

TDSNNs: Competitive Topographic Deep Spiking Neural Networks for Visual Cortex Modeling

Deming Zhou^{1*}, Yuetong Fang^{1*}, Zhaorui Wang¹, Renjing Xu^{1†}

¹The Hong Kong University of Science and Technology (Guangzhou)
 {dzhou704, yfang870, zwang408}@connect.hkust-gz.edu.cn, renjingxu@hkust-gz.edu.cn

Abstract

The primate visual cortex exhibits topographic organization, where functionally similar neurons are spatially clustered, a structure widely believed to enhance neural processing efficiency. While prior works have demonstrated that conventional deep ANNs can develop topographic representations, these models largely neglect crucial temporal dynamics. This oversight often leads to significant performance degradation in tasks like object recognition and compromises their biological fidelity. To address this, we leverage spiking neural networks (SNNs), which inherently capture spike-based temporal dynamics and offer enhanced biological plausibility. We propose a novel Spatio-Temporal Constraints (STC) loss function for topographic deep spiking neural networks (TDSNNs), successfully replicating the hierarchical spatial functional organization observed in the primate visual cortex from low-level sensory input to high-level abstract representations. Our results show that STC effectively generates representative topographic features across simulated visual cortical areas. While introducing topography typically leads to significant performance degradation in ANNs, our spiking architecture exhibits a remarkably small performance drop (No drop in ImageNet top-1 accuracy, compared to a 3% drop observed in TopoNet, which is the best-performing topographic ANN so far) and outperforms topographic ANNs in brain-likeness. We also reveal that topographic organization facilitates efficient and stable temporal information processing via the spike mechanism in TDSNNs, contributing to model robustness. These findings suggest that TDSNNs offer a compelling balance between computational performance and brain-like features, providing not only a framework for interpreting neural science phenomena but also novel insights for designing more efficient and robust deep learning models.

Extended version — <https://arxiv.org/abs/2508.04270>

Introduction

The primate visual cortex processes information hierarchically, with the ventral stream comprising a series of cortical areas that facilitate visual recognition. This pathway begins in the primary visual cortex (V1) and progresses through intermediate regions, such as V4 (mid-tier visual cortical area),

*These authors contributed equally.

†Corresponding author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

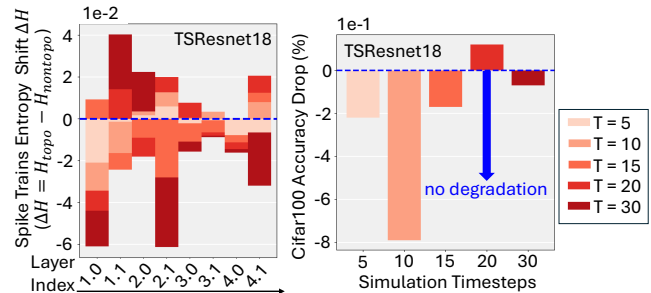


Figure 1: **TDSNNs leverage temporal information.** (Left) Spike train entropy shifts across layers reveal topography-dependent temporal dynamics with various inference timesteps. (Entropy is derived from the neurons’ firing probabilities). (Right) TDSNNs’ spiking mechanisms inherently solve topographic ANNs’ persistent recognition degradation by leveraging these temporal patterns.

culminating in high-level areas such as the inferior temporal (IT) cortex in macaques (Rolls 2000; Livingstone and Hubel 1984). Across different cortical areas, neurons that perform similar functions tend to be spatially grouped together, forming distinct neural clusters (Hubel and Wiesel 1962). This topographic organization results in primary region neurons being tuned to orientation, spatial frequency, and color (Ringach, Shapley, and Hawken 2002; De Valois, Albrecht, and Thorell 1982; Zeki 1983). Specifically, higher-level areas like IT feature neurons can further capture category-specific responses (e.g., faces, bodies) (Tsao et al. 2006; Downing et al. 2001).

In neuroscience, deep learning (LeCun, Bengio et al. 1995) facilitates the modeling of neural responses and the elucidation of underlying brain mechanisms (Dobs et al. 2022; Achterberg et al. 2023; Yamins and DiCarlo 2016). Topographic organization is one of the most significant aspects. Existing topographic artificial neural networks (ANNs) have successfully modeled hierarchical organization within visual pathways, from V1 to IT, encompassing regions beyond primary sensory cortex (Jacobs and Jordan 1992; Margalit et al. 2024; Qian et al. 2024; Deb, Deb, and Murty 2025). The temporal processing capability is a fundamental feature of biological brains (Mauk and Buonomano

2004). However, existing topographic ANNs largely overlook the full spatiotemporal dimension inherent in visual systems. This oversight, particularly the insufficient integration of temporal processing, demonstrably compromises their performance (Fig. 1(a)(b)).

Spiking neural networks (SNNs) (Maass 1997) represent a further step towards biologically plausible ANNs, where each fundamental computing unit is modeled as a neuron. SNNs’ neural dynamics enable intrinsic temporal processing, unlike ANNs, which rely on network architecture for extrinsic time handling. It is a straightforward idea to capture brain-like representations (Brette et al. 2007) and predict neural responses (Huang et al. 2023, 2024) by employing SNNs. However, existing works often overlook the emergence of topographic organization especially for deep SNN architectures (Zhong et al. 2024).

In this paper, we present *topographic deep spiking neural networks* (TDSNNs). We address the critical limitations of current topographic models: the absence of temporal dynamics in multi-layered topographic ANNs and the confinement of topographic SNNs to primary encoding layers. By integrating hierarchical depth and spatiotemporal processing, TDSNNs enable the first systematic investigation into the emergence and functional implications of hierarchical topographic organization within a spatiotemporal context. We demonstrate that TDSNNs not only replicate biologically observed topographic features but also achieve high task performance, shedding light on the underlying mechanisms of efficient information processing and enhanced robustness, providing a potential framework to understand the spatiotemporal dynamics of visual cortical function.

