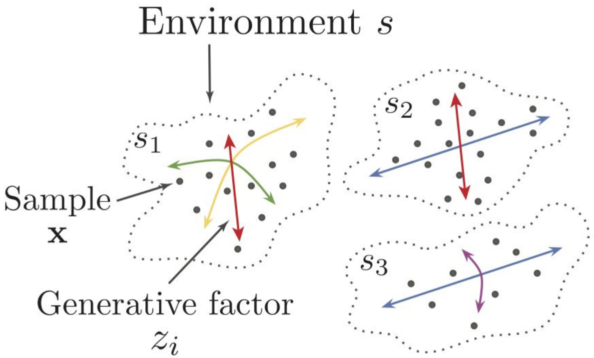
Life-Long Disentangled Representation Learning with Cross-Domain Latent Homologies

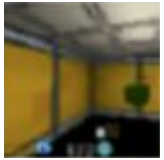
Alessandro Achille, Tom Eccles, Loic Matthey, Christopher P. Burgess,
Nick Watters, Alexander Lerchner, Irina Higgins

DeepMind

UCLA

Life-long learning of disentangled representations





Automatically **detect shifts** in the data distribution

Allocate spare **representational capacity** to learn about the new data

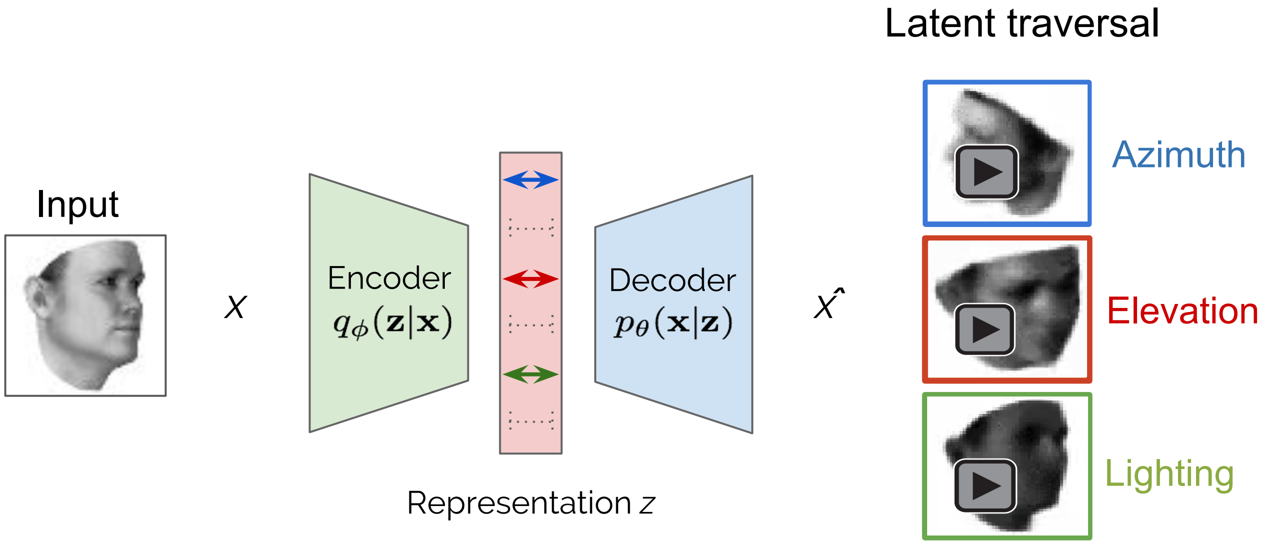
Prevent catastrophic forgetting of previously learnt representations

Share latent dimensions between datasets where appropriate

Disentangled representations with CCI-VAE

Independent factors can be recovered by slowly increasing the representation capacity:

$$\mathcal{L}_{\text{MDL}}(\phi,\theta) = \underbrace{\mathbb{E}_{\mathbf{z}^s\sim q_\phi(\cdot|\mathbf{x}^s)}[-\log p_\theta(\mathbf{x}\mid \mathbf{z}^s, s)]}_{\text{Reconstruction error}} + \gamma|\underbrace{\mathbb{KL}(q_\phi(\mathbf{z}^s|\mathbf{x}^s)||p(\mathbf{z}))}_{\text{Representation capacity}} - \underbrace{C}_{\text{Target}}|^2$$

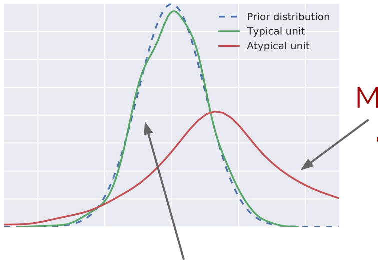


Atypical and shared factors

Which factors can be reused when the environment changes?

