ETH zürich

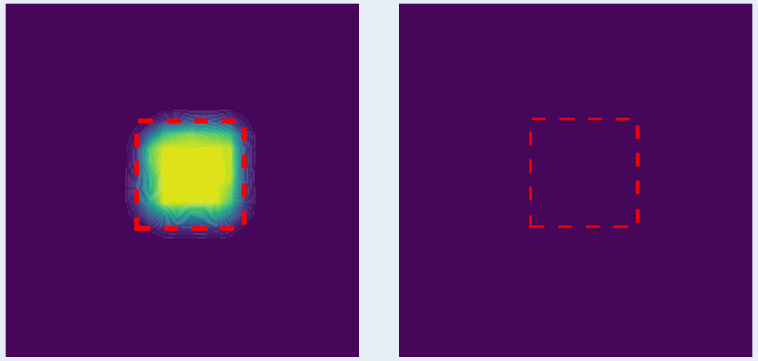
ETH AI CENTER

Institute of
Neuroinformatics
Grewe Lab

Learning Compositional Latent Structure with Vector Networks

Niclas Pokel, Benjamin F. Grewe
Institute of Neuroinformatics, UZH and ETH Zurich
NeurIPS 2026

Error on a region never seen in training



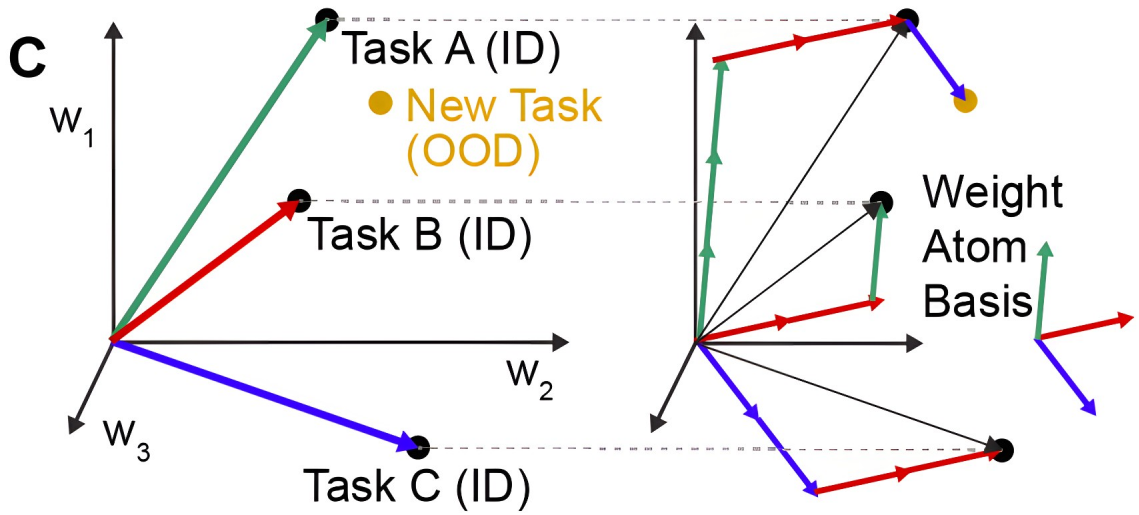
CNN **VN (ours)**

NEURAL INFORMATION PROCESSING SYSTEMS

Shared weight matrices entangle what a network learns

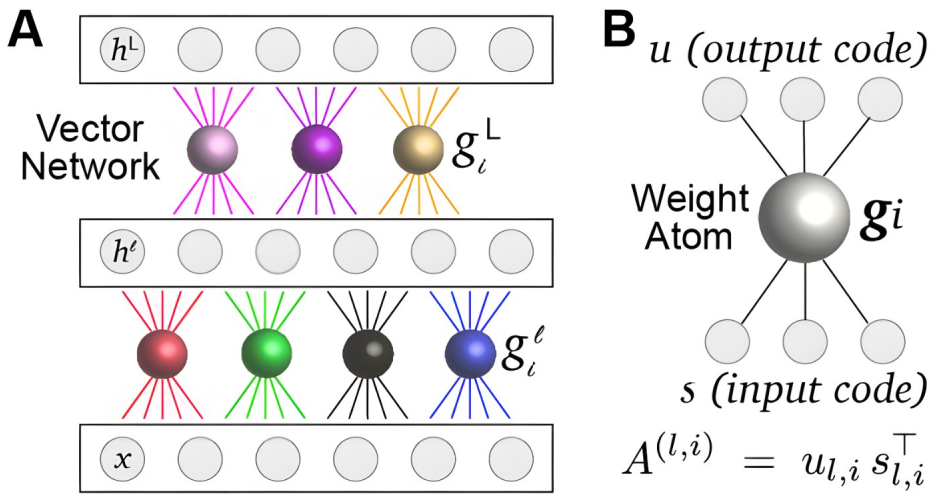
- Deep networks store many computations in one shared weight matrix.
- No part of it can be reused or updated on its own.
- Familiar factors in new combinations then generalise poorly.

Idea: make the transformation itself decomposable, as a library of reusable rank-1 weight atoms recombined for each input.