Related Works

Topographic Vision Models Prior works have explored the topographic organization in neural network models, demonstrating how lateral interactions can self-organize orientation selectivity (Von der Malsburg 1973; Willshaw and Von Der Malsburg 1976). To incorporate topography into deep neural networks, several studies have introduced auxiliary objectives, inspired by biological constraints (e.g., wiring cost minimization) (Jacobs and Jordan 1992; Koulakov and Chklovskii 2001), to reduce the spatial distance between neighboring units mapped onto a cortical sheet (Margalit et al. 2024; Blauch, Behrmann, and Plaut 2022). However, these models often overlook critical biological mechanisms (e.g., recurrent connections) and suffer from degraded performance in classification tasks. Recent advancements have further refined deep topographic models by integrating insights from a connectionist perspective. Specifically, (Dehghani et al. 2024) redesigned self-organizing maps with top-down learning mechanisms, significantly reducing performance trade-offs. (Qian et al. 2024) demonstrates that local lateral connectivity alone is sufficient to drive topographic organization. Furthermore, (Deb, Deb, and Murty 2025) proposes a neural pruning method to balance layer-wise topography and task performance, with demonstrated efficacy across various architectures, including transformers. However, none of these approaches fully incorporate the temporal dimension or ex-

plores spatiotemporal topographic organization. (Blauch, Behrmann, and Plaut 2022) attempts to bridge this gap with interactive topographic networks (ITNs) using recurrent architectures, but their scope is confined to high-level visual regions.

Spiking Neural Networks Spiking Neural Networks (SNNs) leverage bio-inspired computational units (e.g., LIF, HH models) (Maass 1997; Gerstner and Kistler 2002) to integrate temporal dynamics absent in conventional ANNs (Tavanaei et al. 2019). This temporal capability endows SNNs with high computational efficiency (Pei et al. 2019) and robust temporal processing that enhances noise resilience (Ding et al. 2023), making them suitable for applications including object recognition (Hu et al. 2023), image segmentation (Patel et al. 2021), and generative models (Cao et al. 2024). Their biological plausibility renders SNNs valuable for neuroscience research, complementing RNNs in neural circuit modeling (Basu et al. 2022) and brain data analysis (Kasabov 2014). Recent work demonstrates SNNs’ superior representational similarity to the visual cortex over ANNs, highlighting their promise as biologically grounded computational models (Huang et al. 2023, 2024).

Topographic Vision Models with Spiking Mechanism

In computational neuroscience, Spike mechanism have been successfully applied to model early visual cortical areas, particularly the primary visual cortex (V1), incorporating both topographic organization and abundant neural principles (Antolík et al. 2024; Billeh et al. 2020). However, their neurobiological fidelity often precludes end-to-end training, limiting investigations into the developmental emergence of interareal topographic organization. Despite advances in large-scale SNN training (e.g., ANN-SNN conversion (Hu et al. 2023), STBP (Wu et al. 2018)), the integration of topographic organization into deep and trainable SNNs remains underexplored. A notable exception is the Self-Evolving Spiking Neural Network (SESNN) (Zhong et al. 2024), which replicates orientation preference maps in V1, marking an initial step toward topographic SNNs. However, the network is designed with just two layers. In contrast to SESNN that lose topographic organization in deeper layers, TDSNNs preserve spatial relationships across all visual hierarchy levels (from V1 to IT).

Approach

Mapping SNN Layer to a Virtual Physical Space

The first step in constructing a topographic SNN is creating a hierarchical spatial structure modeled after the ventral visual cortex. We adopt a cortical sheet design similar to that of (Margalit et al. 2024), facilitating biologically plausible retinotopic organization in our SNN model with LIF neurons (See Appendix). Specifically, given a layer with a dimension (C, H, W) in the SNN (suppose there are C channels and the size of each feature map is $H \times W$), we non-uniformly embed these units into a cortical sheet (Fig. 2(a)). Let U_{SNN} be the set of all units in the SNN layer:

$$U_{SNN} = \{(c, h', w') \mid 1 \leq c \leq C, 1 \leq h' \leq H, 1 \leq w' \leq W\}. \quad (1)$$