Atypicality

$$\alpha_i = \mathbb{KL}(\underbrace{\mathbb{E}_x[q_\phi(z_i|x)]}_{\text{Marginal}} \,||\, \underbrace{p(z_i)}_{\text{Prior}})$$



If a factor is atypical in one environment, it should be **disabled** to prevent retraining.

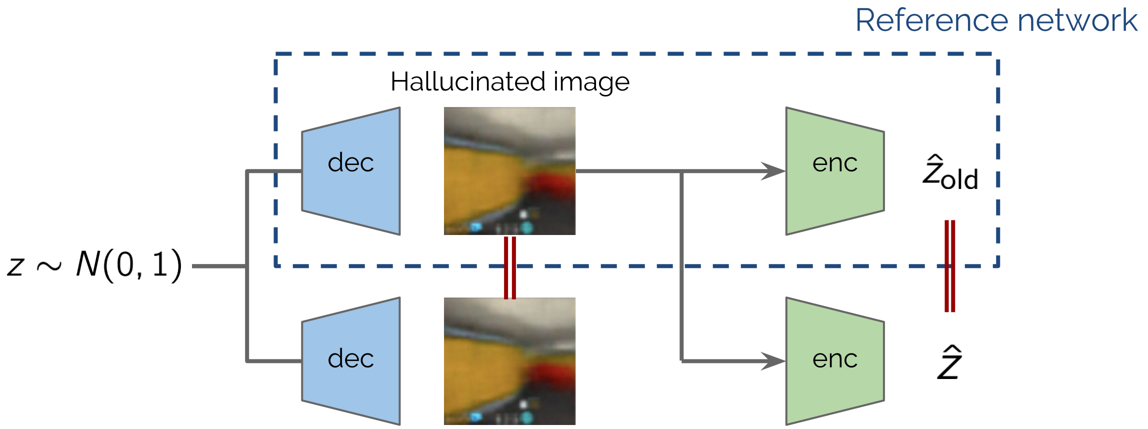
Typical and atypical factors can be used to **detect changes of environment** and **re-identify past environments**.

DeepMind

Imagination Feed-Back Loop

Need to **prevent forgetting** of atypical (disabled) factors while training on new environments.

Idea: Train on hallucinated data from old environments and force equality to past network

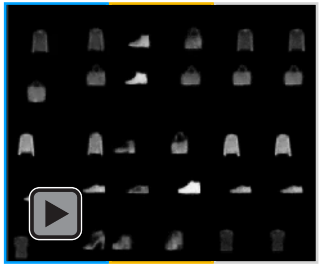


$$\mathcal{L}_{\text{past}}(\phi,\theta) = \mathbb{E}_{z,s',x'}\Big[\underbrace{D[q_\phi(z|x'),\ q_{\phi'}(z'|x')]}_{\text{Encoder proximity}} + \underbrace{D[q_\theta(x|z,s'),\ q_{\theta'}(x'|z,s')]}_{\text{Decoder proximity}}\Big]$$

Sharing latent factors without forgetting

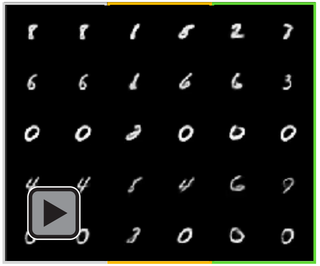
Sharing and reusing semantic factors in multiple environment

