DYNAMICAL MEMORY RETRIEVAL IN ATTRACTOR NETWORKS AS A DENOISING DIFFUSION PROCESS

Anonymous authors
Paper under double-blind review

ABSTRACT

Successful memory retrieval requires the network state to fall into the "attractor basin" of the target memory, which presents a challenge in setting the basin's size: a basin too broad leads to imprecise retrieval, while a basin too narrow fails to capture the network state. Here, we propose that the neural system can solve this problem by dynamically adjusting the basin's size, starting wide and gradually narrowing. We use continuous attractor neural networks (CANNs) as the memory model, where the width of the neural activity bump shrinks over time, thereby narrowing the attractor basin. Intriguingly, we find that this dynamical memory retrieval process is equivalent to the reverse dynamics of denoising diffusion probabilistic models (DDPMs), where the decrease in noise variance effectively corresponds to the contraction of the basin's size in CANNs. We carry out experiments demonstrating the advantages of this dynamical retrieval process; and for hierarchical memories, it gives rise to information retrieval from coarse to fine. We hope this study provides insight into dynamical memory retrieval in the neural system and its link to diffusion generative models in AI.

1 INTRODUCTION

Memory retrieval in the brain is a dynamical process. In classical attractor network models, recurrent interactions organize neural population activity into stable states to store memories. When a cue is incomplete or corrupted, the network can correct it as its state evolves into the corresponding attractor (Hopfield, 1982). The collection of network states that converge to a given memory forms the attractor basin of this memory, and it determines how much deviation from the stored representation the retrieval can tolerate (Storkey & Valabregue, 1999). Previous theoretical research on associative memory has focused on analyzing the memory capacity of attractor networks and how they relate to brain functions, but has often overlooked a dilemma regarding the size of the attractor basin. Simply stated, a basin too broad leads to imprecise memory retrieval, as incorrect cues may wrongly drop into the memory state; a basin too narrow leads to inefficient associative memory, as it misses many informative cues. How to properly choose the size of the attractor basin is critical for the brain to leverage attractor networks in order to recall memories.

In this study, we propose that the neural system can solve this dilemma by adjusting the basin's size dynamically, starting wide and gradually narrowing. A broad basin initially allows cue states over a large range to be captured, while its gradual narrowing enables retrieval to progress from coarse representations toward finer memories. To validate this idea, we consider continuous attractor neural networks (CANNs) as the memory model. CANNs, whose recurrent connectivity supports a continuous family of localized activity bumps (attractors), have been widely regarded as a canonical model of neural information representation (Amari, 1977; Zhang, 1996; Samsonovich & McNaughton, 1997; Wu et al., 2016). In particular, in the middle entorhinal cortex (MEC), the CANN formed by grid cells has been regarded as providing a memory scaffold which binds sensory inputs from hippocampal place cells (HPC) to hold episodic memories. In response to external inputs associated with stored memories, the bump moves along the attractor manifold toward a corresponding memory state, implementing the memory retrieval process (Fig. 1A). The bump width, which determines the spatial range of interaction from nearby network states, effectively determines the size of the attractor basin. We find that if the bump width in a CANN progressively shrinks over time, the attractor basin also shrinks accordingly, and the neural system can capture noisy cues

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

over a broad range initially and distinguish finer memories subsequently with increasing precision, achieving efficient dynamical memory retrieval.

In a separate line of AI research, denoising diffusion probabilistic models (DDPMs) have emerged as a powerful class of generative models. DDPMs first define a forward process that gradually perturbs data with Gaussian noise, and then learn a reverse process that progressively removes noise to generate samples from the data distribution (Sohl-Dickstein et al., 2015; Ho et al., 2020) (Fig. 1B). In the continuous-time formulation, this generative process can be associated with a deterministic probability-flow ordinary differential equation (ODE), whose velocity field is determined by the score of the probability distribution (Song et al., 2021).

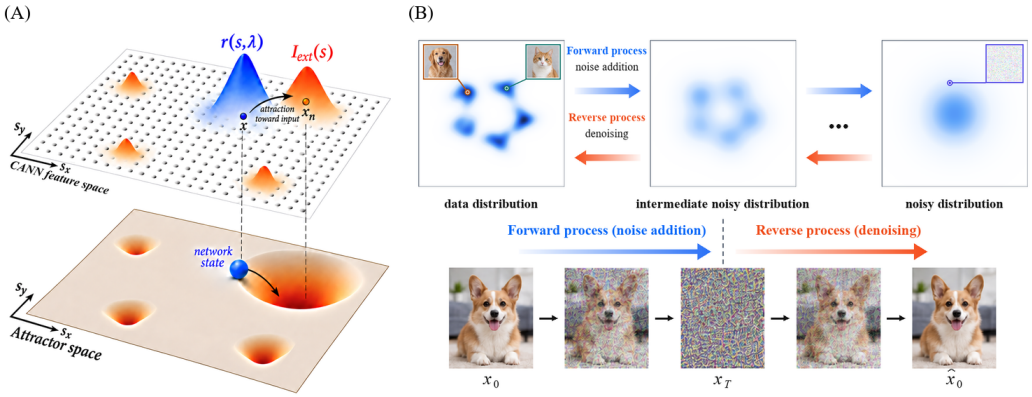

Figure 1: (A) Illustration of memory retrieval in the CANN feature space and the corresponding attractor space. In the CANN feature space, neurons (shown as gray dots) are arranged at different positions according to their preferred features. The network state is represented by the center of an activity bump (in blue), whose motion is jointly driven by multiple external inputs (in orange). In the attractor space, the same dynamics can be equivalently depicted as the network state evolving over an attractor landscape. Convergence to a specific attractor basin represents successful retrieval of the corresponding memory. (B) Illustration of the overall DDPM process. The upper row shows the evolution of the data distribution, where the forward process progressively adds noise to the dataset, transforming it into a noisy distribution, while the reverse process progressively removes the noise and recovers the original data distribution. The lower row shows the corresponding evolution of an individual image sample from this distribution throughout the forward and reverse processes.

Intriguingly, we find that the dynamical memory retrieval process in CANNs is effectively equivalent to the reverse dynamics of DDPMs, where the decrease of noise variance in DDPMs corresponds to the contraction of the basin's size in CANNs. We carry out simulation experiments confirming this equivalence and demonstrating the advantages of dynamical memory retrieval in CANNs. Furthermore, for hierarchical memories, we show that dynamical memory retrieval achieves coarse-to-fine information recall. We hope this study provides insight into dynamical memory retrieval in the brain and its association with diffusion generative models in AI.

Related works. Previous research on memory retrieval in neuroscience has usually considered attractor networks with a fixed attractor landscape during the retrieval process (Hopfield, 1982; Storkey & Valabregue, 1999; Khona & Fiete, 2022). In the classic Hopfield model, for example, the recurrent connections remain fixed during retrieval, and the corresponding attractor basins therefore do not change over time. Recent studies in AI have connected diffusion models with associative memories. Diffusion denoising has been described in the framework of energy-based associative memory (Hoover et al., 2023), and for discrete stored patterns, its energy function has been shown to be asymptotically identical to that of a continuous modern Hopfield network (Ramsauer et al., 2021; Ambrogioni, 2024). These studies have established a link between diffusion models and associative memory, but they have not identified the progressive change in noise variance as a retrieval mechanism that dynamically regulates attractor basins and retrieval resolution.