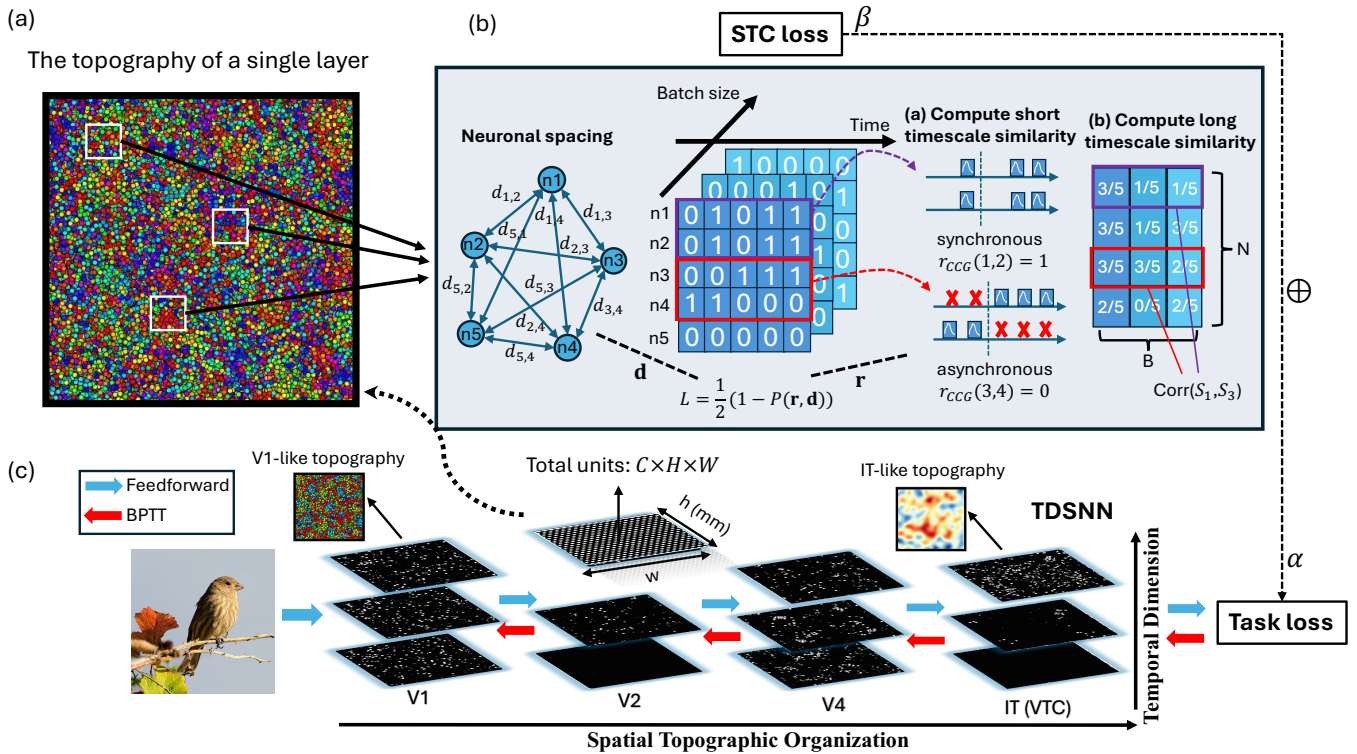


Figure 2: **Overview of the methodology for inducing visual cortex-like neural organization in SNN architectures.** (a) Illustration of the virtual 2D cortical sheet assigned to each layer of the SNN. Each scatter represents a neuron. (b) Spatio-Temporal Constraints (STC) is designed to promote similar response patterns in spatially nearby neurons across both long-time and short-time scales. (c) Schematic of the Training Pipeline for TDSNNs.

Each unit $u_{c,h',w'} \in U_{SNN}$ is assigned a unique two-dimensional coordinate (x, y) on the cortical sheet of size $h \times w$ (where h and w are predefined dimensions of the cortical sheet, in millimeters). This non-uniform embedding can be formally expressed as an injective mapping:

$$\mathcal{M} : U_{SNN} \rightarrow [0, h] \times [0, w], \quad (2)$$

where for each unit $u_{c,h',w'}$, its assigned coordinate is $(x_{c,h',w'}, y_{c,h',w'}) = \mathcal{M}(c, h', w')$.

Neuronal Positions Pre-optimization

To achieve topographic organization from randomly assigned 2D unit coordinates, pre-optimization of unit positions is necessary (Margalit et al., 2024). (1) Pre-train an auxiliary SNN on a task objective using Back-propagation Through Time (BPTT) with surrogate gradient (Neftci, Mostafa, and Zenke 2019). (2) Stochastically swap unit positions based on the pre-trained SNN’s firing rates in response to sine grating stimuli, to promote similar response patterns in adjacent units. (3) All cortical sheets remain fixed for subsequent loss calculation. (Details including the necessity of pre-optimization are provided in the Appendix).

Spatio-Temporal Constraints Loss

We propose the *Spatio-Temporal Constraints* (STC) loss to promote topographic organization in SNNs. This is motivated by the biological principle that neural networks evolve

under competing pressures: minimizing metabolic costs of physical connectivity while maximizing information-theoretic efficiency (Bullmore and Sporns 2009). Building on this, prior studies have demonstrated that balancing task performance with metabolic (spatial) constraints, often through auxiliary wiring cost functions, can induce visual cortex-like features or small-world topologies (Orlov, Makin, and Zohary 2010; Achterberg et al. 2023). Furthermore, biological systems like the macaque visual cortex exhibit both long-timescale (firing rate-based) and short-timescale (synchrony-based) representations (Kohn and Smith 2005), with millisecond-scale neuronal synchronization being a ubiquitous and computationally fundamental feature (Lestienne 2001; Sharma et al. 2022). Inspired by these combined spatial and temporal insights, our auxiliary STC loss, coupled with a task-specific loss, fosters spatial and temporal similarity among adjacent neurons.

The STC method enhances response similarity among neighboring neurons in SNNs by jointly optimizing both long-timescale firing rates and short-timescale spike timing synchrony (see Fig. 2(b)). Here we consider a set of N neurons in a single layer of SNNs, where the spiking activity of a neuron at time t is represented by $S(t)$ with T time steps in total, taking the value 1 if the neuron fires and 0 otherwise. To quantify long-timescale correlations between neurons, we first compute each neuron’s mean fir-

ing rate vector across B trials (where B corresponds to the batch size in SNN training). For neuron i , the firing rate vector is $\mathbf{S}_i = (\langle S_i^1 \rangle_T, \dots, \langle S_i^B \rangle_T)$, where $\langle S_i^b \rangle_T = T^{-1} \sum_{t=1}^T S_i^b(t)$. Here, $S_i^b(t)$ denotes the spike train of neuron i in trial b at time t , and $\langle \cdot \rangle_T$ represents temporal averaging over the simulation window T . For each of the $\binom{N}{2}$ neuron pairs (i, j) , we calculate the Pearson correlation coefficient, $\mathbf{r} = \text{PearsonCorr}(\mathbf{S}_i, \mathbf{S}_j)$, yielding a correlation vector \mathbf{r} . Additionally, leveraging the predefined coordinates of neurons in the cortical sheet, we compute the pairwise Euclidean distances to form an inverse distance vector \mathbf{d} . To further directly model the relationship between similarity and spatial proximity, we construct a loss function following the approach of (Margalit et al. 2024):

$$\mathcal{L}_L = \frac{1}{2} (1 - \text{P}(\mathbf{r}, \mathbf{d})), \quad (3)$$

where $\text{P}(\cdot)$ denotes Pearson’s r . The Long-timescale loss term \mathcal{L}_L decreases as the firing rates of spatially adjacent neurons in the cortical sheet become more similar, and increases otherwise. To capture information encoded in short-timescale spike timing, we incorporate the spike train cross-correlogram (CCG) (Cutts and Eglen 2014) to measure temporal synchrony between neuron pairs:

$$\text{CCG}(i, j) = \sum_{\tau=-W}^W \left[\frac{1}{B\lambda(\tau, T)} \sum_{b=1}^B \sum_{t=1}^{T-|\tau|} S_i^b(t) S_j^b(t + \tau) \right],$$

$$\lambda(\tau, T) = \max(0, T - |\tau|). \quad (4)$$

The CCG employs a specified time window size W , with varying shift indices $\tau \in [-W, W]$ to capture the short-timescale synchrony of neuron i ’s firing events relative to neuron j . The normalization factor $\lambda(\cdot)$ compensates for the reduction in available spike train data due to time lags. Additionally, the CCG requires autocorrelation normalization:

$$r_{\text{CCG}}(i, j) = \frac{\text{CCG}(i, j)}{\sqrt{\text{ACG}(i) \cdot \text{ACG}(j)}}, i \neq j \quad (5)$$