Moving Fashion Items



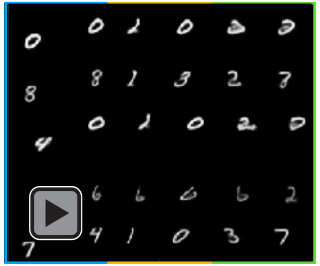
Position Shape ~~Digit~~

Fixed Digits



~~Position~~ Shape Digit

Moving Digits

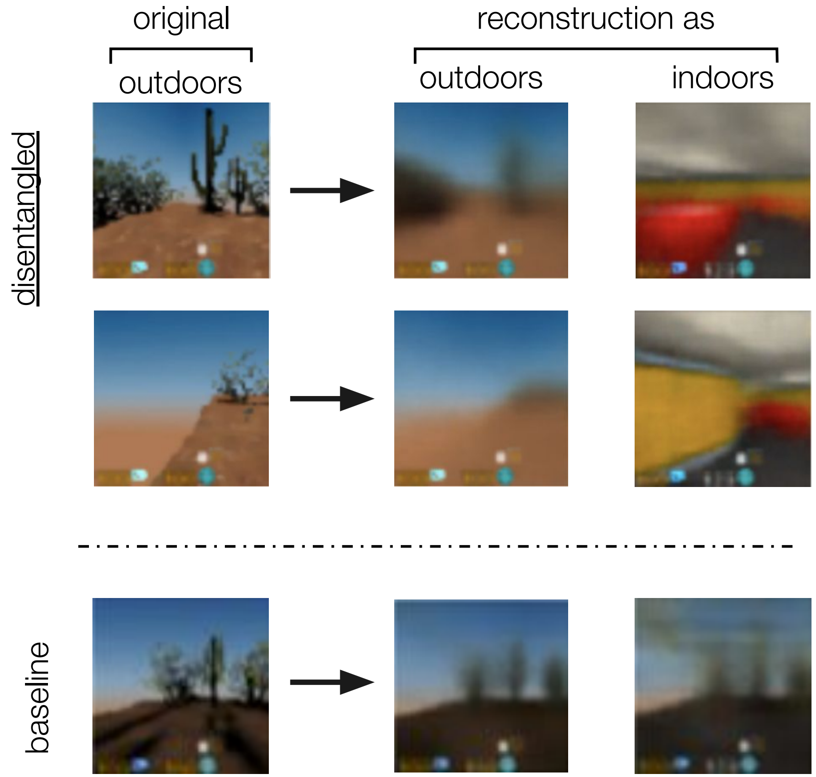


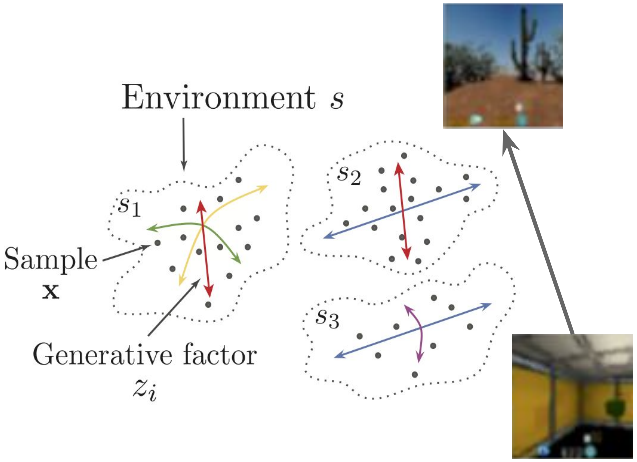
Position Shape Digit

	Disentangled				Entangled			
	Object ID Accuracy		Position MSE		Object ID Accuracy		Position MSE	
Ablation	Max (%)	Change (%)	Min (*1E-4)	Change (*1E-4)	Max (%)	Change (%)	Min (*1E-4)	Change (*1E-4)
-	88.6 (±0.4)	-15.2 (±2.8)	3.5 (±0.05)	24.8 (±13.5)	91.8 (±0.4)	-12.1 (±0.8)	4.2 (±0.7)	10.5 (±2.6)
S	88.9 (±0.5)	-13.9 (±1.9)	3.4 (±0.05)	22.5 (±12.2)	91.7 (±0.4)	-12.2 (±0.03)	4.5 (±0.8)	10.9 (±3.1)
D	88.6 (±0.3)	-14.4 (±1.9)	3.3 (±0.04)	21.4 (±4.9)	91.8 (±0.4)	-12.4 (±0.7)	4.3 (±0.7)	11.7 (±3.2)
A	86.7 (±1.9)	-24.5 (±1.0)	3.3 (±0.04)	67.6 (±107.0)	88.6 (±0.3)	-19.7 (±0.5)	4.5 (±0.7)	47.1 (±26.2)
SA	87.1 (±1.8)	-28.1 (±0.08)	3.3 (±0.04)	78.9 (±109.0)	89.9 (±1.3)	-18.3 (±0.4)	4.8 (±0.7)	41.8 (±20.6)
DA	86.3 (±2.5)	-25.2 (±0.5)	3.3 (±0.04)	72.2 (±90.0)	88.8 (±0.3)	-19.4 (±0.4)	4.6 (±0.7)	40.2 (±19.2)
SD	88.3 (±0.3)	-12.9 (±1.9)	3.4 (±0.05)	20.0 (±3.5)	91.4 (±0.3)	-11.7 (±0.6)	4.3 (±0.5)	11.6 (±1.9)
SD-[41]	-	-	-	-	91.9 (±0.1)	-11.6 (±1.1)	4.7 (±0.8)	10.2 (±1.8)
VASE (SDA)	88.6 (±0.4)	**-5.4 (±0.3)**	3.2 (±0.03)	**3.0 (±0.2)**	91.5 (±0.1)	-6.5 (±0.7)	4.2 (±0.4)	3.9 (±1.1)

