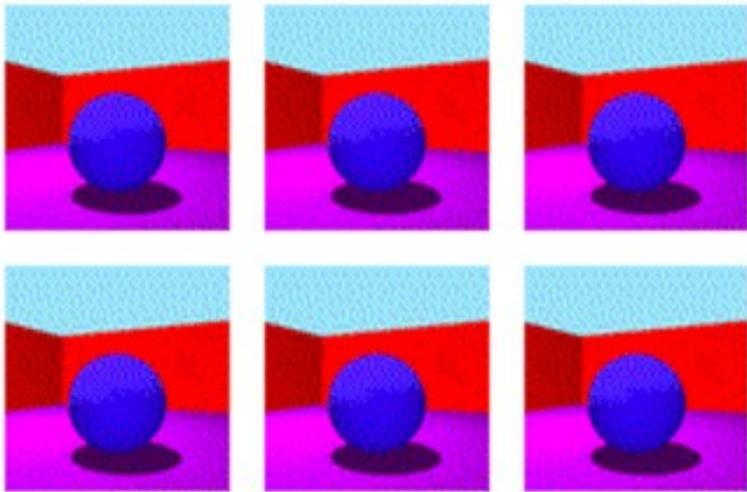
Ivan Grubišić, Tomislav Lipić, Benjamin Mesić, Tomislav Šmuc

Motivation

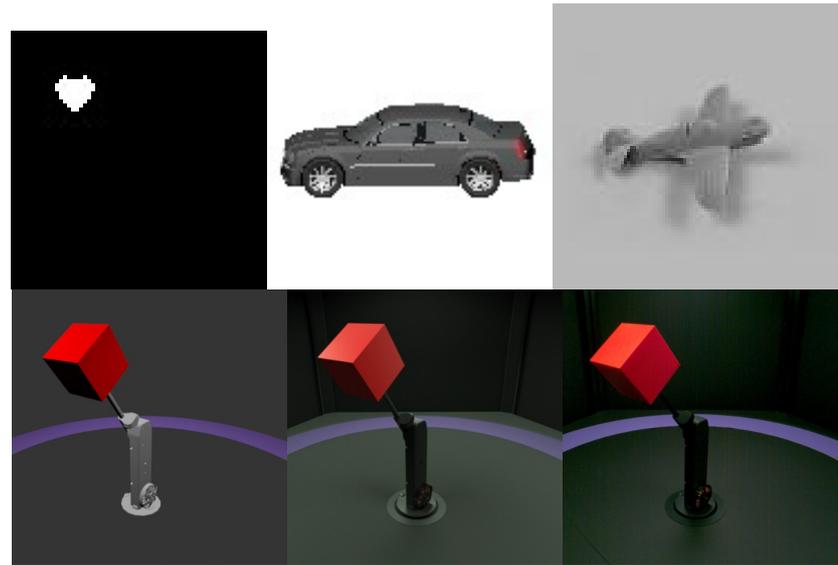
- Explainable machine learning (ML)
- ML robustness and repeatability
- Learn more with less (single-shot, few-shot)
- ML concept understanding
- Understanding interconnections between disentangled and concept based representations

Disentangled representation learning

- Existing data sets: disentanglement_lib, dSprites, Cars3D, NORB, mpi3d ...



disentanglement_lib



Concept learning

- Existing data sets: OMNIGLOT, CUB, CURI, Animals with attributes, ...

polar bear

black: no
white: yes
brown: no
stripes: no
water: yes
eats fish: yes



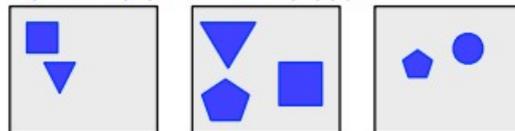
zebra

black: yes
white: yes
brown: no
stripes: yes
water: no
eats fish: no



for-all $x \in S$ ($\text{color?}(x) = \text{"blue"}$) and ($\text{all}(\text{size?}(S) = \text{size?}(x))$)