The autocorrelograms (ACG) in Eq. 5 are calculated similarly to the CCG, except that each spike train of a neuron is compared with itself, such that $\text{ACG}(i) = \text{CCG}(i, i)$. When $i = j$, r_{CCG} equals to 1. Notably, through normalization, the range of r_{CCG} is constrained to $[0, 1]$. Values closer to 1 indicate higher synchronization of neuronal responses, and vice versa. Thus, we design a loss function \mathcal{L}_S that reflects response similarity and spatial proximity on a short timescale:

$$\mathcal{L}_S = \frac{1}{2} (1 - \text{P}(\mathbf{r}_{\text{CCG}}, \mathbf{d})). \quad (6)$$

Training for the Emergence of Topography

The final step involves training a SNN model from scratch to obtain the TDSNN (refer to Fig. 2(c)). Following weights initialization, SNNs are optimized by minimizing a total loss function comprising a task-specific loss ($\mathcal{L}_{\text{task}}$, e.g., cross-entropy) and the STC loss terms. In practice, for computational efficiency, the STC loss terms \mathcal{L}_L and \mathcal{L}_S are computed by randomly sampling small clusters of neurons

within each layer and averaging their contributions as indicated by (Margalit et al. 2024). Given K layers, M random selected neuron clusters in a layer, the final loss is:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \frac{1}{M} \sum_{k=1}^K \sum_{m=1}^M [\alpha \mathcal{L}_L(k, m) + \beta \mathcal{L}_S(k, m)], \quad (7)$$

where α and β are weighting factors to control the STC loss term. As shown in Fig. 2(c), in each training iteration, the STC loss, computed for each appointed SNN layer and summed, forms an additional objective optimized jointly with the primary task loss using BPTT with surrogate gradient. (Training details are provided in the Appendix).

Results and Analysis

Experimental Settings

As primary TDSNNs architectures, we employ the feed-forward Spiking ResNet-18 (SResnet18) (Hu, Tang, and Pan 2021) and Spikformer (Zhou et al. 2022), alongside Spiking CORnet-RT (SCornet) (Kubilius et al. 2019), a spiking recurrent network featuring inter-layer self-connections. We refer to their topographic versions as TSResNet18, TSpikformer and TSCornet. For SResnet18, we construct a 2D cortical sheet (see Fig. 2(c)) with specified height and width for the feature map of the final layer in each residual block. A similar approach is applied to SCornet and Spikformer, with the distinction that, for Spikformer, the cortical sheet is constructed for the linear layer following the attention module. To streamline experimental analysis, we set the number of time steps to 4 for feedforward SNNs SResnet18 and Spikformer, and to 10 for SCornet. All networks employed LIF neurons with a membrane time constant of 2.0. We set weighting factors of STC loss as 50.0 (i.e., α and β in Eq. 7). To reduce computational load, neuron clusters for computing STC loss in each SNN layer are confined to fixed-size square regions, with multiple clusters randomly sampled. All SNN and TDSNN models are trained on the ImageNet dataset (Deng et al. 2009) directly using BPTT with surrogate gradient, unless specified (Details in Appendix).

TDSNNs Exhibit V1-like Topography

Neurons in the primate V1 demonstrate a well-organized topographic structure, featuring systematic maps of preferred stimulus orientation, spatial frequency, and color (Livingstone and Hubel 1984; Hubel and Wiesel 1962), with those sharing similar response properties grouped into vertically oriented ”columns” relative to the cortical surface. We employed four distinct models as baselines: TSResnet18 and TSCornet, which were trained with the proposed STC loss combined with a task objective to induce topography (Eq. 7), and SResnet18 and SCornet, trained solely with the task loss. To assess whether these models exhibited V1-like topographic organization, we adopted the methodology established in (Margalit et al. 2024; Dehghani et al. 2024). Our analysis focused on characterizing neuronal preferences for orientation, spatial frequency, and color (Fig. 3(a)).

We initially constructed tuning curves for individual neurons in the model’s V1 layer. The preferred stimuli from

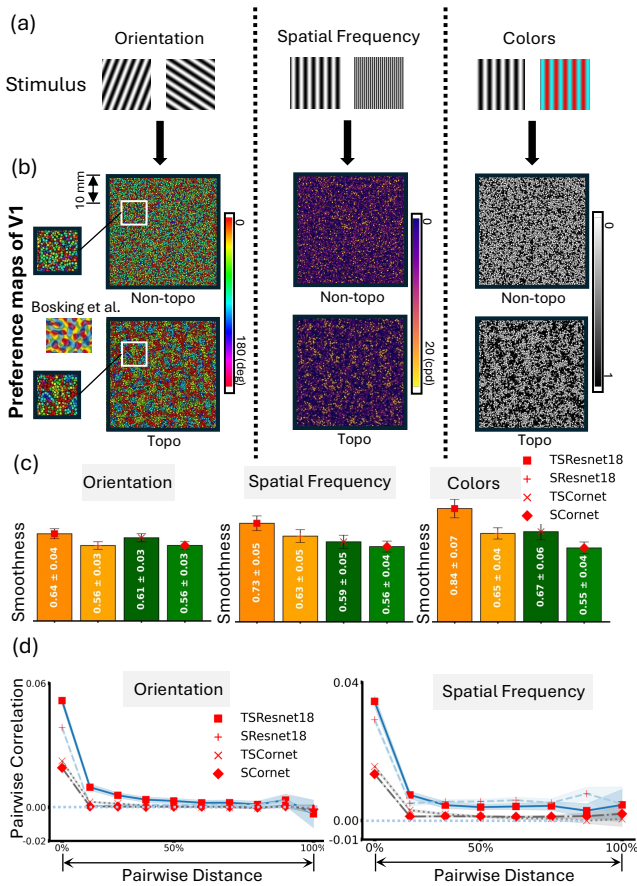


Figure 3: Analysis of V1-like topography of TDSNNs. (a) Sine grating stimuli used to probe neural responses, as described in (Margalit et al. 2024). (b) Preference maps for orientation, spatial frequency, and color in Layer 2.0. Top row: non-topographic SResNet18. Bottom row: topographic TSResNet18. Orientation preference maps were generated via vector summation of angle-specific response data (Bosking et al. 1997). (c) Smoothness analysis of orientation, spatial frequency, and color preferences. Higher smoothness denote greater similarity in responses among closely located neurons, indicating smoother transitions in preference maps. (See Appendix for details; error bars: SEM). (d) Pairwise firing rate correlation as a function of spatial distance for orientation preference (with 95% confidence intervals).

these curves were visualized as 2D preference maps, e.g., the orientation preference map in Fig. 3(b). Prominent pinwheel patterns, typical of biological V1, were consistently observed in TDSNNs. For quantitative analysis, we examined the relationship between neuronal firing similarity and spatial distance. Fig. 3(d) illustrates that the pairwise correlation of orientation preferences between neurons decreases as spatial separation increases, showing an inverse relationship between preference similarity and distance. The pairwise correlation between nearby neurons is higher in TD-SNNs than in non-topological SNNs. Additionally, the topographic organization in TDSNNs led to noticeably higher

smoothness in spatial and color variation of neuronal preferences than in non-topographic counterparts (Fig. 3(c)). (See more experimental results in Appendix).