2 CANN AND DDPM PROBABILITY-FLOW ODE

2.1 CONTINUOUS ATTRACTOR NEURAL NETWORKS AND MEMORY RETRIEVAL

A CANN represents a continuous variable through a localized population activity bump in a feature space (Wu et al., 2016). Each neuron is indexed by its preferred feature $\mathbf{s} \in \mathbb{R}^d$ to which it activates the most; for example, $\mathbf{s}$ may denote a preferred head direction or spatial location. The center of the activity bump $\mathbf{x}$ specifies the network state, and movement of the bump corresponds to a continuous change of the represented state. Denote $u(\mathbf{s}, \lambda)$ as the synaptic input at time λ to neurons indexed at $\mathbf{s}$ in the feature space. The dynamics of a CANN is written as (Wu et al., 2002):

$$\tau_u \frac{\partial u(\mathbf{s}, \lambda)}{\partial \lambda} = -u(\mathbf{s}, \lambda) + \rho \int J(\mathbf{s}, \mathbf{s}') r(\mathbf{s}', \lambda)\, d\mathbf{s}' + I_{ext}(\mathbf{s}, \lambda), \tag{1}$$

where τ_u is the synaptic time constant, ρ the neuronal density, and $r(\mathbf{s}, \lambda)$ the neuronal firing rate. On the right-hand side, the first term describes the decay of synaptic input, the second term the neuronal recurrent interactions, and the last term the external input. We set the recurrent connectivity strength to be $J(\mathbf{s}, \mathbf{s}') = J_0 \exp\left[-\|\mathbf{s} - \mathbf{s}'\|^2/(2a^2)\right]$, with J_0 determining the recurrent interaction strength and a controlling the spatial range of interaction. $J(\mathbf{s}, \mathbf{s}')$ is translationally invariant in the feature space, enabling the network to hold a continuous family of attractors. The firing rate is determined from the synaptic input through the equation,

$$r(\mathbf{s}, \lambda) = \frac{u^2(\mathbf{s}, \lambda)}{1 + k\rho \int u^2(\mathbf{s}', \lambda)\, d\mathbf{s}'}, \tag{2}$$

where k controls the strength of global inhibition. The normalization limits the overall population activity while allowing a localized group of neurons to remain strongly active.

Previous studies have shown that, when the external input is sufficiently weak, the network admits an analytically solvable continuous family of Gaussian stationary bumps for both u and r (Wu et al., 2008; Fung et al., 2010):

$$u^*(\mathbf{s} \mid \mathbf{x}) = u_0 \exp\left[-\frac{\|\mathbf{s} - \mathbf{x}\|^2}{4a^2}\right], \qquad r^*(\mathbf{s} \mid \mathbf{x}) = r_0 \exp\left[-\frac{\|\mathbf{s} - \mathbf{x}\|^2}{2a^2}\right]. \tag{3}$$

It shows that the interaction scale a determines the width of the stationary population bump.

We consider that the external input takes the form of the summation over a number of memory states centered at $\mathbf{x}_1, \ldots, \mathbf{x}_N$ respectively,

$$I_{ext}(s, \lambda) = \sum_{n=1}^{N} \alpha_n \exp\left[-\frac{\|\mathbf{s} - \mathbf{x}_n\|^2}{4b_n^2}\right], \tag{4}$$

with each having its own strength α_n and spatial width b_n, allowing different represented states to exert different strengths and ranges of influence on the network. In the context of memory retrieval, this external input drives the activity bump toward a stable memory state and can be regarded as a retrieval signal associated with memories. In the MEC-HPC system, it corresponds to the feedforward projection from HPC to MEC. The memory retrieval process can therefore be characterized by the dynamics of the bump center.

2.2 DDPM AND ITS REVERSE DYNAMICS

Denoising diffusion probabilistic models (DDPMs) are generative models that learn a diffusion process to generate samples from a data distribution (Sohl-Dickstein et al., 2015; Ho et al., 2020). They consist of a forward diffusion process that gradually adds Gaussian noise to the data and a learned reverse process that progressively removes this noise to recover the data distribution.

During training, the model learns how a noisy sample should be moved toward the data distribution at different noise levels. This information can be described by the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, where $p_t(\mathbf{x})$ denotes the probability density of the noisy data at time t. The score gives the direction in which the probability density increases most rapidly. The reverse dynamics can therefore use the learned score to progressively move noisy states toward regions of higher probability under the data distribution.

We consider a set of training samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ drawn from an underlying data distribution $p_{\text{data}}(\mathbf{x})$. In practice, we assume an extra Gaussian sampling error σ_{data}, so that each sample $\mathbf{x}_n$ is associated with a Gaussian distribution $\mathcal{N}\left(\mathbf{x}_n, \sigma_{\text{data}}^2 \mathbf{I}\right)$. The forward diffusion process then progressively adds Gaussian noise to this distribution with an accumulated variance $\bar{\sigma}_t^2 = \int_0^t \sigma_s^2 ds$. Consequently, the distribution associated with the n-th sample at time t becomes

$$\mathbf{x}_t \mid \mathbf{x}_n \sim \mathcal{N}\left(\mathbf{x}_n, w_t^2 \mathbf{I}\right), \qquad w_t^2 = \bar{\sigma}_t^2 + \sigma_{\text{data}}^2. \tag{5}$$

The resulting noisy data distribution can therefore be written as the Gaussian mixture

$$p_t(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{(2\pi w_t^2)^{d/2}} \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{2w_t^2}\right]. \tag{6}$$

Song et al. (Song et al., 2021) showed that the evolution of a noisy state during the denoising process can be described by a deterministic probability-flow ODE (hereinafter referred to as DDPM ODE). Defining the reverse time $\lambda = -t$, the DDPM ODE can be written as

$$\frac{d\mathbf{x}}{d\lambda} = \frac{\sigma_t^2}{2} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) = \frac{\sigma_t^2}{2w_t^2} \frac{\sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{x}) \exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2w_t^2}\right]}{\sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2w_t^2}\right]}. \tag{7}$$

During the reverse process, w_t progressively decreases as the noise is removed, while the state $\mathbf{x}$ follows the score field toward the learned data distribution.

2.3 Correspondence between CANNs and DDPMs

We now establish a mathematical correspondence between the memory retrieval dynamics of CANNs and the reverse dynamics of DDPMs. For simplicity, we consider that all input components share the same strength α and width b, so that the CANN dynamics can be written as

$$\tau_u \frac{\partial u(\mathbf{s}, \lambda)}{\partial \lambda} = -u(\mathbf{s}, \lambda) + \rho \int J(\mathbf{s}, \mathbf{s}') r(\mathbf{s}', \lambda) d\mathbf{s}' + \alpha \sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{s} - \mathbf{x}_n\|^2}{4b^2}\right]. \tag{8}$$

Previous studies have shown that in a CANN, the dynamics of the network state can be well described by several dominating motion modes, and the dynamics of each mode can be obtained by projecting the network dynamics onto this mode (Wu et al., 2008; Fung et al., 2010). By projecting Eq. 8 onto the position mode $\boldsymbol{\phi}(\mathbf{s}|\mathbf{x}) = (\mathbf{s} - \mathbf{x}) \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right]$, we can derive the dynamics of the bump center (see Appendix A for details):

$$\frac{d\mathbf{x}}{d\lambda} = \frac{2\alpha a^2}{\tau_u u_0 (a^2 + b^2)} \left(\frac{2b^2}{a^2 + b^2}\right)^{d/2} \sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{x}) \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{4(a^2 + b^2)}\right]. \tag{9}$$

Comparing this expression with DDPM ODE in Eq. 7, we find that the two vector fields become identical if

$$2(a^2 + b^2) = w_t^2 = \bar{\sigma}_t^2 + \sigma_{\text{data}}^2, \tag{10}$$

and

$$\alpha(\mathbf{x}, t) = \frac{\tau_u u_0 \sigma_t^2}{8a^2} \left(\frac{w_t}{2b}\right)^d \frac{1}{\sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2w_t^2}\right]}. \tag{11}$$

Under these correspondences, a CANN can exactly reproduce DDPM ODE and retrieve a specific memory among multiple stored patterns with an error of σ_{data}.

