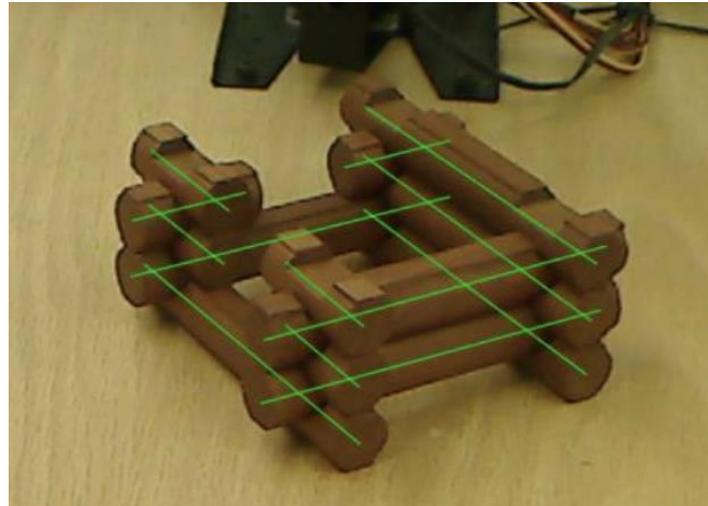
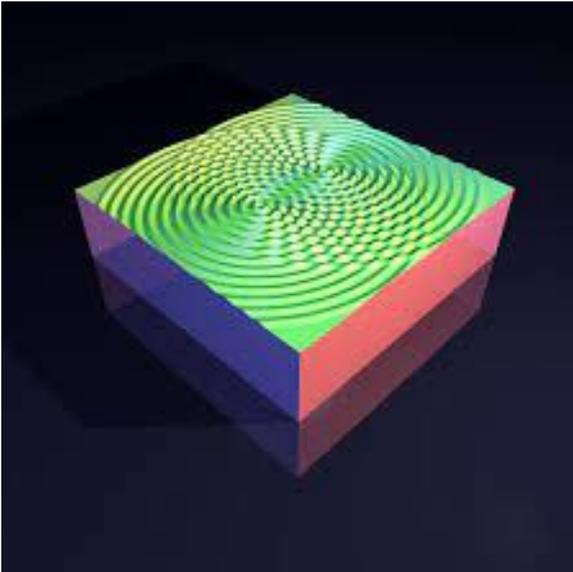
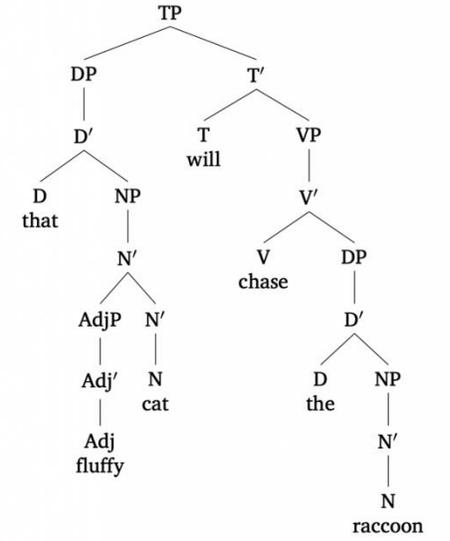
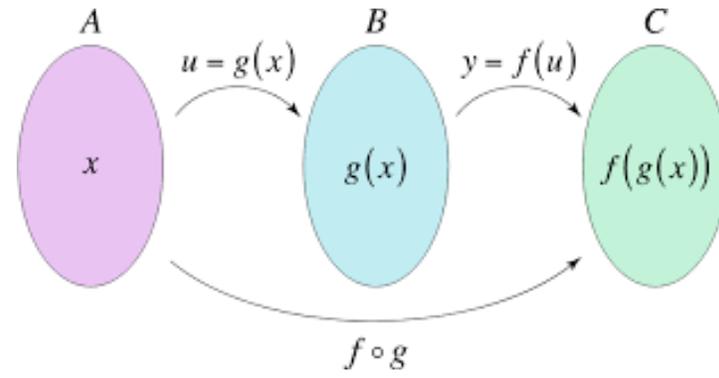