IT-Analogous Category Selectivity in Deep Layers

The functional organization of IT/VTC is characterized by the spatial clustering of neurons tuned to ecologically relevant categories (e.g., faces, places, limbs, visual wordforms) into distinct patches, exhibiting specific sizes, numbers, and inter-patch spacing (Grill-Spector and Weiner 2014). Studies have demonstrated that under the combined influence of spatial constraints and task objectives, category selectivity emerges in deeper layers of ANNs (Margalit et al. 2024; Deb, Deb, and Murty 2025). Extending this, we examine if SNNs, influenced by STC, develop IT-like organization characterized by category selectivity. We employed three stimulus datasets: the fLoc dataset (Stigliani, Weiner, and Grill-Spector 2015), Big-small (Konkle and Oliva 2012), and Origin-Transform (Long, Yu, and Konkle 2018) (Fig. 4(a)), to characterize the selectivity profiles of neural responses and construct corresponding selectivity maps. Relative to non-topo SNNs, the last layer of TD-SNNs exhibited a more pronounced spatial clustering of responses, forming spatially contiguous "continent-like" patterns (Fig. 4(a)). Neurons in close proximity demonstrate greater selectivity similarity compared to a random spatial arrangement (Fig. 4(b)). Furthermore, the category selectivity of adjacent LIF neurons in TDSNNs varied more smoothly across the layer (Fig. 4(c)). Concurrently, we observed that selectivity for faces and bodies was spatially co-localized (overlap correlation comparison: 0.63/TSResnet18 vs 0.15/SResnet18). Conversely, selectivity for characters and places was spatially segregated (overlap correlation comparison: 0.16/TSResnet18 vs 0.47/SResnet18). (More details in Appendix).

TDSNNs Achieve High Task Performance and Brain-likeness

Applying topographic architecture to ANNs involves a trade-off between model performance (task performance and brain-likeness score) and the extent of topography. To evaluate this trade-off in topographic SNNs, we used two methods: prediction accuracy degradation and Brain-Score (Schrimpf et al. 2018).

Object Recognition Task Performance To date, topographic vision ANNs have largely adopted the ResNet18 architecture. For comparative purposes, TSResNet18 is utilized. As reported by previous work, the performance degradation on the ImageNet dataset between topographic and non-topographic models typically ranges from 3% (TopoNet) to 16.57% (LLCNN-G) (Fig. 5(a)). We observe no performance degradation in our TDSNN.

Analysis of L_L and L_S in STC Experiments with TSResnet18 on Imagenet reveal that the short-timescale loss term (L_S , controlled by β) is crucial for promoting temporal coding. When $\beta = 0$ (i.e., no L_S term), the network primarily induces topography in a rate-coded manner, akin to ANNs.

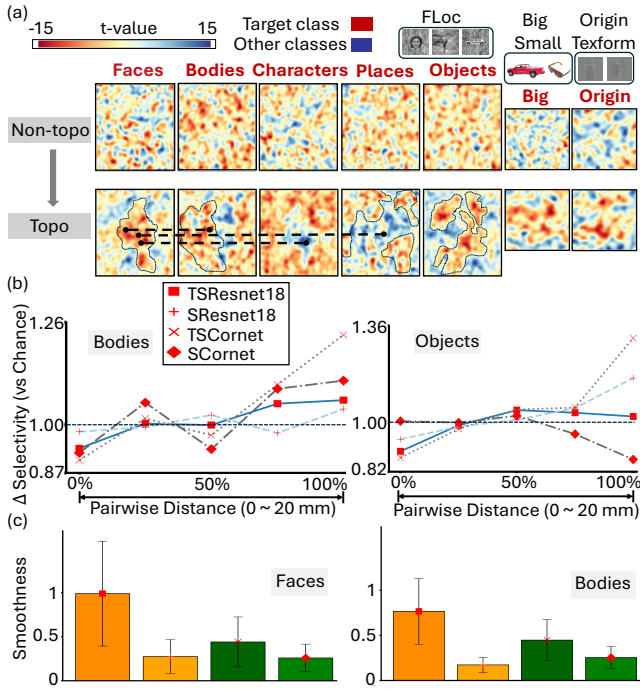


Figure 4: **Analysis of IT-like topography of TDSNNs.** (a) Category t-value selectivity maps of the final layer are shown for SResNet18 (non-topo) and TSResNet18 (topo). The topographic organization in TSResNet18 exhibits a more clustered "continent" form, indicating larger neural clusters for similar functional representations. Areas of high selectivity for faces overlap with those for bodies, whereas areas for characters and places are spatially segregated, as indicated by black dots (See Appendix for t-value definition). (b) Difference in selectivity as a function of pairwise neuronal distance for bodies and objects. (c) Smoothness analysis of faces and bodies t-value maps.

However, L_S functions as a spike timing regulator, significantly enhancing temporal coding, as evidenced by improved smoothness (0.7674 for $\alpha 50 - \beta 50$ vs. 0.755 for $\alpha 50 - \beta 0$) and classification accuracy (58.34% for $\alpha 50 - \beta 50$ vs. 58.21% for $\alpha 50 - \beta 0$). Similar improvements were observed for the $\alpha 10 - \beta 10$ vs. $\alpha 10 - \beta 90$ comparison.