In real neural systems, memory representations and external cues possess static physical meanings in feature space. The characteristic width of a fixed stimulus or memory pattern cannot vary over time. Therefore, we constrain b to be a time-invariant constant rather than a time-dependent quantity. We can write this as

$$b^2 = \beta \sigma_{\text{data}}^2, \quad a^2 = (\frac{1}{2} - \beta)\sigma_{\text{data}}^2 + \frac{\bar{\sigma}_t^2}{2}, \quad 0 < \beta < \frac{1}{2}, \tag{12}$$

indicating that for a CANN which implements a DDPM-like retrieval process, its recurrent interaction scale a and its bump width should progressively decrease.

3 Width Contraction Improves Retrieval Resolution

Section 2.3 shows that the decreasing noise scale w_t in DDPM corresponds directly to the contraction of the bump width in CANN. This correspondence allows us to characterize the effect of width contraction in CANN through the evolution of the DDPM reverse dynamics during denoising.

Hoover et al. (Hoover et al., 2023) interpreted diffusion denoising within an energy-based associative memory framework, in which the reverse dynamics can be described as gradient descent on a time-dependent energy landscape. For the Gaussian-mixture distribution considered here, the energy function can be written as

$$E(\mathbf{x}, t) = -\log \sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{2w_t^2}\right] + C(t), \qquad \frac{d\mathbf{x}}{d\lambda} = -\frac{\sigma_t^2}{2}\nabla_{\mathbf{x}} E(\mathbf{x}, t). \tag{13}$$

At each reverse time λ, points satisfying $\nabla_{\mathbf{x}} E = 0$ are the instantaneous equilibria of DDPM, and the Hessian $\nabla_{\mathbf{x}}^2 E$ evaluated at an equilibrium determines its local stability. Under the correspondence established in Section 2.3, these equilibria correspond to the equilibrium bump centers of the matched CANN at the corresponding bump width, with the same local stability. The evolution of the equilibrium points and their stability as w_t decreases directly describes how bump width contraction reorganizes the attractor structure of the CANN.

Analysis of the equilibrium condition and Hessian (see Appendix B for details) shows that broad bumps favor collective representations while narrow bumps prefer precise memories. As a result, when w_t is large relative to the separation between memories, the equilibrium lies close to their mean. This tendency becomes exact in the limit $w_t \to \infty$, where the landscape will have only one stable equilibrium at $\mathbf{x}^* \to \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n$, which is the average of all memories. However, as the bump width decreases, this coarse collective representation progressively loses stability, marked by the emergence of negative curvature in the Hessian, while new stable equilibria form around finer memory representations. Figure 2 illustrates this transition, where the central collective equilibrium gradually loses stability and finer stable equilibria emerge as the width decreases.

These results reveal that the bump width acts as a dynamical measure of the network's resolution. A broad bump stabilizes coarse attractors jointly determined by multiple memories, allowing the population state to first settle into a collective representation. As the bump contracts, the coarse attractors lose stability and finer attractors progressively emerge, pushing the network to scales of better precision.

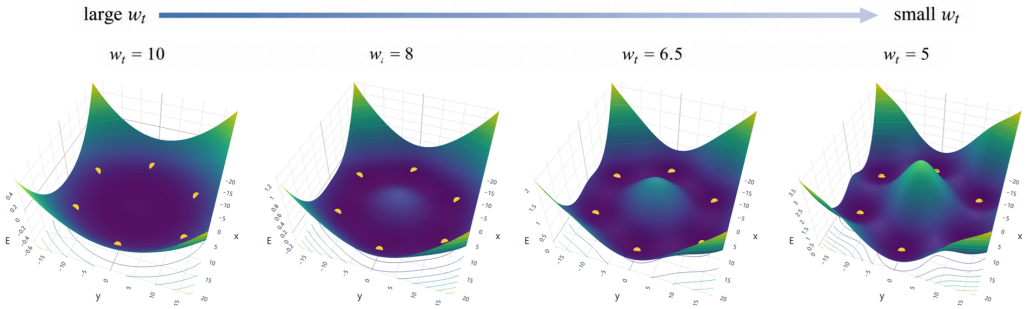

Figure 2: Evolution of the energy landscape during width contraction. We consider a hexagonal Gaussian memory configuration, where the yellow points indicate memory centers and the distance between two adjacent centers is set to 15. As the width decreases from $w_t = 10$ to 8, 6.5, and 5, the stable equilibrium at the center gradually loses stability, while new stable equilibria emerge around each finer memory.

However, when two memories are sufficiently close, bump contraction can no longer resolve them into distinct attractors, meaning that the network has a maximum resolution. To quantify this limit, we consider two identical Gaussian memories positioned along the x-axis and separated by a dis-

tance Δ, with their midpoint taken as the origin. The energy function here is formulated as:

$$E(\mathbf{x}, t) = \frac{x^2}{2w_t^2} - \log\cosh\left(\frac{\Delta x}{2w_t^2}\right) + C(t). \tag{14}$$

Intriguingly, differentiating this energy function gives a critical distance $\Delta_c = 2\sigma_{data}$ such that both the coarse equilibrium located between the two memories remains stable throughout retrieval, and no separate stable equilibria emerge around the two individual memories if $\Delta < \Delta_c$. Corresponding this to CANN, $\Delta_c = 2w_t|_{t=0} = 2\sqrt{2(a_t^2 + b^2)}|_{t=0}$ would be the maximum resolution of the network. The input width b reflects that both external stimuli and memory representations carry information within a finite range, instead of being ideal delta functions, whereas the bump width a reflects the resolution dynamically selected by the network during retrieval. It shows that the finest resolution attainable by a CANN is jointly constrained by the precision of the stored representation and the narrowest population bump that the network can sustain.

4 Experiments

4.1 Numerical Validation of the CANN-DDPM Correspondence

We first validate the CANN-DDPM correspondence derived in Section 2.3 using computer simulations. Both systems are implemented on a two-dimensional torus $\mathbb{S}^1 \times \mathbb{S}^1$; their dynamical properties are in good agreement with the theoretical results derived on $\mathbb{R}^2$, and this geometry eliminates boundary effects. We construct a mixed memory configuration on the torus and initialize DDPM ODE and the CANN from identical positions. The parameters in the CANN are determined according to the theoretical correspondences in Section 2.3, matched to the DDPM ODE. Simulation results show that across different initial positions, the trajectories of the bump center generated by the CANN closely resemble those of DDPM ODE and converge to the same memory states (Fig. 3).