# Soft Tensor Product Representations for Fully Continuous, Compositional Visual Representations

Bethia Sun, Maurice Pagnucco, Yang Song

School of Computer Science and Engineering, UNSW Sydney



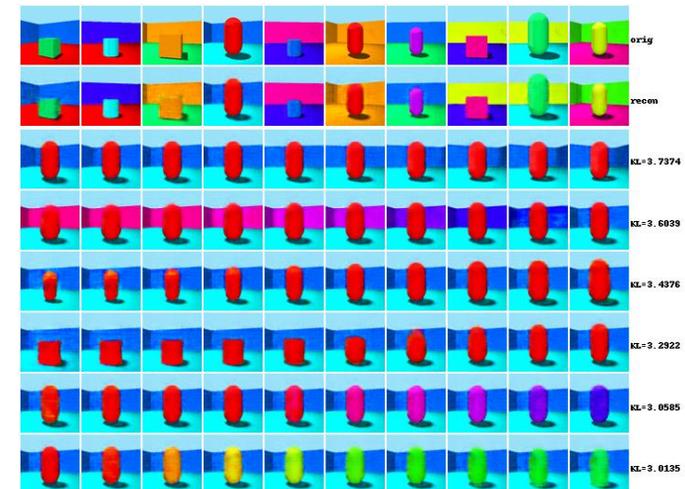


# Motivation

- **Theoretical:** long philosophical tradition (Fodor, Chomsky) of inductively arguing from key properties of human cognition that **cognition** itself *must* be underpinned by a **compositional system** [1, 2].

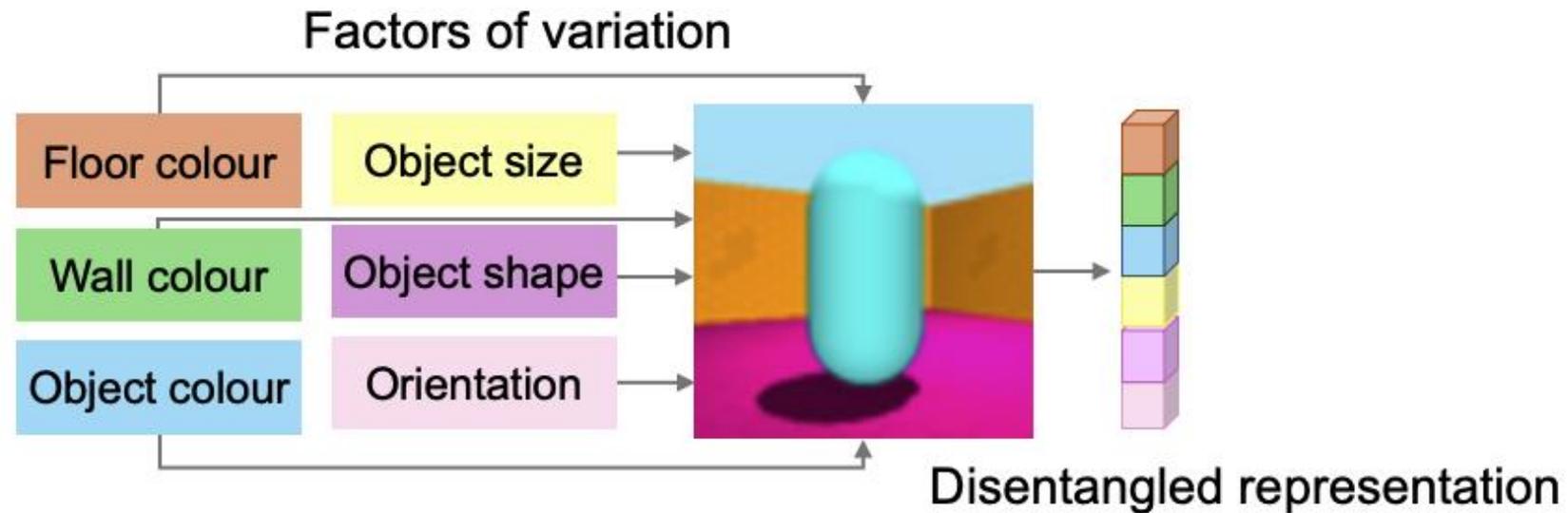


- **Empirical:** compositional representations enhance interpretability [4, 5], sample efficiency [6, 7], fairness [8, 9, 10], robustness to OOD settings [7, 11, 12].