Topographic Organization Extent and Task Performance Following (Deb, Deb, and Murty 2025), we investigated the relationship between topographic organization (quantified by V1-like layer smoothness in orientation preference maps) and model performance. By adjusting hyperparameters α and β in Eq. 7 that control the nontopographic-to-topographic transition, We observed that TSResnet18's prediction accuracy even increased, reaching a maximum of 58.72%, when α and β were set to 10.0 and 90.0 (trained on Imagenet), respectively. Meanwhile V1-like layer smoothness significantly improved (0.57 to 0.76). Notably, on CIFAR100 (Krizhevsky, Hinton et al. 2009), topographic organization formation also improved TSResnet18's accuracy (see Fig. 5(b), non-topographic SResnet18: 73.01% vs.

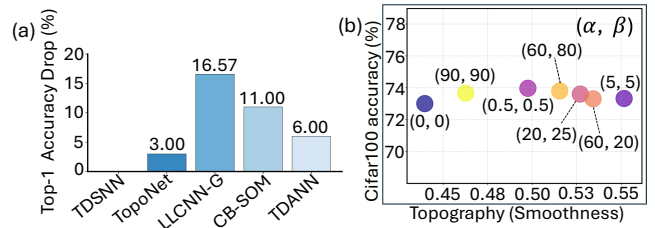


Figure 5: **Performance comparisons between TDSNNs and topographic ANNs.** (a) TDSNN has no top-1 accuracy drop in ImageNet compared to topographic ANNs (Deb, Deb, and Murty 2025; Qian et al. 2024; Dehghani et al. 2024; Margalit et al. 2024). (b) TDSNN maintains competitive prediction accuracy on CIFAR100 across varying topographic organization strengths.

Model	V1	V2	V4	IT
TopoNet*	0.7116	0.3038	0.2923	0.5723
TDANN*	0.6932	0.1775	0.2792	0.4259
ANN*	0.6913	0.3038	0.2346	0.5953
SNN	0.6823	0.3079	0.3970	0.7102
TDSNN (ours)	0.6845	0.3021	0.3886	0.7127

Table 1: BrainScore results (* denotes ANN architecture)

TSResnet18 with α and β set to 0.5: 73.97%). We also introduce topography into non-CNN architectures, specifically Spikformer, and observe no drop in prediction accuracy. TDSNNs not only achieve robust prediction performance but also concurrently exhibit desirable topographic organization. (See more details in Appendix).

Brain-likeness BrainScore employs benchmarks and various evaluation metrics (e.g., neural data prediction) to quantify *the extent to which neural network models replicate brain mechanisms for core object recognition*. For fair comparison, brain-likeness evaluation is conducted by using the benchmark selection of (Deb, Deb, and Murty 2025) and the same structure (Resnet18). SNNs exhibit greater brain-likeness and better V2, V4, IT performance over ANNs, highlighting the importance of temporal information (Tab. 1). Our TDSNNs, nearly matching SNNs in V2/V4 (0.3021/0.3079, 0.3886/0.3970) while outperforming them in V1/IT (+0.22%, +0.25%), proving the additional topology constraint contributes to brain-likeness.

Topography-Driven Information Hierarchy

The topographic organization in the brain is believed to enhance the efficiency of neural processing (Karbasforoushan, Tian, and Baker 2022). Studies (Margalit et al. 2024; Deb, Deb, and Murty 2025; Qian et al. 2024; Zhong et al. 2024) exhibit that topographic networks (leveraging lateral connections and spike timing) are more parameter efficient than non-topographic counterparts, achieving higher accuracy after L1 pruning and demonstrating robustness to mild noise. Building on these findings, we explore the functional

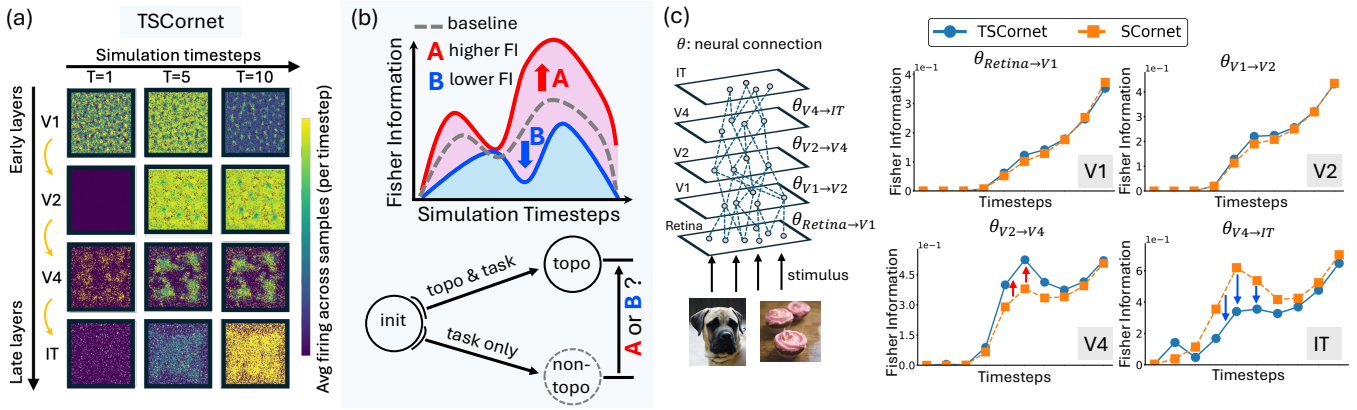


Figure 6: **Inducing topography fundamentally reshapes temporal information processing in SNNs.** (a) The spiking activity pattern across all layers at each timestep (ImageNet validation set). (b) A comparative experiment was conducted to analyze specific differences in neural connections (modes A or B) between topographic and non-topographic SNNs. (c) Analysis of Fisher information across visual regions during network inference. (See Appendix for the results of TSResnet18)

changes induced by topographic organization in SResnet18 and SCornet (TDSNNs vs SNNs as comparison experiments). We used the following computational metric:

- **Fisher information:** Fisher information (FI) offers an intuitive means to elucidate the temporal significance of a model’s internal parameters, thereby *revealing the concentration of functionally critical neural connections within the network* (Fisher 1925; Achille, Rovere, and Soatto 2018). Specifically, for SNNs, FI across the time domain is computed as:

$$I_t = \frac{1}{N} \sum_{n=1}^N \left\| \nabla_{\theta} \log f_{\theta}(y | x_{i \leq t}^n) \right\|^2. \quad (8)$$

$f_{\theta}(y | x)$ represents the posterior probability distribution of the SNN parameterized by model weights (neural connection strength) θ , with output $y \sim f_{\theta}(y | x)$. Given N training samples, the FI of the model at a specific time t is characterized by I_t (see Appendix for details).