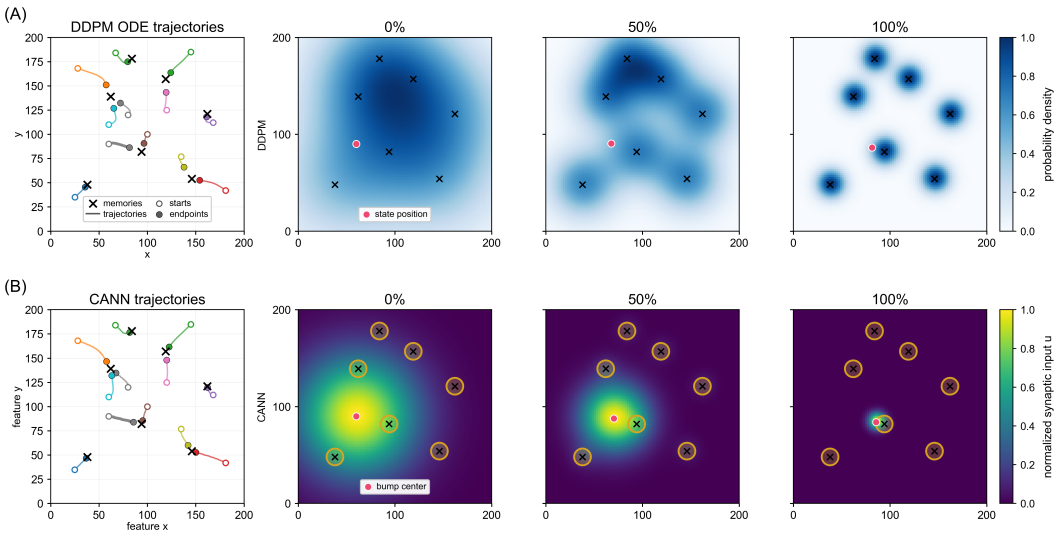

Figure 3: Simulations of the correspondence between CANN and DDPM ODE. (A) The left panel shows retrieval trajectories generated by DDPM ODE, starting at various locations. The right side presents snapshots of one of the DDPM ODE trajectories on the left (specifically, the gray one in bold), illustrating the probability density and the state position at 0%, 50%, and 100% of the reverse process. (B) The left shows retrieval trajectories generated by the corresponding CANN from the same initial positions as in (A). The snapshots on the right record the CANN bump center as well as the synaptic input field $u(\mathbf{s}, \lambda)$ along the trajectory initialized at the same position as the gray trajectory in (A), at 0%, 50%, and 100% of the retrieval process. It shows that as the bump contracts, the bump center moves toward a specific input center from the distance, while reproducing the corresponding DDPM trajectory. Yellow circles indicate regions within one standard deviation of each Gaussian input center.

4.2 Computational Advantages of the Dynamic Bump Width

A central implication of the theory in Section 3 is that bump width controls two complementary properties of retrieval: a broad bump integrates information over a large region of feature space and therefore provides a large capture range, whereas a narrow bump can resolve nearby memories into distinct attractors. Previous CANN models adopted fixed bump widths and therefore could not simultaneously maximize both properties. We here directly compare a dynamically contracting CANN with two controls whose widths are fixed to be broad and narrow respectively during the whole process, and we show that the dynamic CANN provides greater computational flexibility by combining long-range capture and fine-scale discrimination within a single retrieval process.

We still carry out the simulations on a 2D torus, where six identical Gaussian inputs are densely distributed as a hexagon, representing similar memories. We add a small amount of noise to each model such that the bump would be able to escape the central equilibrium when it loses stability. As is shown in Fig. 4A, the dynamic-width CANN is initialized at the distance as a wide bump. The bump first travels far and drops into the coarse energy basin at the center of the hexagon. As the width subsequently contracts, it gradually converges to a finer memory representation. The second CANN with a fixed wide width is initialized at the same place as the dynamic one. Although it also succeeds in converging to the central basin from the distance, it cannot effectively resolve and retrieve the finer memories (Fig. 4B). With a permanently narrow bump, finer states can be retrieved precisely if the bump is nearby, whereas a distant bump would only receive little effective drive and fail to reach the memory cluster (Fig. 4C).

The results of the two fixed-width controls reproduce the dilemma regarding the size of the attractor basin we mentioned in Section 1: a broad basin supports long-range captures but limits the retrieval precision, whereas a narrow basin provides fine discrimination at the cost of capture range. Our simulation shows that a contracting bump can combine the advantages of the two together within a single retrieval process.

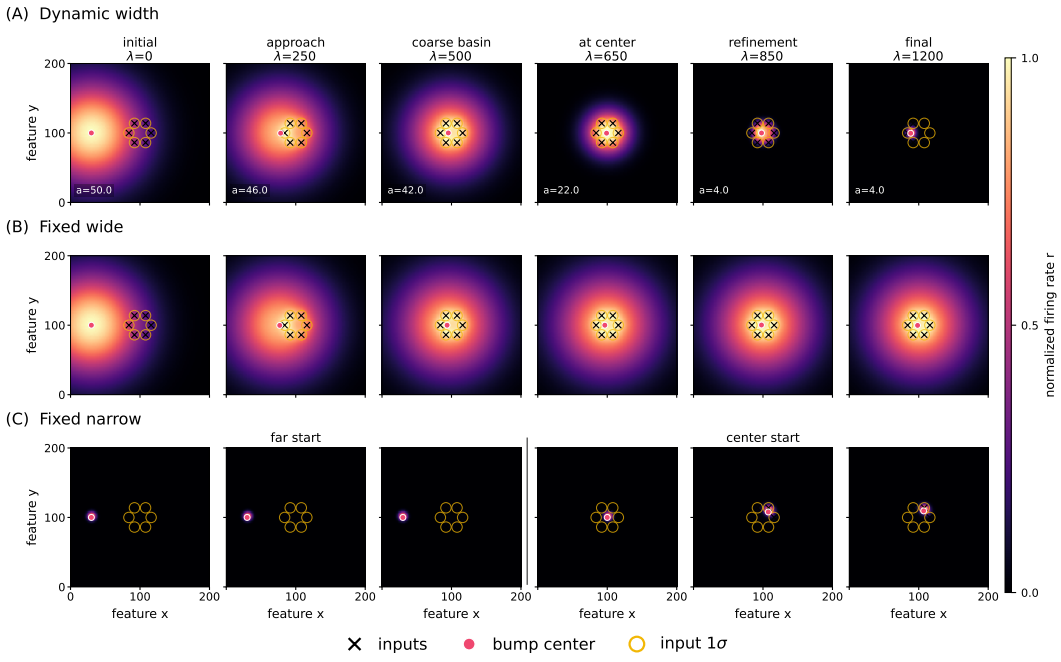

Figure 4: Comparison of the retrieval dynamics under dynamic and fixed bump widths. (A) Dynamic width enables the network to first approach the memory cluster with a broad bump, and then progressively refine its state toward an individual memory as the bump further contracts. (B) With a fixed broad width, the network can approach the memory cluster from a distant initial state, but it remains at the coarse collective representation subsequently. (C) A narrow bump initialized far away cannot effectively converge into the memory regime (shown in left half), but can distinguish individual memories when initialized nearby (shown in right half).