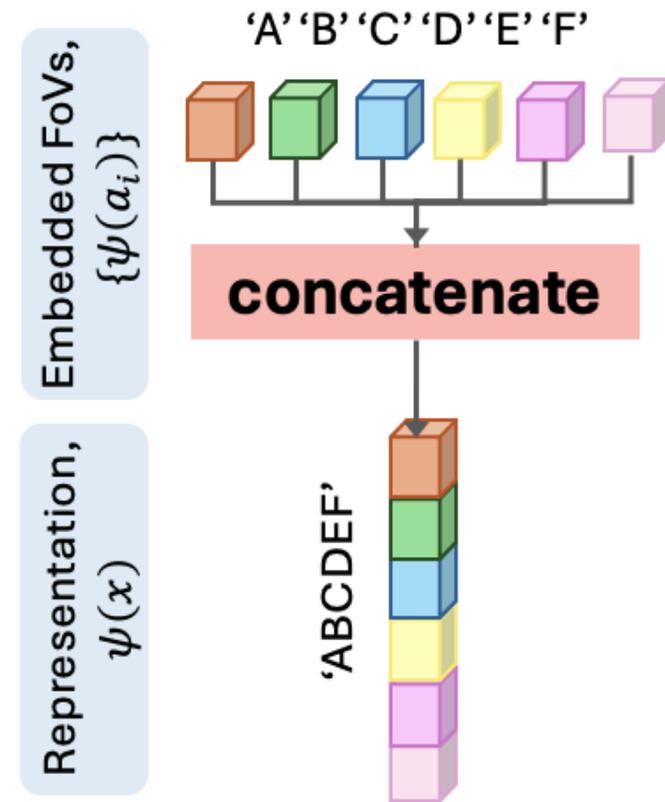
# Existing Work

- **Disentanglement** is a key approach for compositional representation learning.
- Aims to **isolate** underlying factors of variation (FoVs) into **distinct parts** of the representation.
  - i.e., FoVs should be 1-1 mapped to **representational parts** – the Jacobian requirement of [13].



# Disentanglement and Symbolic Compositionality

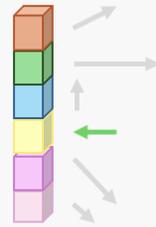
- Disentanglement enforces a fundamentally **symbolic** treatment of compositional structure.
  - This is because disentanglement essentially allocates FoVs to **distinct representational slots**.
  - The overall representation is thus analogous to a **string** formed by the **concatenation** of FoV slots (**tokens**).



# Our Key Hypothesis

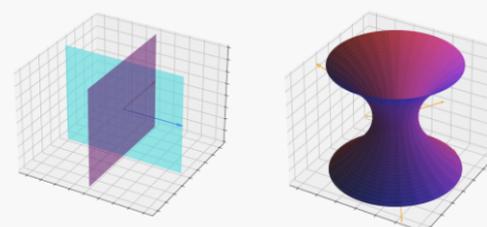
- Symbolic compositional representations may be **fundamentally incompatible** with the **continuous vector spaces** of deep learning:

## Gradient Flow & Learning



- This localist encoding offered by symbolic compositional representations may restrict smooth flow of gradient across all *dimensions* of the representation - e.g., when updating a *single* FoV (shown in yellow).
- Updates to a single FoV therefore provide minimal feedback to other FoVs, preventing joint and interdependent FoV updates.
- Furthermore, the localist encoding produces *abrupt* and *discontinuous* changes in the representation space when transitioning between FoV updates, potentially complicating convergence.