Increased FI signifies a neural connection’s **higher importance** to incoming input features. While beneficial for precise processing, this sensitivity simultaneously makes the connection **less robust** against substantial input signal perturbations (Kim et al. 2023).

We first visualized the internal network activity across layer of TDSNNs and found a stronger tendency for neighboring neurons to fire synchronously compared to SNNs (Fig. 6(a) and Appendix). Furthermore, to understand how topography influences temporal information processing by altering neuronal firing patterns, we analyzed the Shannon entropy (i.e., information capacity of spike trains) across each layer of both TDSNNs and SNNs. We observed a significant shift in information capacity between SNNs and TDSNNs, with variations from early to late layers (Appendix and Fig. 1(a)). Next, we investigate changes in neural connectivity across visual regions (Fig. 6(b), topo vs non-topo). As shown in

Fig. 6(c), these results reveal a dynamic hierarchical progression: starting with early layers (V1 and V2) that preserve raw signal fidelity (stable FI); transitioning through V4, which critically amplifies discriminative features (mode A, increased FI in across all inference timesteps); and culminating in IT’s stabilized, noise-resistant encoding (mode B, decreased FI in throughout inference). The formation of topography significantly reshapes temporal function, evident in the alterations of inter-layer synaptic connections, with this reshaping mainly occurring in the deeper layers.

Robustness We conducted four types of attacks: Gaussian noise, FGSM, PGD, and random pixel masking at each timestep (Appendix for more details). With $(\alpha, \beta) = 10.0$, TSResnet18 consistently outperformed non-topo SResnet18 in robustness (25.8% vs 24.5%, 24.4% vs 23.6%, 10.7% vs 9.97%, 21.0% vs 20.8%) despite similar clean accuracy (58.5%). This suggests topographic connections enhance the robustness of decision making in recognition task.

Conclusions

We present Topographic Deep Spiking Neural Networks (TDSNNs), overcoming key limitations in existing topographic models by unifying hierarchical depth with spatiotemporal processing. TDSNNs successfully replicate biologically observed topographic features and achieve high performance, revealing the fundamental mechanism underlying efficient processing and enhanced robustness. Our work provides a novel perspective for understanding the evolution of biological neural systems by developing large-scale, trainable deep learning models that incorporate biologically plausible elements (e.g., long-range connections, distinct populations of excitatory and inhibitory neurons).

Acknowledgements

This work was supported by Guangdong S&T program (2025A0505000036), the Guangzhou-HKUST(GZ) Joint Funding Program (Grant No. 2023A03J0682), the National

Natural Science Foundation of China (Grant No. 62405255), Guangdong Basic and Applied Basic Research Foundation (No. 2023A1515110679), and partially supported by the collaborative project with Brain Mind Innovation, Inc.

References

- Achille, A.; Rovere, M.; and Soatto, S. 2018. Critical learning periods in deep networks. In *International Conference on Learning Representations*.
- Achterberg, J.; Akarca, D.; Strouse, D.; Duncan, J.; and Astle, D. E. 2023. Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings. *Nature Machine Intelligence*, 5(12): 1369–1381.
- Antolík, J.; Cagnol, R.; Rózsa, T.; Monier, C.; Frégnac, Y.; and Davison, A. P. 2024. A comprehensive data-driven model of cat primary visual cortex. *PLOS Computational Biology*, 20(8): e1012342.
- Basu, A.; Deng, L.; Frenkel, C.; and Zhang, X. 2022. Spiking neural network integrated circuits: A review of trends and future directions. In *2022 IEEE Custom Integrated Circuits Conference (CICC)*, 1–8. IEEE.
- Billeh, Y. N.; Cai, B.; Gratiy, S. L.; Dai, K.; Iyer, R.; Gouwens, N. W.; Abbasi-Asl, R.; Jia, X.; Siegle, J. H.; Olsen, S. R.; et al. 2020. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, 106(3): 388–403.
- Blauch, N. M.; Behrmann, M.; and Plaut, D. C. 2022. A connectivity-constrained computational account of topographic organization in primate high-level visual cortex. *Proceedings of the National Academy of Sciences*, 119(3): e2112566119.
- Bosking, W. H.; Zhang, Y.; Schofield, B.; and Fitzpatrick, D. 1997. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of neuroscience*, 17(6): 2112–2127.
- Brette, R.; Rudolph, M.; Carnevale, T.; Hines, M.; Beeman, D.; Bower, J. M.; Diesmann, M.; Morrison, A.; Goodman, P. H.; Harris, F. C.; et al. 2007. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of computational neuroscience*, 23: 349–398.
- Bullmore, E.; and Sporns, O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3): 186–198.
- Cao, J.; Wang, Z.; Guo, H.; Cheng, H.; Zhang, Q.; and Xu, R. 2024. Spiking denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 4912–4921.
- Cutts, C. S.; and Eglén, S. J. 2014. Detecting pairwise correlations in spike trains: an objective comparison of methods and application to the study of retinal waves. *Journal of Neuroscience*, 34(43): 14288–14303.
- De Valois, R. L.; Albrecht, D. G.; and Thorell, L. G. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5): 545–559.
- Deb, M.; Deb, M.; and Murty, N. 2025. TopoNets: High performing vision and language models with brain-like topography. *arXiv preprint arXiv:2501.16396*.
- Dehghani, A.; Qian, X.; Farahani, A.; and Bashivan, P. 2024. Credit-based self organizing maps: training deep topographic networks with minimal performance degradation. In *The Thirteenth International Conference on Learning Representations*.
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.
- Ding, J.; Yu, Z.; Huang, T.; and Liu, J. K. 2023. Spike timing reshapes robustness against attacks in spiking neural networks. *arXiv preprint arXiv:2306.05654*.
- Dobs, K.; Martinez, J.; Kell, A. J.; and Kanwisher, N. 2022. Brain-like functional specialization emerges spontaneously in deep neural networks. *Science advances*, 8(11): eabl8913.
- Downing, P. E.; Jiang, Y.; Shuman, M.; and Kanwisher, N. 2001. A cortical area selective for visual processing of the human body. *Science*, 293(5539): 2470–2473.
- Fisher, R. A. 1925. Theory of statistical estimation. In *Mathematical proceedings of the Cambridge philosophical society*, volume 22, 700–725. Cambridge University Press.
- Gerstner, W.; and Kistler, W. M. 2002. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press.
- Grill-Spector, K.; and Weiner, K. S. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8): 536–548.
- Hu, Y.; Tang, H.; and Pan, G. 2021. Spiking deep residual networks. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 5200–5205.
- Hu, Y.; Zheng, Q.; Jiang, X.; and Pan, G. 2023. Fast-snn: Fast spiking neural network by converting quantized ann. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12): 14546–14562.
- Huang, L.; Ma, Z.; Yu, L.; Zhou, H.; and Tian, Y. 2023. Deep spiking neural networks with high representation similarity model visual pathways of macaque and mouse. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 31–39.
- Huang, L.; Ma, Z.; Yu, L.; Zhou, H.; and Tian, Y. 2024. Long-Range Feedback Spiking Network Captures Dynamic and Static Representations of the Visual Cortex under Movie Stimuli. *Advances in Neural Information Processing Systems*, 37: 11432–11455.
- Hubel, D. H.; and Wiesel, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, 160(1): 106.
- Jacobs, R. A.; and Jordan, M. I. 1992. Computational consequences of a bias toward short connections. *Journal of cognitive neuroscience*, 4(4): 323–336.
- Karbasforoushan, H.; Tian, R.; and Baker, J. 2022. There is a topographic organization in human cortico-pontine connectivity. *Brain Communications*, 4(2): fcac047.