4.3 Coarse-to-Fine Retrieval on Hierarchical Natural Images

Further experiments show that the coarse-to-fine dynamics we derive above can operate on memories with an underlying hierarchical structure. We construct a three-level, $4 \times 3 \times 3$ visual memory space containing 4 major categories, 12 middle categories, and 36 leaf memories. The four major categories are dog, bird, fish, and snake, with each divided into three middle categories such as retriever dogs, parrots and sharks, and each middle category further divided into three leaf memories such as golden retriever, macaw and hammerhead shark. Each leaf is represented by an ImageNet exemplar and embedded into a four-dimensional continuous feature space. The coordinates are manually arranged such that leaves within the same middle category are closer than those belonging to different middle categories, while different major categories are separated at a larger scale. Importantly, only the 36 leaf states are stored as memory inputs to the CANN; middle and major states are defined by the "averages" of their constituent leaves and are not introduced as additional attractors. To visualize the continuous network state, we map the four-dimensional coordinates of the CANN into the latent space of a pretrained StyleGAN-XL generator. Details are provided in Appendix C.

We then initialize the CANN at random positions and gradually decrease the bump width throughout retrieval, with only the 36 leaf memories providing Gaussian inputs of identical strength and width. Simulation results show that at broad widths, the network is attracted toward a collective state associated with one of the four major categories. As the width decreases, the state moves toward one of the three middle categories within the selected major class; a further reduction in width finally resolves the individual leaf memories and drives the state toward a specific stored exemplar (Fig. 5). Overall, the whole process exhibits clear coarse-to-fine retrieval dynamics across these hierarchies.

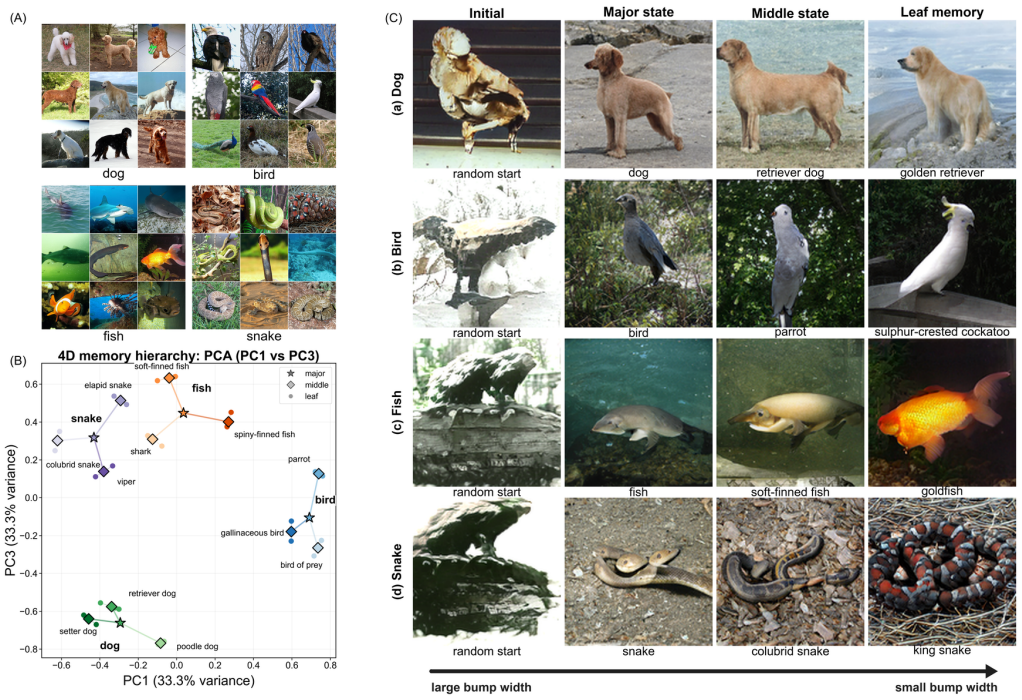

Figure 5: (A) Thirty-six ImageNet exemplars are used as leaf memories, organized into four major categories: dog, bird, fish, and snake. Within each 3×3 block, each row corresponds to one middle category and each image represents one leaf memory. (B) PCA visualization of the four-dimensional hierarchical feature space. Stars, diamonds, and circles denote major, middle and leaf memories. Major and middle memories are defined by the mean coordinates of their constituent leaves and are shown only as a reference; only the 36 leaf memories are stored as memory inputs to the CANN. (C) Representative coarse-to-fine retrieval processes for the four major categories. From left to right, the snapshots show the network state evolving from a random initialization toward increasingly specific representations. By contracting the bump width, the latent CANN first approaches the major level, then the middle level, and finally converges to a specific leaf memory.

5 Conclusion and Discussion

In this work, we establish a correspondence between the memory retrieval dynamics of CANNs and the reverse dynamics of DDPMs, through which the decreasing noise scale in diffusion corresponds to the contraction of the neural activity bump. Beyond this correspondence, our results identify bump width as a computational variable that controls the attractor basin's size and the resolution of retrieval. The energy landscape provides an intuitive picture of this process, where a broad bump integrates the influences of multiple nearby memories and supports coarse collective equilibria, whereas decreasing the width destabilizes these coarse states and allows finer attractors to emerge. Our resolution analysis further separates two sources of precision in this process. The input width reflects the finite information conveyed by the external cue, while the bump width reflects the resolution dynamically selected by the network. The finest distinction that the network can make is therefore jointly constrained by the precision of external cue and the width of the attractor activity.

An important consequence of this picture is that coarse representations do not have to be stored independently from fine memories, nor does the network require explicitly organized hierarchical memory layers for hierarchical structures to emerge. In our hierarchical image experiment, only the leaf exemplars are introduced as memory inputs, whereas the major and middle states arise from the collective influence of nearby leaf memories. A single set of stored memories can support both categorical and specific representations without changing the stored memories themselves; changing the resolution of retrieval is sufficient to reveal different levels of their underlying structure.

Limitations. For clarity, the present theoretical analysis assigns equal strength and width to different memory inputs, allowing the computational role of bump width to be isolated. In practice, this simplification is unlikely satisfied. Differences in input strength would change the relative influence of memories on the population state, whereas differences in input width would change the spatial range over which individual memories affect the dynamics. These factors can reshape the attraction landscape, alter the positions and stability of collective states, and bias which fine attractor is eventually retrieved. Extending the present framework to memory inputs of different strengths and widths will provide a more realistic account of retrieval. A remaining question concerns the biological implementation of the width modulation. In the present framework, bump width is treated as an effective computational variable, while the underlying circuit mechanism is left unspecified. In neural systems, such modulation may arise through changes in neuronal response properties, recurrent interactions, or inhibitory regulation. Identifying the circuit mechanisms that dynamically control population width will be important for connecting the present theory to biological neural networks.

Future directions. The monotonic contraction considered in this work is required by the correspondence with the DDPM ODE, whose noise scale progressively decreases during denoising. However, the bump width can be treated as an independently controllable variable and modulated non-monotonically according to computational demands in real CANNs. The network may increase or decrease its bump width flexibly according to its current task and thereby move between different retrieval resolutions. One possible function of such flexible width modulation is state switching. A network initially stabilized at a fine attractor with a narrow bump may first broaden toward a coarser collective state and then narrow again toward another fine attractor. In a hierarchical memory space, this process can also support associations between related memories, where the extent of width-broadening determines how far the association spans the representational hierarchy. More generally, flexible modulation of the bump width may support a broader range of computations through the dynamic control of retrieval resolution, many of which remain to be explored.

Another broader direction is extending the dynamical framework to other contexts of neural computation. Although we formulate the present theory in terms of memory retrieval, the underlying framework is not restricted to memory. The bump center can generally represent a "state" in a continuous feature space, while stable attractors correspond to candidate states that the system may retrieve. Similar dynamics may therefore arise in other neural computations that require a system to recover, select, or transition toward a stable state from incomplete or ambiguous information, such as perceptual inference, categorical recognition, internal-state estimation, or decision making. From this perspective, memory retrieval represents one instance of a broader class of state-retrieval dynamics that may be described within the same framework.