## Expressivity & Robustness to Noise

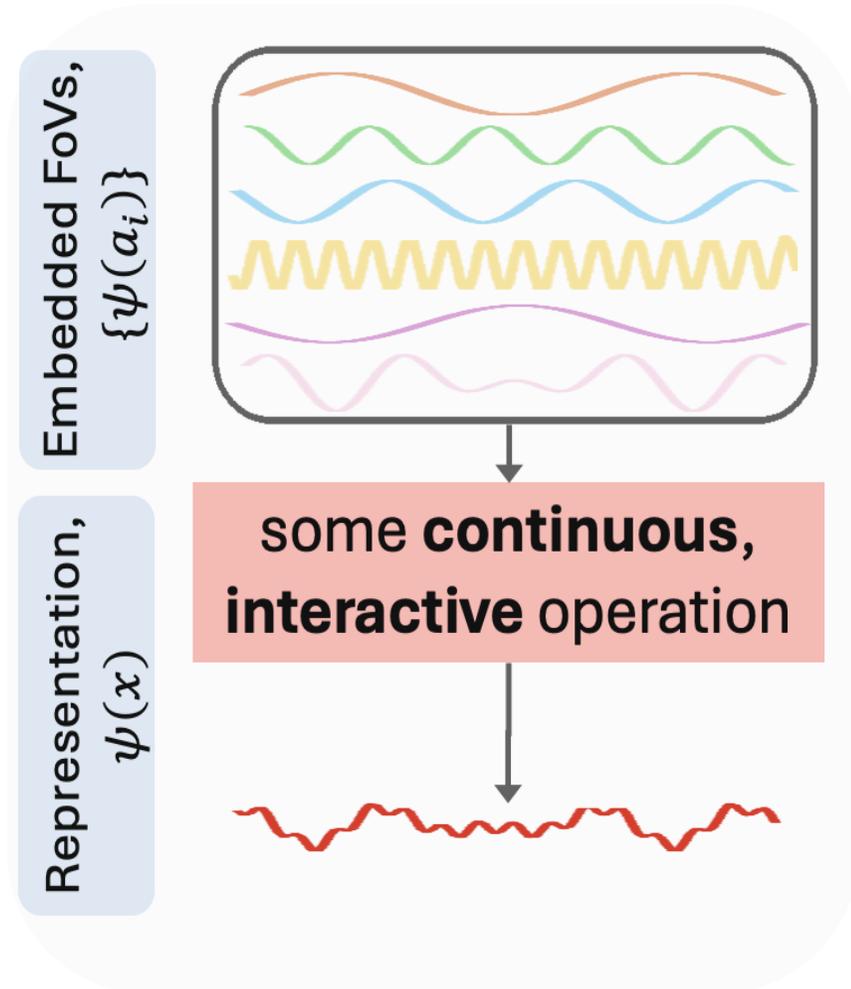


- Consider a symbolic compositional representation of  $n$  FoVs, in  $d$ -dimensional space.
- Such a representation has constrained practical expressivity, as each FoV only has  $d/n$  dimensions to encode information (contrast the case on the LHS where each FoV can only use a *single* dimension vs the RHS where FoVs can use *two* dimensions)
- Furthermore, such an encoding scheme is highly vulnerable to dimension-wise noise, as such noise will directly impair the representation of the corresponding FoV (i.e., there is no overlapping or redundant encoding to distribute the effect of such noise)

- The symbolic/continuous mismatch may manifest in broadly **suboptimal deep learning model behaviour.**

# A New Way of Treating Compositional Structure?

- Can we align compositional structure with continuous vector spaces, by formulating a fundamentally **distributed compositional representation**?
  - Such an approach **smoothly blends** FoVs into the representation – like the **continuous superimposition** of **multiple waves** into an aggregate wave (in red on the RHS)



# Soft TPR Framework

- To do this, we propose a **new compositional representation learning framework**, the *Soft Tensor Product Representation (TPR) framework*, which comprises:
  1. *Soft TPR*: a new, inherently **distributed compositional representational form**.
  2. *Soft TPR Autoencoder*: a theoretically-principled **method** for **learning Soft TPRs**.

# Soft TPR

- Our Soft TPR form is a new mathematical specification that represents compositional structure in a **distributed fashion**:

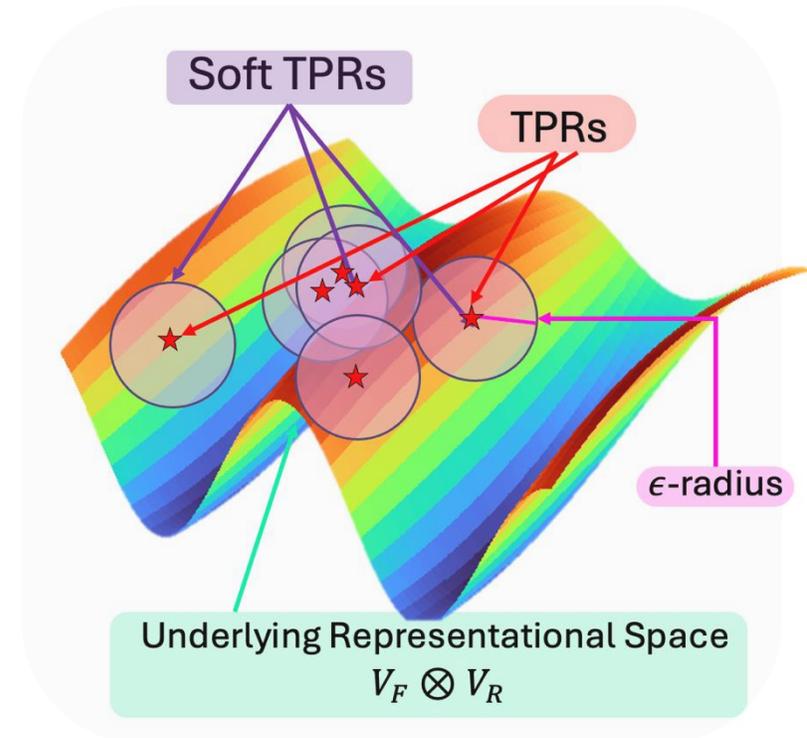
Soft TPR Form:

$$z \in \{x \in V_F \otimes V_R \mid \|x - \psi_{tpr}\|_F \leq \epsilon\}$$

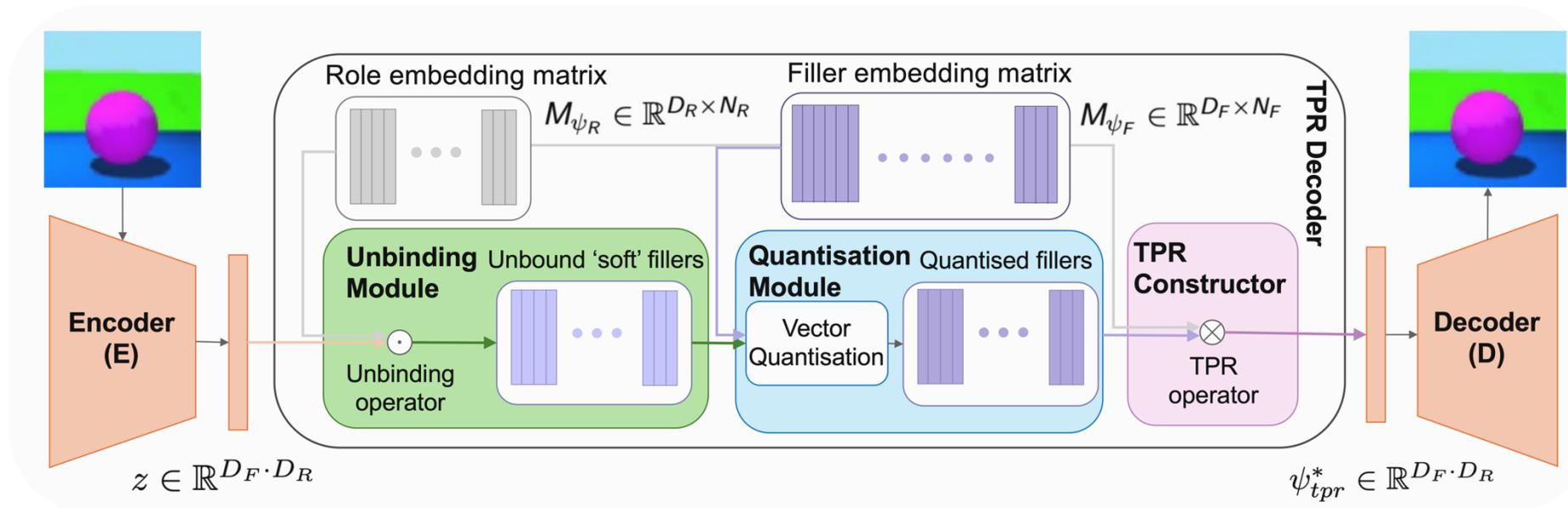
where  $\|A\|_F$  denotes Frobenius norm of  $A$ ,  
 $\epsilon$  some small, +ve scalar-valued constant,

$\psi_{tpr}$  a (traditional) TPR produced by TPR function mapping from data to TPRs in  $V_F \otimes V_R$

- It extends upon Smolensky's Tensor Product Representation [3].
- Soft TPR preserves the traditional TPR's useful mathematical & structural properties (see paper for proofs & further details).
- Soft TPRs have the added benefits of being **easier to learn** and more **representationally flexible** than TPRs.
  - This allows Soft TPRs to be **applied in broader settings** compared to traditional TPRs [14, 15, 16, 17, 18, 19, 20] e.g., the non-formal domain of *vision* with a more realistic *weak supervision* requirement.



# Soft TPR Autoencoder



- A novel framework introduced to learn Soft TPRs. 3 main components (please see paper for more details):
- **Encoder**: Produces a **candidate Soft TPR**,  $z$ .
- **TPR Decoder**: Leverages the mathematical properties of the Soft TPR/TPR framework to encourage  $z$  to have the **correct mathematical form** of a **Soft TPR** (unsupervised loss).
- **Weak Supervision**: Apply a weakly supervised loss inspired by prior disentanglement work [21, 22, 23, 24, 25] to encourage  $z$  to contain the **correct semantic content**.

# Results

- Our results empirically suggest that the **enhanced** vector space alignment produced by Soft TPRs is **broadly** beneficial for deep learning models (both representation learners & downstream models).
  - Please see the Appendix in our paper for an extensive suite of experimental results.

# Result #1: Structural

- Structurally, Soft TPRs are **more explicitly compositional** than baselines (as quantified by disentanglement metrics).
  - SoTA disentanglement (DCI boosts of **29%+**, **74%+** on Cars3D/MPI3D).

Table 1: FactorVAE and DCI scores. Additional results in Section C.3.3

Models	Cars3D		Shapes3D		MPI3D	
	FactorVAE score	DCI score	FactorVAE score	DCI score	FactorVAE score	DCI score
Symbolic scalar-tokened compositional representations						
Slow-VAE	$0.902 \pm 0.035$	$0.509 \pm 0.027$	$0.950 \pm 0.032$	$0.850 \pm 0.047$	$0.455 \pm 0.083$	$0.355 \pm 0.027$
Ada-GVAE-k	$0.947 \pm 0.064$	$0.664 \pm 0.167$	$0.973 \pm 0.006$	<b><math>0.963 \pm 0.077</math></b>	$0.496 \pm 0.095$	$0.343 \pm 0.040$
GVAE	$0.877 \pm 0.081$	$0.262 \pm 0.095$	$0.921 \pm 0.075$	$0.842 \pm 0.040$	$0.378 \pm 0.024$	$0.245 \pm 0.074$
ML-VAE	$0.870 \pm 0.052$	$0.216 \pm 0.063$	$0.835 \pm 0.111$	$0.739 \pm 0.115$	$0.390 \pm 0.026$	$0.251 \pm 0.029$
Shu	$0.573 \pm 0.062$	$0.032 \pm 0.014$	$0.265 \pm 0.043$	$0.017 \pm 0.006$	$0.287 \pm 0.034$	$0.033 \pm 0.008$
Symbolic vector-tokened compositional representations						
VCT	$0.966 \pm 0.029$	$0.382 \pm 0.080$	$0.957 \pm 0.043$	$0.884 \pm 0.013$	$0.689 \pm 0.035$	$0.475 \pm 0.005$
COMET	$0.339 \pm 0.008$	$0.024 \pm 0.026$	$0.168 \pm 0.005$	$0.002 \pm 0.000$	$0.145 \pm 0.024$	$0.005 \pm 0.001$
Fully continuous compositional representations						
Ours	<b><math>0.999 \pm 0.001</math></b>	<b><math>0.863 \pm 0.027</math></b>	<b><math>0.984 \pm 0.012</math></b>	$0.926 \pm 0.028$	<b><math>0.949 \pm 0.032</math></b>	<b><math>0.828 \pm 0.015</math></b>