All objects are blue and have the same size



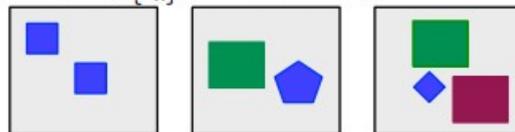
for-all $x \in S$ ($\text{all}(\text{color?}(x) = \text{color?}(S))$)

All objects in the scene have the same color



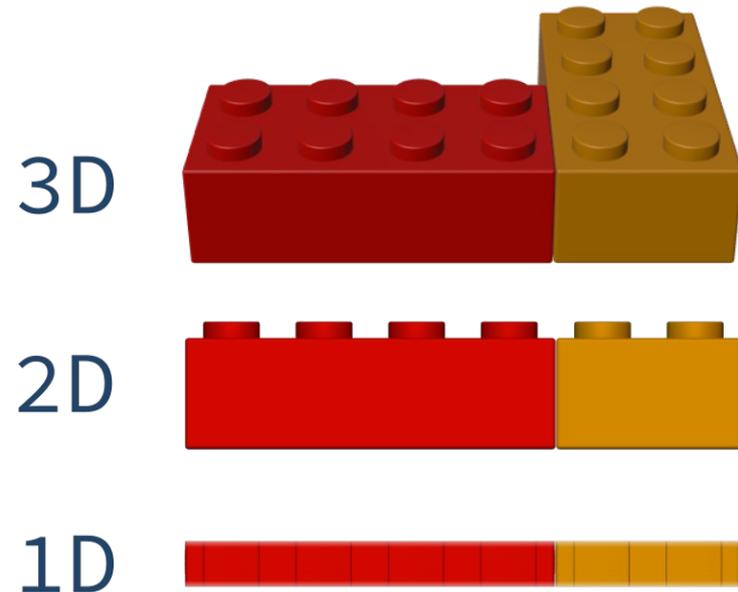
exists $x \in S$ ($\text{color?}(x) = \text{"blue"}$) and $\text{all}(\text{shape?}(S_{\{-x\}}) = \text{"square"})$

There exists a blue object in the scene and the rest of the objects are squares



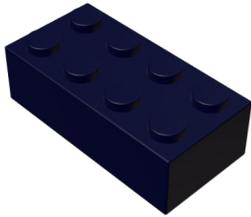
Disentangled concept data set – BricksND

- Data set in different spaces (1D, 2D, 3D)
- Variable properties: shape, colour (hue, saturation, value), position, scale, number
- Concept representation
- Hierarchical concept complexity (simple to complex)



Complexity representation

- Simple concept



	Disentangled	Concept
Shape	shape_01	Long wide thick (LWT)
Colour	H=168, S=130, V=70	Navy blue (NB)
Scale	1.5	Large (L)
Position	X=0, Y=0	Center (C)

- Complex concept



	Concept		
Object ID	1	2	3
Shape	LWT	LWT	NWT
Colours	NB	Y	R
Connected	2, 3	1, 3	1, 2
Beside	3		1
Above		1, 3	
Bellow	2		2

References

- [1] F. Locatello, et al. Challenging common assumptions in the unsupervised learning of disentangled representations. In international conference on machine learning, pages 4114–4124, 2019.
- [2] M. W. Gondal, et al. On the transfer of inductive bias from simulation to the real world: a new disentanglement dataset. arXiv preprint arXiv:1906.03292, 2019.
- [3] B. M. Lake, et al. The omniglot challenge: a 3-year progress report. *Current Opinion in Behavioral Sciences*, 29:97–104, 2019.
- [4] L. Matthey, et al. dsprites: Disentanglement testing sprites dataset. <https://github.com/deepmind/dspritesdataset/>, 2017.
- [5] R. Vedantam, et al. Curi: A benchmark for productive concept learning under uncertainty. ArXiv, abs/2010.02855, 2020.
- [6] F. X. Yu, et al. Designing category-level attributes for discriminative visual recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 771–778, 2013.