- Kasabov, N. K. 2014. NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural networks*, 52: 62–76.
- Kim, Y.; Li, Y.; Park, H.; Venkatesha, Y.; Hambitzer, A.; and Panda, P. 2023. Exploring temporal information dynamics in spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8308–8316.
- Kohn, A.; and Smith, M. A. 2005. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14): 3661–3673.
- Konkle, T.; and Oliva, A. 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, 74(6): 1114–1124.
- Koulakov, A. A.; and Chklovskii, D. B. 2001. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron*, 29(2): 519–527.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.
- Kubilius, J.; Schrimpf, M.; Kar, K.; Rajalingham, R.; Hong, H.; Majaj, N.; Issa, E.; Bashivan, P.; Prescott-Roy, J.; Schmidt, K.; et al. 2019. Brain-like object recognition with high-performing shallow recurrent ANNs. *Advances in neural information processing systems*, 32.
- LeCun, Y.; Bengio, Y.; et al. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10): 1995.
- Lestienne, R. 2001. Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in neurobiology*, 65(6): 545–591.
- Livingstone, M. S.; and Hubel, D. H. 1984. Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4(1): 309–356.
- Long, B.; Yu, C.-P.; and Konkle, T. 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, 115(38): E9015–E9024.
- Maass, W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9): 1659–1671.
- Margalit, E.; Lee, H.; Finzi, D.; DiCarlo, J. J.; Grill-Spector, K.; and Yamins, D. L. 2024. A unifying framework for functional organization in early and higher ventral visual cortex. *Neuron*, 112(14): 2435–2451.
- Mauk, M. D.; and Buonomano, D. V. 2004. The neural basis of temporal processing. *Annu. Rev. Neurosci.*, 27(1): 307–340.
- Orlov, T.; Makin, T. R.; and Zohary, E. 2010. Topographic representation of the human body in the occipitotemporal cortex. *Neuron*, 68(3): 586–600.
- Patel, K.; Hunsberger, E.; Batir, S.; and Eliasmith, C. 2021. A spiking neural network for image segmentation. *arXiv preprint arXiv:2106.08921*.
- Pei, J.; Deng, L.; Song, S.; Zhao, M.; Zhang, Y.; Wu, S.; Wang, G.; Zou, Z.; Wu, Z.; He, W.; et al. 2019. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature*, 572(7767): 106–111.
- Qian, X.; Dehghani, A. O.; Farahani, A. B.; and Bashivan, P. 2024. Local lateral connectivity is sufficient for replicating cortex-like topographical organization in deep neural networks. *bioRxiv*, 2024–08.
- Ringach, D. L.; Shapley, R. M.; and Hawken, M. J. 2002. Orientation selectivity in macaque V1: diversity and laminar dependence. *Journal of neuroscience*, 22(13): 5639–5651.
- Rolls, E. T. 2000. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, 27(2): 205–218.
- Schrimpf, M.; Kubilius, J.; Hong, H.; Majaj, N. J.; Rajalingham, R.; Issa, E. B.; Kar, K.; Bashivan, P.; Prescott-Roy, J.; Geiger, F.; et al. 2018. Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, 407007.
- Sharma, D.; Ng, K. K.; Birznieks, I.; and Vickery, R. M. 2022. Perceived tactile intensity at a fixed primary afferent spike rate varies with the temporal pattern of spikes. *Journal of Neurophysiology*, 128(4): 1074–1084.
- Stigliani, A.; Weiner, K. S.; and Grill-Spector, K. 2015. Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience*, 35(36): 12412–12424.
- Tavanaei, A.; Ghodrati, M.; Kheradpisheh, S. R.; Masquelet, T.; and Maida, A. 2019. Deep learning in spiking neural networks. *Neural networks*, 111: 47–63.
- Tsao, D. Y.; Freiwald, W. A.; Tootell, R. B.; and Livingstone, M. S. 2006. A cortical region consisting entirely of face-selective cells. *Science*, 311(5761): 670–674.
- Von der Malsburg, C. 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2): 85–100.
- Willshaw, D. J.; and Von Der Malsburg, C. 1976. How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 194(1117): 431–445.
- Wu, Y.; Deng, L.; Li, G.; Zhu, J.; and Shi, L. 2018. Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in neuroscience*, 12: 331.
- Yamins, D. L.; and DiCarlo, J. J. 2016. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3): 356–365.
- Zeki, S. 1983. Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, 9(4): 741–765.
- Zhong, H.; Wang, H.; Huang, M.; Dai, W. P.; Huang, Y.; An, M.; Roe, A. W.; and Yu, Y. 2024. Emergence of Orientation Pinwheels in a Self-Evolving Spiking Neural Network: Enhancing Visual Coding Efficiency and Reliability. *bioRxiv*, 2024–03.
- Zhou, Z.; Zhu, Y.; He, C.; Wang, Y.; Yan, S.; Tian, Y.; and Yuan, L. 2022. Spikformer: When spiking neural network meets transformer. *arXiv preprint arXiv:2209.15425*.