# Result #2: Representation Learner Convergence

- Soft TPRs have **faster representation learner convergence**.
- Representations **useful** for **downstream tasks** can be consistently learned with substantially **fewer** representation learner **training iterations**.
  - We consider the 2 standard downstream tasks used in disentanglement: FoV regression and abstract visual reasoning.
  - Note that at 100 iterations of representation learner training, Soft TPRs (in blue) achieve performance (Fig 20 & Fig 22) that is only achieved with **2 orders' magnitude more training iterations** by the most competitive baseline.

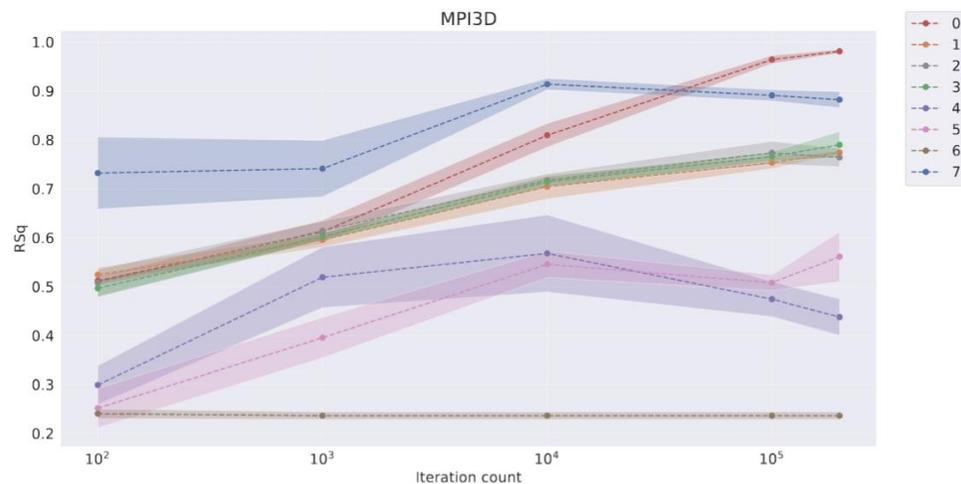


Figure 20: Convergence of representation learners as measured by FoV regression on the MPI3D dataset (dimensionality-controlled setting)

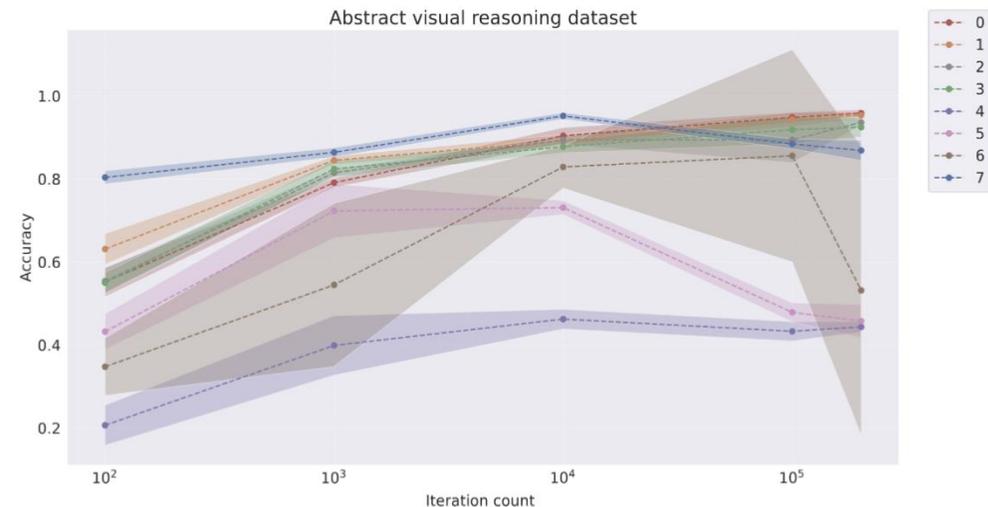


Figure 22: Convergence of representation learners as measured by classification performance on the abstract visual reasoning dataset (dimensionality-controlled setting)

# Result #3: Downstream Performance

- Soft TPRs have **substantially superior downstream sample efficiency** (e.g., **93%+**) and **low-sample regime performance** (e.g., **138%+**, **168%+**).
  - Again, we consider the 2 standard downstream tasks of FoV regression and abstract visual reasoning, a subset of results below:

Table 4: Downstream FoV  $R^2$  scores (odd columns) and sample efficiencies (even columns) on the MPI3D dataset.