A dense W can only interpolate between learned tasks; a weight-atom basis reaches a new task by recombination.

Vector Networks infer which weight atoms to use for each input



A VN layer: atoms with inferred coefficients g (A); each atom pairs an input code s with an output code u (B).

1 **Compose.** Each input recruits a few atoms.

$$W^\star(x) = \sum_i g_i^\star(x)\, u_i\, s_i^\top$$

2 **Settle.** Sparse inference on a layer-local energy.

$$E(g) = \tfrac{1}{2}\|x - Sg\|_2^2 + \lambda\|g\|_1 + \tfrac{1}{2}\|h_{\text{target}} - Ug\|_2^2$$

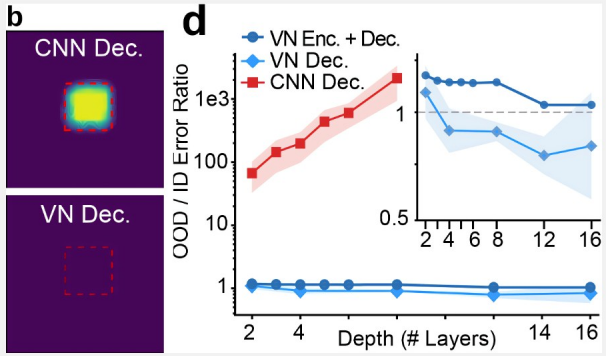
3 **Learn.** Atomic-Hebb updates only the active atoms.

$$S \leftarrow S + \rho\, r^x g^{\star\top}, \qquad U \leftarrow U + \rho\, r^h g^{\star\top}$$

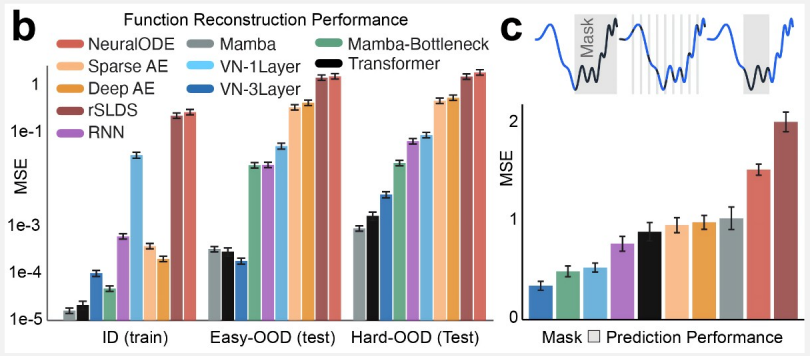
4 **Recombine.** Sums of atoms decode exactly.

$$S(g_a + g_b) = Sg_a + Sg_b$$

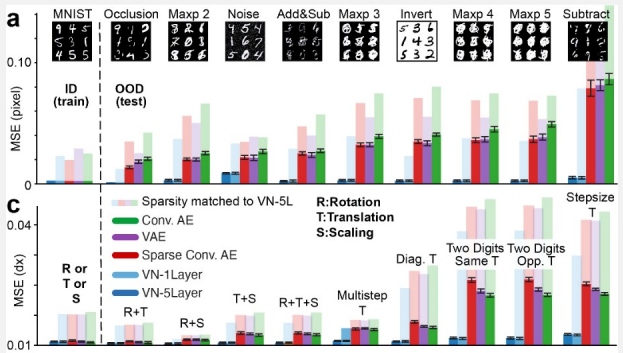
Tested on new combinations, VN often has ~10× lower error



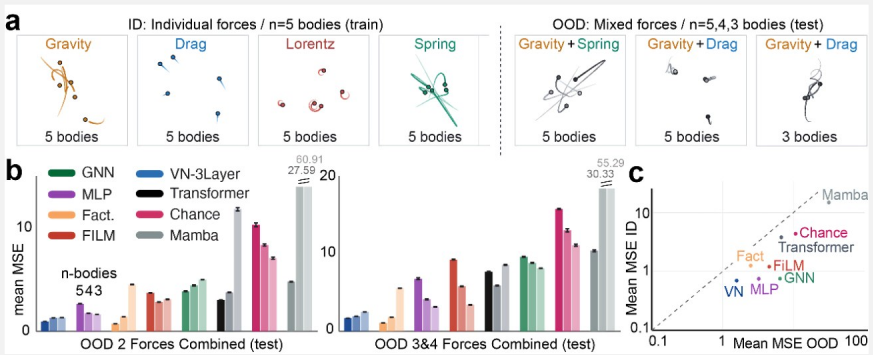
Spatial. The CNN fails on the held-out square; VN does not.



Functions. Masked prediction: VN ranks first on Hard-OOD.



Images. Hardest cases: ~10× lower error than autoencoders.



N-body. Mixed forces: VN stays closest to the ID = OOD line.

Take-home messages

- Each layer is a library of reusable rank-1 weight atoms.
- Fast sparse inference picks the atoms; slow Atomic-Hebb learning updates only those.
- Comparable in distribution, often ~10× lower error on new combinations.
- Limits: iterative inference is slower; guarantees cover additive compositions.

Paper on arXiv

arxiv.org/abs/2605.28007