AI USE STATEMENT

In this work, we used generative AI tools to assist with drafting and polishing portions of the manuscript, including improving clarity, grammar, and organization. We also used generative AI tools to support literature retrieval and discovery of potentially relevant prior work, as well as code development and debugging. All AI-suggested references were independently checked against the original sources, and all AI-assisted mathematical arguments, code, experimental procedures, and results were reviewed and verified by the authors. We made all research decisions and take full responsibility for the final content of this work, including its text, claims, analyses, and results.

REPRODUCIBILITY STATEMENT

The assumptions and mathematical derivations underlying the CANN–DDPM correspondence are provided in Sections 2 and 3, with additional derivation details in Appendices A and B. The experimental settings and model configurations used to validate the correspondence, compare dynamic and fixed bump widths, and demonstrate coarse-to-fine retrieval are described in Section 4, Appendix C and the corresponding figure captions. These descriptions specify the theoretical assumptions, math equations, memory configurations, and experimental procedures used to obtain the reported results.

REFERENCES

Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87, 1977. doi: 10.1007/BF00337259.

Luca Ambrogioni. In search of dispersed memories: Generative diffusion models are associative memory networks. *Entropy*, 26(5):381, 2024. doi: 10.3390/e26050381.

C. C. Alan Fung, K. Y. Michael Wong, and Si Wu. A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Computation*, 22(3):752–792, 2010. doi: 10.1162/neco.2009.07-08-824.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Benjamin Hoover, Hendrik Strobelt, Dmitry Krotov, Judy Hoffman, Zsolt Kira, and Duen Horng Chau. Memory in plain sight: Surveying the uncanny resemblances of associative memories and diffusion models. *arXiv preprint arXiv:2309.16750*, 2023.

John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554.

Mikail Khona and Ila R. Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23:744–766, 2022. doi: 10.1038/s41583-022-00642-0.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David P. Kreil, Michael K. Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. In *International Conference on Learning Representations*, 2021.

Alexei Samsonovich and Bruce L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–5920, 1997. doi: 10.1523/JNEUROSCI.17-15-05900.1997.

Axel Sauer, Katja Schwarz, and Andreas Geiger. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 Conference Proceedings*, 2022. doi: 10.1145/3528233.3530738.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265. PMLR, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Amos J. Storkey and R. Valabregue. The basins of attraction of a new hopfield learning rule. *Neural Networks*, 12(6):869–876, 1999. doi: 10.1016/S0893-6080(99)00038-6.

Si Wu, Shun-ichi Amari, and Hiroyuki Nakahara. Population coding and decoding in a neural field: A computational study. *Neural Computation*, 14(5):999–1026, 2002. doi: 10.1162/089976602753633367.

Si Wu, Kosuke Hamaguchi, and Shun-ichi Amari. Dynamics and computation of continuous attractors. *Neural Computation*, 20(4):994–1025, 2008. doi: 10.1162/neco.2008.10-06-378.

Si Wu, K. Y. Michael Wong, C. C. Alan Fung, Yuanyuan Mi, and Wenhao Zhang. Continuous attractor neural networks: Candidate of a canonical model for neural information representation. *F1000Research*, 5:156, 2016. doi: 10.12688/f1000research.7387.1.

Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996. doi: 10.1523/JNEUROSCI.16-06-02112.1996.

A Position Mode Projection Method for CANN

We consider the CANN in Section 2.3, whose equations can be written as:

$$\tau_u \frac{\partial u(\mathbf{s},\lambda)}{\partial \lambda} = -u(\mathbf{s},\lambda) + \rho \int J(\mathbf{s},\mathbf{s}') r(\mathbf{s}',\lambda) d\mathbf{s}' + \alpha \sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{4b^2}\right], \tag{15}$$

$$r(\mathbf{s},\lambda) = \frac{u^2(\mathbf{s},\lambda)}{1+k\rho\int u^2(\mathbf{s}',\lambda)d\mathbf{s}'}. \tag{16}$$

Previous studies have shown that when the external input is sufficiently small,

$$u(\mathbf{s},\lambda) = u_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] \tag{17}$$

is a stationary solution for synaptic input $u(\mathbf{s},\lambda)$. Substitute this into Eq. 16, and since

$$\int u^2(\mathbf{s}',\lambda)d\mathbf{s}' = u_0^2 \int \exp\left[-\frac{\|\mathbf{s}'-\mathbf{x}\|^2}{2a^2}\right] d\mathbf{s}' = u_0^2 (2\pi a^2)^{d/2}, \tag{18}$$

we derive that the firing rate $r(\mathbf{s},\lambda)$ is also a Gaussian:

$$r(\mathbf{s},\lambda) = \frac{u_0^2}{1+k\rho u_0^2 (2\pi a^2)^{d/2}} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{2a^2}\right] \equiv r_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{2a^2}\right]. \tag{19}$$

Then we calculate the form of the recurrent connection term $\rho\int J(\mathbf{s},\mathbf{s}')r(\mathbf{s}',\lambda)d\mathbf{s}'$. Substituting $J(\mathbf{s},\mathbf{s}')$ and $r(\mathbf{s}',\lambda)$ into the recurrent term gives

$$\rho \int J(\mathbf{s},\mathbf{s}')r(\mathbf{s}',\lambda)\, d\mathbf{s}' = \rho J_0 r_0 \int \exp\left[-\frac{\|\mathbf{s}-\mathbf{s}'\|^2}{2a^2} - \frac{\|\mathbf{s}'-\mathbf{x}\|^2}{2a^2}\right] d\mathbf{s}'. \tag{20}$$

The quadratic terms in the exponent can be rewritten as

$$\frac{\|\mathbf{s}-\mathbf{s}'\|^2}{2a^2} + \frac{\|\mathbf{s}'-\mathbf{x}\|^2}{2a^2} = \frac{1}{a^2}\left\|\mathbf{s}' - \frac{\mathbf{s}+\mathbf{x}}{2}\right\|^2 + \frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}. \tag{21}$$

Therefore,

$$\rho \int J(\mathbf{s},\mathbf{s}')r(\mathbf{s}',\lambda)\, d\mathbf{s}' = \rho J_0 r_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] \int \exp\left[-\frac{1}{a^2}\left\|\mathbf{s}' - \frac{\mathbf{s}+\mathbf{x}}{2}\right\|^2\right] d\mathbf{s}'. \tag{22}$$

The Gaussian integral evaluates to

$$\int \exp\left[-\frac{1}{a^2}\left\|\mathbf{s}' - \frac{\mathbf{s}+\mathbf{x}}{2}\right\|^2\right] d\mathbf{s}' = (\pi a^2)^{d/2}. \tag{23}$$

Hence, the recurrent term becomes

$$\rho \int J(\mathbf{s},\mathbf{s}')r(\mathbf{s}',\lambda)\, d\mathbf{s}' = \rho J_0 (\pi a^2)^{d/2} r_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right]. \tag{24}$$

So under the stationary solution, the equation of CANN would be:

$$\tau_u \frac{\partial u(\mathbf{s},\lambda)}{\partial \lambda} = -u_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] + \rho J_0 (\pi a^2)^{d/2} r_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] + \alpha \sum_{n=1}^{N} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{4b^2}\right]. \tag{25}$$

