Neuromorphic Hyperdimensional Visual Scene Factorization

Alpha Renner[1,2], Lazar Supic[3], Andreea Danielescu[3], Giacomo Indiveri[1], Bruno A. Olshausen[4], Yulia Sandamirskaya[5], Friedrich T. Sommer[4,5], E. Paxon Frady[5]

1. Institute of Neuroinformatics, University of Zurich and ETH Zurich 2. Forschungszentrum Jülich, Germany 3. Accenture Labs 4. Redwood Center for Theoretical Neuroscience, UC Berkeley 5. Intel Neuromorphic Computing Lab

Summary. In this talk at NNPC 2023, I will present a modular neuromorphic algorithm leveraging recent advancements in hyperdimensional computing/ Vector Symbolic Architectures (VSAs). VSAs have been proposed as a framework for symbolic reasoning, spatial, and graph operations on neuromorphic hardware. They make use of a small set of computational primitives that are robust, efficient, and compatible with diverse hardware. Our algorithm approaches scene understanding as a factorization problem and employs the resonator network to extract object identities and transformations. This is achieved by reflecting the algebraic structure of 2d rigid transforms (translations and rotation) in the neural VSA representation. Finally, we use a spike-timing-based implementation of phasor neurons to show an efficient proof of concept implementation on neuromorphic hardware and employ the model in a robotics task for visual odometry (visual SLAM).

Simultaneously inferring the identity and position of objects and their rigid transformations is still a challenging problem in visual scene understanding. We propose a neuromorphic solution leveraging three key concepts:

(1) A computational framework based on Vector Symbolic Architectures (VSA) with complex-valued vectors; VSAs provide a framework [5] for developing scalable algorithms that make use of massive parallelism, sparse event-based/asynchronous computation, in-memory compute, and local connectivity, hallmarks of neuromorphic computation. VSAs are suitable for various neuromorphic substrates; we demonstrate an implementation on Intel's asynchronous digital neuromorphic research chip Loihi [8, 7].

(2) Hierarchical Resonator Networks (HRNs) as an extension of the resonator model [3, 4] to handle the non-commutative nature of translation and rotation in visual scenes when both are used in combination.

(3) A spike-timing-based phasor neuron model for implementing complex-valued vector binding on neuromorphic hardware.

To demonstrate visual scene factorization, we focus on synthetic images of simple visual scenes composed of object (letter) templates (d) that are translated (h,v) and colored (c). The network's task is to estimate the object identities, colors, and locations from an image (see the network's output in Fig. 1D). We use VSA binding operations to produce generative image models in which binding ($\odot$, equivalent to multiplication) acts as the equivariant operation for geometric transformations [1] (Fig. 1A).

In order to feed an image (such as Fig. 1A) into the network, we encode images into hypervectors using complex-valued Fourier Holographic (FHRR) vectors [6]. A pixel at the Cartesian image coordinates x and y is represented by the index vector $\mathbf{h}^x \odot \mathbf{v}^y$ according to [1] (with v_j and $h_j = e^{\imath\phi_j}, \phi_j \sim \mathcal{U}[0, 2\pi]$). This kind of image encoding has pivotal properties for our proposed scene factorization algorithm as it ensures that the *equivariant vector operation* for image translation is the binding operation.

The generative model allows one to easily compose and render a synthetic scene, but inference in generative models is computationally expensive [10] as it involves a combinatorial search across all templates in all possible poses. Conveniently, the VSA encoding permits a fast parallel implementation of this search. The resonator iteratively estimates each factor by removing estimates of the other factors (Fig. 1C and E). The closer to the correct solution, the better the estimates of the factors support each other and "resonate." Note that the network is not trained on the combined factors, but thanks to the VSA binding, it is able to factorize combinations it has never seen (which could be called zero-shot compositional generalization). For a detailed model explanation, see [8].

To implement the resonator network on neuromorphic hardware, we use populations of phasor neurons [2] to represent the complex FHRR vectors, with the binding operation corresponding to spike timing shifts [8, 7]. Our proof of concept on Intel's neuromorphic research chip Loihi achieves more than two orders of magnitude improvement in energy consumption compared to a CPU implementation.

Furthermore, to incorporate rotation and scaling in addition to translation and color, we develop the hierarchical resonator (HRN), a partitioned architecture in which vector binding is equivariant for horizontal and vertical translation within one partition and for rotation and scaling within the other.

Finally, the HRN is put to use in a real-world robotics application scenario to solve visual odometry, i.e., estimation of the camera pose from visual input [9]. We use the HRN as a recursive filter that estimates three degrees of freedom of camera movement and allows sensory fusion of the visual and inertial modalities. Our approach outperforms trained neural network approaches on an event-based vision dataset. Our work is a step towards scalable, robust neuro-symbolic algorithms for neuromorphic hardware.