DeepMind

Meaningful cross-domain translation

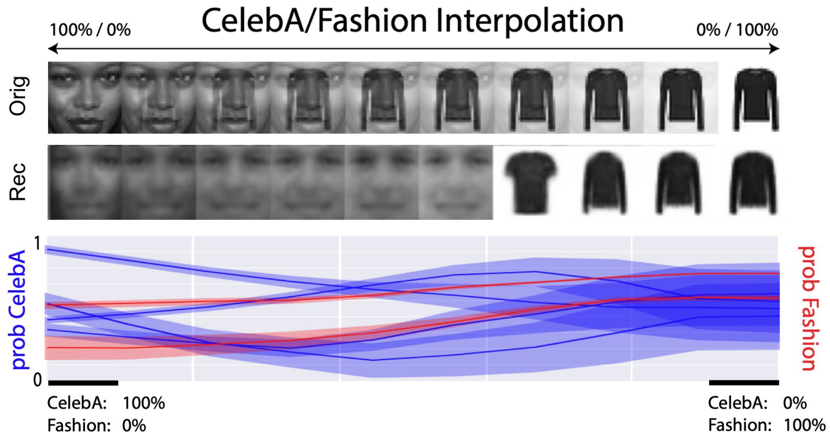




DeepMind

Dealing with ambiguity

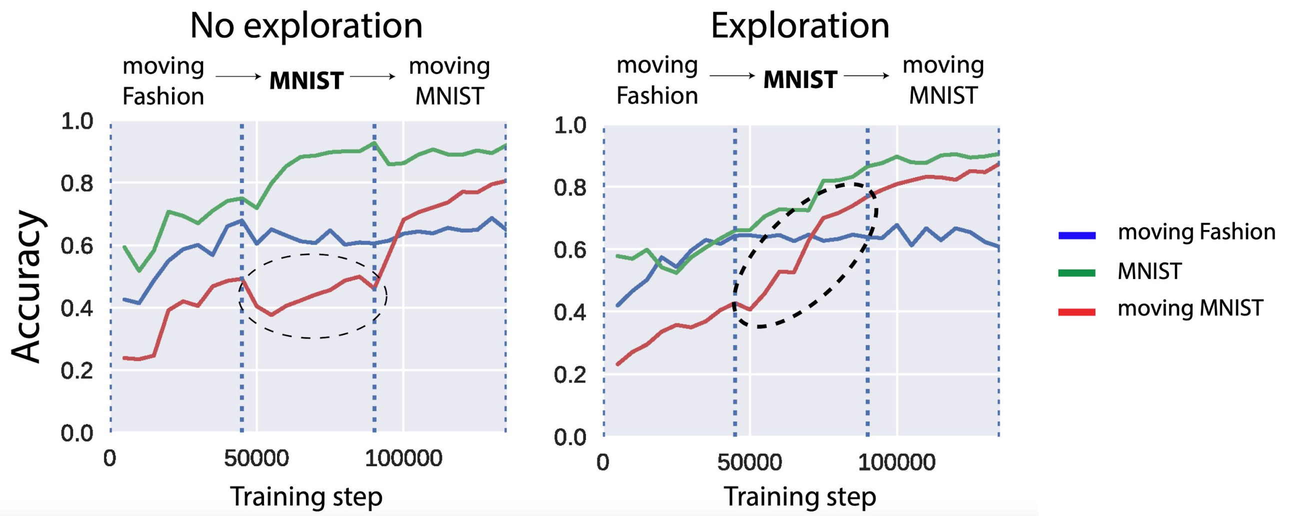
Presented with ambiguous stimuli our model express **uncertainty** through feature **variance**, but can reconstruct the without ambiguity.



Emergence of "**categorical perception**".

Imagination-driven exploration

An agent can act on the environment to realize a state it imagines possible given its past experience.



This imagination-driven exploration can improve the zero-shot performance on new environments.

Conclusions

Learn **disentangled factors** in a **life-long learning** setting.

Atypicality allows to detect environment changes and to share factors.

Imagination Feedback Loop to avoid catastrophic forgetting.

Compositional representation is robust and can be adapted to solve tasks in unseen environment.

Can **share** factors between environment in a **semantically meaningful** way.