We then project Eq. 25 onto the position mode $\boldsymbol{\phi}(\mathbf{s}|\mathbf{x}) = (\mathbf{s}-\mathbf{x}) \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right]$, which is equivalent to multiplying both sides by $\boldsymbol{\phi}(\mathbf{s}|\mathbf{x})$ and then integrating with respect to $\mathbf{s}$. Note that since both the leakage term and the recurrent connection term are symmetric with respect to $\mathbf{x}$, the integral of these two terms is 0. The projected equation is therefore

$$\tau_u \int (\mathbf{s}-\mathbf{x}) \frac{\partial u(\mathbf{s},\lambda)}{\partial \lambda} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] d\mathbf{s} = \alpha \sum_{n=1}^{N} \int (\mathbf{s}-\mathbf{x}) \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{4b^2} - \frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] d\mathbf{s}. \tag{26}$$

The derivative of u with respect to λ is

$$\frac{\partial u(\mathbf{s},\lambda)}{\partial \lambda} = \dot{u}_0 \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right] + \frac{u_0 (\mathbf{s}-\mathbf{x})^{\mathrm{T}} \dot{\mathbf{x}}}{2a^2} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2}\right], \tag{27}$$

so the left side is

$$\tau_u \int (\mathbf{s}-\mathbf{x}) \left[\dot{u}_0 + \frac{u_0 (\mathbf{s}-\mathbf{x})^{\mathrm{T}} \dot{\mathbf{x}}}{2a^2}\right] \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{2a^2}\right] d\mathbf{s}. \tag{28}$$

The term with $\dot{u}_0$ is an odd function, and

$$\int (\mathbf{s}-\mathbf{x})(\mathbf{s}-\mathbf{x})^{\mathrm{T}} \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{2a^2}\right] d\mathbf{s} = a^2 (2\pi a^2)^{d/2} \boldsymbol{I}. \tag{29}$$

Then finally, the left side equals to

$$\frac{\tau_u u_0}{2} (2\pi a^2)^{d/2} \dot{\mathbf{x}}. \tag{30}$$

For the n-th external input, its contribution to the position mode projection is

$$\int (\mathbf{s}-\mathbf{x}) \exp\left[-\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2} - \frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{4b^2}\right] d\mathbf{s}. \tag{31}$$

To evaluate this integral, we first combine the two Gaussian factors. Completing the square gives

$$\frac{\|\mathbf{s}-\mathbf{x}\|^2}{4a^2} + \frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{4b^2} = \frac{a^2+b^2}{4a^2 b^2} \left\|\mathbf{s} - \frac{b^2\mathbf{x} + a^2 \mathbf{x}_n}{a^2+b^2}\right\|^2 + \frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{4(a^2+b^2)}.$$

We define the center of the product Gaussian as

$$\boldsymbol{\mu}_n = \frac{b^2\mathbf{x} + a^2 \mathbf{x}_n}{a^2+b^2}. \tag{32}$$

Its displacement relative to the current bump center is therefore

$$\boldsymbol{\mu}_n - \mathbf{x} = \frac{a^2}{a^2+b^2}(\mathbf{x}_n - \mathbf{x}). \tag{33}$$

Then the projection integral can be written as

$$\exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{4(a^2+b^2)}\right] \int [(\mathbf{s}-\boldsymbol{\mu}_n) + (\boldsymbol{\mu}_n - \mathbf{x})] \exp\left[-\frac{a^2+b^2}{4a^2 b^2}\|\mathbf{s}-\boldsymbol{\mu}_n\|^2\right] d\mathbf{s}. \tag{34}$$

The first term vanishes by symmetry, and the remaining d-dimensional Gaussian integral is

$$\int \exp\left[-\frac{a^2+b^2}{4a^2b^2}\|\mathbf{s}-\boldsymbol{\mu}_n\|^2\right] d\mathbf{s} = \left(\frac{4\pi a^2 b^2}{a^2+b^2}\right)^{d/2}. \tag{35}$$

Hence the n-th external input contributes

$$\frac{a^2}{a^2+b^2}(\mathbf{x}_n-\mathbf{x})\left(\frac{4\pi a^2 b^2}{a^2+b^2}\right)^{d/2}\exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{4(a^2+b^2)}\right], \tag{36}$$

and Eq. 26 can be written as:

$$\frac{\tau_u u_0}{2}(2\pi a^2)^{d/2}\dot{\mathbf{x}} = \alpha\frac{a^2}{a^2+b^2}\left(\frac{4\pi a^2 b^2}{a^2+b^2}\right)^{d/2}\sum_{n=1}^{N}(\mathbf{x}_n-\mathbf{x})\exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{4(a^2+b^2)}\right]. \tag{37}$$

By dividing both sides by $\tau_u u_0(2\pi a^2)^{d/2}/2$, we derive the center dynamics of the bump center by this method of projection:

$$\frac{d\mathbf{x}}{d\lambda} = \frac{2\alpha a^2}{\tau_u u_0(a^2+b^2)}\left(\frac{2b^2}{a^2+b^2}\right)^{d/2}\sum_{n=1}^{N}(\mathbf{x}_n-\mathbf{x})\exp\left[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{4(a^2+b^2)}\right]. \tag{38}$$

B Analysis of Equilibria and Stability

The energy function of DDPM ODE is

$$E(\mathbf{x},t) = -\log\sum_{n=1}^{N}\exp[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2w_t^2}] + C(t). \tag{39}$$

An equilibrium point requires $\nabla_{\mathbf{x}}E = 0$. Specifically,

$$\nabla_{\mathbf{x}}E(\mathbf{x},t) = \frac{1}{w_t^2}\sum_{n=1}^{N}q_n(\mathbf{x})(\mathbf{x}-\mathbf{x}_n), \quad q_n(\mathbf{x}) = \frac{\exp[-\frac{\|\mathbf{x}-\mathbf{x}_n\|^2}{2w_t^2}]}{\sum_{m=1}^{N}\exp[-\frac{\|\mathbf{x}-\mathbf{x}_m\|^2}{2w_t^2}]}. \tag{40}$$

So any equilibrium point (denoted as $\mathbf{x}^*$) should satisfy

$$\mathbf{x}^* = \sum_{n=1}^{N}q_n(\mathbf{x}^*)\mathbf{x}_n. \tag{41}$$

And we can furthermore calculate the Hessian of the energy function at $\mathbf{x}^*$, which indicates the equilibrium point's stability:

$$\nabla_{\mathbf{x}}^2 E(\mathbf{x}^*,t) = \frac{1}{w_t^2}\boldsymbol{I} - \frac{1}{w_t^4}\sum_{n=1}^{N}q_n(\mathbf{x}^*)(\mathbf{x}_n-\mathbf{x}^*)(\mathbf{x}_n-\mathbf{x}^*)^{\mathrm{T}}. \tag{42}$$

When w_t is large compared with the spatial separation between memories (e.g. in the limit $w_t \to \infty$), $q_n(\mathbf{x})$ become nearly uniform. Hence, there is only one equilibrium point which is stable and approaches the mean of all memories.

$$\mathbf{x}^* \to \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n, \quad \nabla_{\mathbf{x}}^2 E(\mathbf{x}^*,t) \succ 0. \tag{43}$$