Models	100 samples	100 samples/all	250 samples	250 samples/all
	Symbolic scalar-tokened compositional representations			
Slow-VAE	$0.127 \pm 0.050$	$0.130 \pm 0.051$	$0.152 \pm 0.011$	$0.155 \pm 0.011$
Ada-GVAE-k	$0.206 \pm 0.031$	$0.270 \pm 0.037$	$0.213 \pm 0.023$	$0.279 \pm 0.026$
GVAE	$0.181 \pm 0.030$	$0.234 \pm 0.035$	$0.217 \pm 0.023$	$0.282 \pm 0.027$
ML-VAE	$0.182 \pm 0.013$	$0.236 \pm 0.019$	$0.222 \pm 0.024$	$0.288 \pm 0.030$
Shu	$0.151 \pm 0.016$	$0.343 \pm 0.024$	$0.211 \pm 0.026$	$0.482 \pm 0.075$
	Symbolic vector-tokened compositional representations			
VCT	$0.086 \pm 0.051$	$0.189 \pm 0.107$	$0.119 \pm 0.070$	$0.246 \pm 0.137$
COMET	$-0.051 \pm 0.015$	$0.000 \pm 0.000$	$-0.042 \pm 0.018$	$0.000 \pm 0.000$
	Fully continuous compositional representations			
Ours	<b><math>0.490 \pm 0.068</math></b>	<b><math>0.556 \pm 0.078</math></b>	<b><math>0.594 \pm 0.056</math></b>	<b><math>0.665 \pm 0.067</math></b>

Table 5: Abstract visual reasoning accuracy in the low-sample regime of 500 samples.

Models	Symbolic scalar-tokened
Slow-VAE	$0.196 \pm 0.028$
Ada-GVAE-k	$0.203 \pm 0.007$
GVAE	$0.182 \pm 0.013$
ML-VAE	$0.193 \pm 0.012$
Shu	$0.200 \pm 0.010$
	Symbolic vector-tokened
VCT	$0.277 \pm 0.039$
COMET	$0.259 \pm 0.016$
	Fully continuous
Ours	<b><math>0.360 \pm 0.033</math></b>

# Thank you 😊

- In summary:
  1. We propose a **new framework** for learning fully **continuous compositional representations** (Soft TPR + Soft TPR Autoencoder)
  2. Our approach is the **first** to learn fully **continuous compositional representations** in the **non-formal** domain of **vision**
  3. Extensive empirical results highlight the far-reaching benefits of our **representation's enhanced vector space alignment**, for representational structure, representation learners, and downstream models, underscoring the necessity of reconceptualising compositional representations in a fully continuous manner.
- Please see our **full paper** for more details on our approach, including proofs, conceptual motivation, theory, and suggestions for future work.
- **Code is available!**
- Questions? Thoughts? Contact [bethia.sun@unsw.edu.au](mailto:bethia.sun@unsw.edu.au)