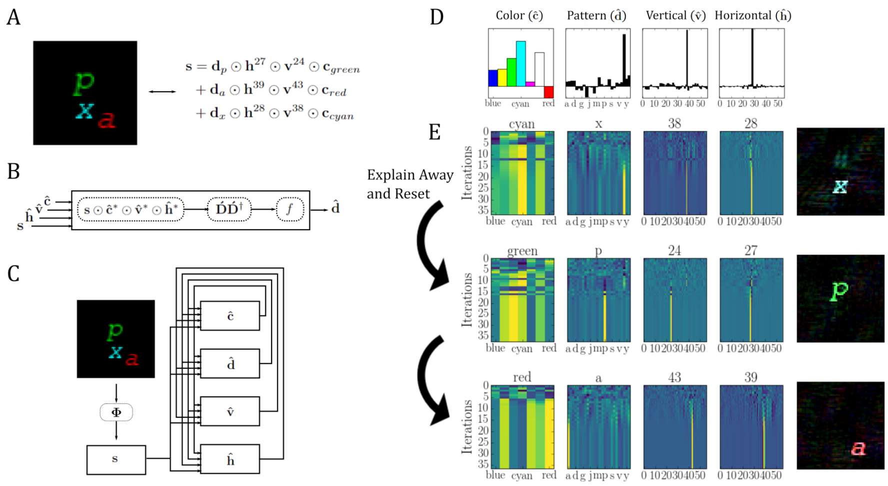

Figure 1: Resonator network for inferring shape, color, and translation. From [8] **A.** A synthetic scene and the VSA representation of the generative model. d:letter template; h:horizontal shift; v:vertical shift; c:color; $\odot$:binding; $+$:superposition. **B.** A resonator module consists of a binding stage, a clean-up stage, and an activation function. **C.** An input scene created by the generative model is encoded into the VSA space and is the input into the resonator network. The resonator modules communicate their estimates for each factor of variation in the generative model. **D.** The weighted factor estimates in each resonator module can be decoded using the codebooks, and the maximum value is taken as the output. **E.** The four dynamic estimates in the resonator network are visualized as heatmaps, with time represented vertically and each component horizontally. After several iterations, the resonator network converges to a solution and remains stable. Once converged, the component with the largest activation is chosen as the output (top of each panel). The decoded output is visualized to the right of each row. The object is then 'explained away' by subtracting the resonator's estimate from the scene vector. The resonator circuit is reset and converges to another solution, which describes a different object in the scene (rows 2 and 3).

References

[1] E. Frady, D. Kleyko, C. Kymn, B. Olshausen, and F. Sommer. Computing on functions using randomized vector representations. *arXiv preprint arXiv:2109.03429*, 2021. URL https://arxiv.org/abs/2109.03429.

[2] E. P. Frady and F. T. Sommer. Robust computation with rhythmic spike patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 116(36):18050–18059, sep 2019. doi: 10.1073/pnas.1902653116. URL http://dx.doi.org/10.1073/pnas.1902653116.

[3] E. P. Frady, S. J. Kent, B. A. Olshausen, and F. T. Sommer. Resonator networks, 1: An efficient solution for factoring high-dimensional, distributed representations of data structures. *Neural Computation*, pages 1–21, oct 2020. doi: 10.1162/neco_a_01331. URL http://dx.doi.org/10.1162/neco_a_01331.

[4] S. J. Kent, E. P. Frady, F. T. Sommer, and B. A. Olshausen. Resonator networks, 2: Factorization performance and capacity compared to optimization-based methods. *Neural Computation*, 32(12):2332–2388, dec 2020. doi: 10.1162/neco_a_01329. URL http://dx.doi.org/10.1162/neco_a_01329.

[5] D. Kleyko, M. Davies, E. Frady, and P. Kanerva. Vector symbolic architectures as a computing framework for nanoscale hardware. *arXiv preprint arXiv:2106.05268*, 2021. URL https://arxiv.org/abs/2106.05268.

[6] T. A. Plate. Holographic reduced representations. *IEEE Transactions on Neural networks*, 6(3):623–641, 1995.

[7] A. Renner, Y. Sandamirskaya, F. T. Sommer, and E. P. Frady. Sparse vector binding on spiking neuromorphic hardware using synaptic delays. In *International Conference on Neuromorphic Systems (ICONS 2022)*. ACM, 2022. doi: 10.1145/3546790.3546820. URL https://dl.acm.org/doi/abs/10.1145/3546790.3546820.

[8] A. Renner, L. Supic, A. Danielescu, G. Indiveri, B. A. Olshausen, Y. Sandamirskaya, F. T. Sommer, and E. P. Frady. Neuromorphic visual scene understanding with resonator networks. *arXiv preprint arXiv:2208.12880*, 2022. doi: 10.48550/arXiv.2208.12880. URL https://arxiv.org/abs/2208.12880.

[9] A. Renner, L. Supic, A. Danielescu, G. Indiveri, F. T. Sommer, E. P. Frady, and Y. Sandamirskaya. Neuromorphic visual odometry with resonator networks. *arXiv preprint arXiv:2209.02000*, 2022. doi: 10.48550/arXiv.2209.02000. URL https://arxiv.org/abs/2208.12880.

[10] Y. W. Teh, M. Welling, S. Osindero, and G. E. Hinton. Energy-based models for sparse overcomplete representations. *Journal of Machine Learning Research*, 4(Dec):1235–1260, 2003.