Università di Genova | MaLGa | NEURAL INFORMATION PROCESSING SYSTEMS

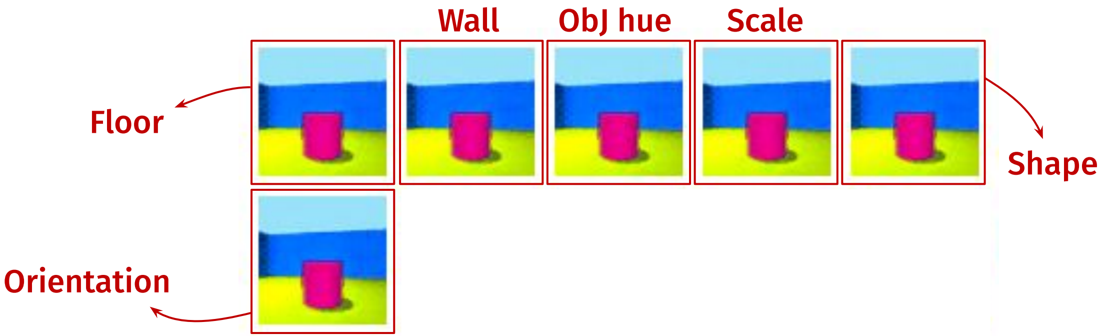
Transferring disentangled representations: bridging the gap between synthetic and real images

Jacopo Dapueto

Nicoletta Noceti

Francesca Odone

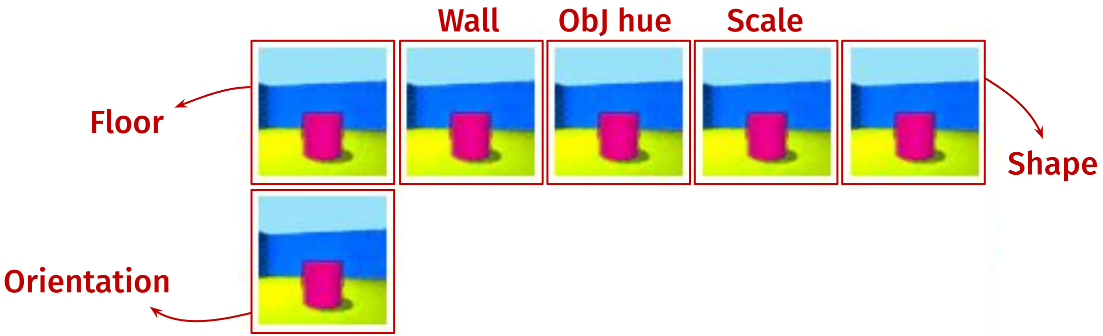
Introduction: intuition



Disentangled representation learning (DRL):

1) Identify the informative Factors of Variations (FoVs) in the data
2) Encode the FoVs in separated parts of the representation

Introduction: intuition



Disentangled representation learning (DRL):

1) Identify the informative Factors of Variations (FoVs) in the data
2) Encode the FoVs in separated parts of the representation

Properties:

- Modular: a FoV affects only a subset of representation
- Compact: a FoV uses as less dimensions as possible → ideally *one FoV, one dim.*
- Explicit: it is possible to retrieve all FoVs from representation

Motivations

Motivations:

1. Fully unsupervised DRL has been shown *unsatisfactory* **but** labelling every single factor of real-world data is **unfeasible** or **impossible**.
2. Most used metrics depend on **Classifiers** or **Mutual Information Estimation**

Motivations

Motivations:

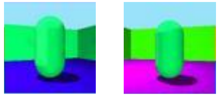
1. Fully unsupervised DRL has been shown *unsatisfactory* **but** labelling every single factor of real-world data is **unfeasible** or **impossible**.
2. Most used metrics depends on **Classifiers** or **Mutual Information Estimation**

Our contributions:

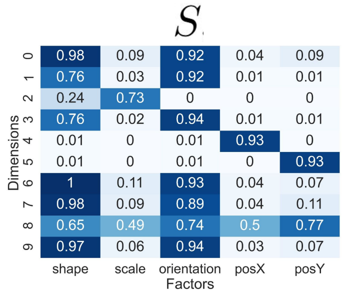
1) Methodology for DR transfer to Target datasets **without FoV annotation**
2) Novel classifier-free and interpretable **metric**
3) Extensive **experimental analysis** that considers different (Source, Target)

OMES: Overlap Multiple Encoding Scores

Intervention-based: couple samples differing in 1 FoV



Association matrix S based on **Correlation**



Provide **interpretable** info about the structure of the representation

$$OMES(S) = \frac{1}{n}\sum_{j=1}^{n} \alpha \, \mathrm{OS}(S, j) + (1-\alpha)\, \mathrm{MES}(S, j)$$