Paper



Code

# References 1/2

- [1] Noam Chomsky. Syntactic Structures. The Hague: Mouton, 1957.
- [2] Jerry A. Fodor. The Language of Thought: A Theory of Mental Representation. Cambridge, MA: Harvard University Press, 1975.
- [3] Paul Smolensky. “Tensor product variable binding and the representation of symbolic structures in connectionist systems”. In: Artificial Intelligence 46.1 (1990), pp. 159–216.
- [4] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”. In: International Conference on Learning Representations. 2017.
- [5] Tameem Adel, Zoubin Ghahramani, and Adrian Weller. “Discovering Interpretable Representations for Both Deep Generative and Discriminative Models”. In: Proceedings of the 35th International Conference on Machine Learning. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 50–59.
- [6] Sjoerd van Steenkiste, Francesco Locatello, Jürgen Schmidhuber, and Olivier Bachem. “Are disentangled representations helpful for abstract visual reasoning?” In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019
- [7] F. Locatello, B. Poole, G. Rätsch, B. Schölkopf, O. Bachem, and M. Tschannen. “Weakly-Supervised Disentanglement Without Compromises”. In: Proceedings of the 37th International Conference on Machine Learning (ICML). Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 6348–6359
- [8] Elliot Creager, David Madras, Joern-Henrik Jacobsen, Marissa Weis, Kevin Swersky, Toniann Pitassi, and Richard Zemel. “Flexibly Fair Representation Learning by Disentanglement”. In: Proceedings of the 36th International Conference on Machine Learning. 2019.
- [9] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. “Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations”. In: Proceedings of the 36th International Conference on Machine Learning. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 4114–4124.
- [10] Sungho Park, Sunhee Hwang, Dohyung Kim, and Hyeran Byun. “Learning Disentangled Representation for Fair Facial Attribute Classification via Fairness-aware Information Alignment”. In: Proceedings of the AAAI Conference on Artificial Intelligence 35 (2021), pp. 2403–2411
- [11] H. Zhang, Y.-F. Zhang, W. Liu, A. Weller, B. Schölkopf, and E. Xing. “Towards Principled Disentanglement for Domain Generalization”. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022, pp. 8024–8034
- [12] Haoyang Li, Xin Wang, Zeyang Zhang, Haibo Chen, Ziwei Zhang, and Wenwu Zhu. “Disentangled Graph Self-supervised Learning for Out-of-Distribution Generalization”. In: Forty-first International Conference on Machine Learning. 2024.
- [13] Frederik Träuble, Elliot Creager, Niki Kilbertus, Francesco Locatello, Andrea Dittadi, Anirudh Goyal, Bernhard Schölkopf, and Stefan Bauer. “On Disentangled Representations Learned from Correlated Data”. In: Proceedings of the 38th International Conference on Machine Learning. Vol. 139. Proceedings of Machine Learning Research. 2021, pp. 10401–10412.

# References 2/2

- [14] Kezhen Chen, Qiuyuan Huang, Hamid Palangi, Paul Smolensky, Kenneth D. Forbus, and Jianfeng Gao. “Natural- to formal-language generation using Tensor Product Representations”. In: CoRR abs/1910.02339 (2019). arXiv: 1910.02339. URL: <http://arxiv.org/abs/1910.02339>
- [15] Qiuyuan Huang, Li Deng, Dapeng Wu, Chang Liu, and Xiaodong He. “Attentive Tensor Product Learning”. In: Proceedings of the AAAI Conference on Artificial Intelligence 33.01 (2019), pp. 1344– 1351
- [16] Imanol Schlag and Jürgen Schmidhuber. “Learning to Reason with Third-Order Tensor Products”. In: Advances in Neural Processing Information Systems. 2019.
- [17] R. Thomas McCoy, Tal Linzen, Ewan Dunbar, and Paul Smolensky. “Tensor Product Decomposition Networks: Uncovering Representations of Structure Learned by Neural Networks”. In: Proceedings of the Society for Computation in Linguistics 2020. Association for Computational Linguistics, 2020, pp. 277–278. URL: <https://aclanthology.org/2020.scil-1.34>.
- [18] Qiuyuan Huang, Paul Smolensky, Xiaodong He, Li Deng, and Dapeng Wu. “Tensor Product Generation Networks for Deep NLP Modeling”. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics.
- [19] Yichen Jiang, Asli Celikyilmaz, Paul Smolensky, Paul Soulos, Sudha Rao, Hamid Palangi, Roland Fernandez, Caitlin Smith, Mohit Bansal, and Jianfeng Gao. “Enriching Transformers with Structured Tensor-Product Representations for Abstractive Summarization”. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: Association for Computational Linguistics, 2021, pp. 4780–4793. DOI: 10.18653/v1/2021.naacl-main.381. URL: <https://aclanthology.org/2021.naacl-main.381>.
- [20] Taewon Park, Inchul Choi, and Minho Lee. “Attention-based Iterative Decomposition for Tensor Product Representation”. In: The Twelfth International Conference on Learning Representations. 2024. URL: <https://openreview.net/forum?id=FDb2JQZsFH>.
- [21] Diane Bouchacourt, Ryota Tomioka, and Sebastian Nowozin. “Multi-Level Variational Autoencoder: Learning Disentangled Representations From Grouped Observations”. In: Proceedings of the AAAI Conference on Artificial Intelligence 32.1 (2018).
- [22] Haruo Hosoya. “Group-based learning of disentangled representations with generalizability for novel contents”. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019, pp. 2506–2513.
- [23] Junxiang Chen and Kayhan Batmanghelich. “Weakly Supervised Disentanglement by Pairwise Similarities”. In: Proceedings of the AAAI Conference on Artificial Intelligence 34.04 (2020), pp. 3495– 3502
- [24] F. Locatello, B. Poole, G. Rätsch, B. Schölkopf, O. Bachem, and M. Tschannen. “Weakly-Supervised Disentanglement Without Compromises”. In: Proceedings of the 37th International Conference on Machine Learning (ICML). Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 6348– 6359.
- [25] Rui Shu, Yining Chen, Abhishek Kumar, Stefano Ermon, and Ben Poole. “Weakly Supervised Disentanglement with Guarantees”. In: International Conference on Learning Representations. 2020.