OMES: Overlap Multiple Encoding Scores

Intervention-based: couple samples differing in 1 FoV  

Association matrix S based on **Correlation**

Provide **interpretable** info about the structure of the representation

S

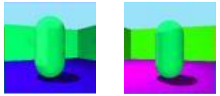
Dimensions	shape	scale	orientation	posX	posY
0	0.98	0.09	0.92	0.04	0.09
1	0.76	0.03	0.92	0.01	0.01
2	0.24	0.73	0	0	0
3	0.76	0.02	0.94	0.01	0.01
4	0.01	0	0.01	0.93	0
5	0.01	0	0.01	0	0.93
6	1	0.11	0.93	0.04	0.07
7	0.98	0.09	0.89	0.04	0.11
8	0.65	0.49	0.74	0.5	0.77
9	0.97	0.06	0.94	0.03	0.07

Factors

$$OMES(S) = \frac{1}{n}\sum_{j=1}^{n} \alpha\, \mathrm{OS}(S, j) + (1-\alpha)\, \mathrm{MES}(S, j)$$

Modularity

OMES: Overlap Multiple Encoding Scores

Intervention-based: couple samples differing in 1 FoV 

Association matrix S based on **Correlation**

Provide **interpretable** info about the structure of the representation

S

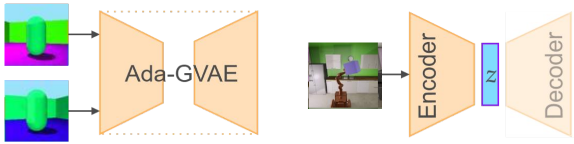
Dimensions	shape	scale	orientation	posX	posY
0	0.98	0.09	0.92	0.04	0.09
1	0.76	0.03	0.92	0.01	0.01
2	0.24	0.73	0	0	0
3	0.76	0.02	0.94	0.01	0.01
4	0.01	0	0.01	0.93	0
5	0.01	0	0.01	0	0.93
6	1	0.11	0.93	0.04	0.07
7	0.98	0.09	0.89	0.04	0.11
8	0.65	0.49	0.74	0.5	0.77
9	0.97	0.06	0.94	0.03	0.07

Factors

$$OMES(S) = \frac{1}{n}\sum_{j=1}^{n} \alpha\, \mathrm{OS}(S,j) + (1-\alpha)\, \mathrm{MES}(S,j)$$

Modularity *Compactness*

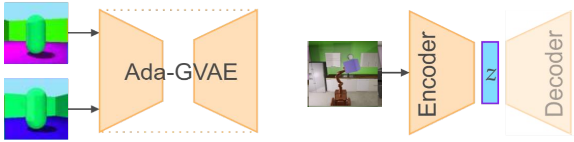
Transfer experiments



We considered **Source** and **Target** datasets covering different challenges

Dataset	Real	3D	Occlusions	#FoV	Independence	Complete annotation	Resolution	#Images
dSprites	✗	✗	✗	5	✓	✓	64×64	737K
Noisy-dSprites	✗	✗	✗	5	✓	✓	64×64	737K
Color-dSprites	✗	✗	✗	6	✓	✓	64×64	4,4M
Noisy-Color-dSprites	✗	✗	✗	6	✓	✓	64×64	4,4M
Shapes3D	✗	✓	✓	6	✓	✓	64×64	480K
Isaac3D	✗	✓	✓	9	✓	✓	128×128	737K
Coil100-Augmented	✓	✓	✓	4	✓	✓	128×128	1,1M
RGB-D Objects	✓	✓	✓	3*	✗	✗	256×256	35K

Transfer experiments



We considered **Source** and **Target** datasets covering different challenges and different scenarios:

- syn2syn, syn2real & real2real
- Additional FoV; FoV of similar semantics; etc.

Dataset	Real	3D	Occlusions	#FoV	Independence	Complete annotation	Resolution	#Images
dSprites	✗	✗	✗	5	✓	✓	64×64	737K
Noisy-dSprites	✗	✗	✗	5	✓	✓	64×64	737K
Color-dSprites	✗	✗	✗	6	✓	✓	64×64	4,4M
Noisy-Color-dSprites	✗	✗	✗	6	✓	✓	64×64	4,4M
Shapes3D	✗	✓	✓	6	✓	✓	64×64	480K
Isaac3D	✗	✓	✓	9	✓	✓	128×128	737K
Coil100-Augmented	✓	✓	✓	4	✓	✓	128×128	1,1M
RGB-D Objects	✓	✓	✓	3*	✗	✗	256×256	35K

Discussion

Our study suggests that

1) Our metric provides **helpful information** to evaluate the transferred representation and it's **more robust** w.r.t. existing metrics.

2) We can design a synthetic dataset to disentangled specific FoVs then **transfer while preserving all the disentanglement properties**.

Github

UniGe

MaLGa