As the bump contracts, the corresponding w_t decreases and the stability of the coarse equilibrium point changes. For an arbitrary unit direction $\mathbf{v}$, the curvature of this direction at an equilibrium point is

$$\mathbf{v}^{\mathrm{T}}\nabla_{\mathbf{x}}^2 E(\mathbf{x}^*,t)\mathbf{v} = \frac{1}{w_t^4}\left\{w_t^2 - \sum_{n=1}^{N}q_n(\mathbf{x}^*)\left[\mathbf{v}^{\mathrm{T}}(\mathbf{x}_n-\mathbf{x}^*)\right]^2\right\}. \tag{44}$$

$\sum_{n=1}^{N}q_n(\mathbf{x}^*)\left[\mathbf{v}^{\mathrm{T}}(\mathbf{x}_n-\mathbf{x}^*)\right]^2$ measures the weighted spatial dispersion of the memories along $\mathbf{v}$. The coarse equilibrium point will lose its stability along that direction once this dispersion exceeds

w_t^2 (or equivalently, $\mathbf{v}^{\mathrm{T}}\nabla_{\mathbf{x}}^2 E(\mathbf{x}^*, t)\mathbf{v} < 0$) and the system's representation will depart. And then as w_t gets smaller and smaller, more directions will lose their stability progressively at the coarse equilibrium point.

Meanwhile, the decreasing width also induces new stable equilibria around memories of better precision. When w_t is small compared with the spatial separation between memories, in the vicinity of a certain memory $\mathbf{x}_i$, $q_i(\mathbf{x})$ approaches 1 and $q_j(\mathbf{x})$ $(j \neq i)$ approaches 0. Thus, Eq. 41 admits an equilibrium around $\mathbf{x}_i$. As for its stability, near $\mathbf{x}_i$, the weighted dispersion $\sum_{n=1}^{N} q_n(\mathbf{x}^*)\left[\mathbf{v}^{\mathrm{T}}(\mathbf{x}_n - \mathbf{x}^*)\right]^2$ decays rapidly toward 0, with a rate faster than w_t^2. Therefore, for sufficiently small w_t, the Hessian of the energy function would be positive definite, which means the new, finer equilibrium point around $\mathbf{x}_i$ would be stable. Taken together, as the bump width in CANN contracts, new stable equilibrium points emerge in the neighborhood of those finer-scale memory representations.

C Details of the Hierarchical Image Experiment

Image readout. The dynamics of the hierarchical image experiment in Section 4.3 is performed entirely in the four-dimensional CANN feature space, while the corresponding images are obtained through a separate continuous readout. Specifically, we use a readout function R_θ followed by a pretrained image generator G,

$$\mathbf{x} \in \mathbb{R}^4 \xrightarrow{R_\theta} \mathbf{w} \in \mathcal{W} \subset \mathbb{R}^{512} \xrightarrow{G} I, \tag{45}$$

where $\mathbf{x}$ denotes the CANN state, $\mathbf{w}$ is a latent code in the StyleGAN $\mathcal{W}$ space, and I is the decoded image. We use the pretrained StyleGAN-XL generator trained on ImageNet at 128×128 resolution (Sauer et al., 2022). The generator is kept fixed throughout the experiment, it only provides a continuous visualization of the state evolution in the four-dimensional feature space.

To associate images with the generator, we first apply the latent-optimization inversion procedure provided with StyleGAN-XL to each of the 36 ImageNet leaf exemplars. For each target image I_n, the inversion produces a single latent code $\mathbf{w}_n \in \mathbb{R}^{512}$ such that $G(\mathbf{w}_n)$ approximately reconstructs the original image. The generator itself is not fine-tuned during this procedure. These inverted latent codes provide the image representations of the 36 leaf memories used in the experiment.

Construction of the hierarchical feature space. The four-dimensional hierarchy is constructed independently of the StyleGAN latent space. We manually assign the 36 leaf memories to anchors $\mathbf{x}_{ijk} \in \mathbb{R}^4$, where $i = 1, \ldots, 4$ indexes the major category, $j = 1, \ldots, 3$ the middle category, and $k = 1, \ldots, 3$ the individual leaf. The coordinates are arranged such that leaves belonging to the same middle category are nearby, different middle categories within the same major category are separated at a larger scale, and different major categories are separated further (as shown in Fig. 6).

For each middle category, we define its center by the mean of its three leaf anchors,

$$\mathbf{x}_{ij}^{\mathrm{mid}} = \frac{1}{3}\sum_{k=1}^{3} \mathbf{x}_{ijk}, \tag{46}$$

and define the center of each major category by the mean of its nine constituent leaves,

$$\mathbf{x}_{i}^{\mathrm{maj}} = \frac{1}{9}\sum_{j=1}^{3}\sum_{k=1}^{3} \mathbf{x}_{ijk}. \tag{47}$$

The corresponding reference points in the StyleGAN latent space are defined analogously,

$$\mathbf{w}_{ij}^{\mathrm{mid}} = \frac{1}{3}\sum_{k=1}^{3} \mathbf{w}_{ijk}, \qquad \mathbf{w}_{i}^{\mathrm{maj}} = \frac{1}{9}\sum_{j=1}^{3}\sum_{k=1}^{3} \mathbf{w}_{ijk}. \tag{48}$$

These middle and major centers are used only to define the hierarchical geometry and the image readout. They are not introduced as additional memory inputs to the CANN; the network receives external inputs only from the 36 leaf anchors.

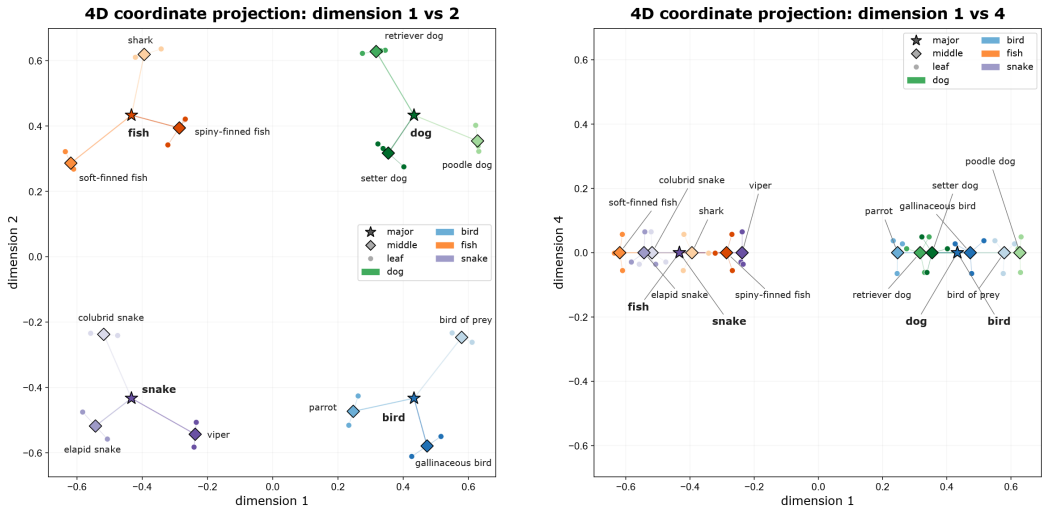

Figure 6: Two-dimensional projections of the four-dimensional hierarchical feature space. Stars, diamonds, and circles denote major, middle, and leaf states respectively. The left and right panels show the projections onto dimensions 1–2 and 1–4.

Continuous mapping from CANN states to StyleGAN latent codes. We parameterize $R_\theta : \mathbb{R}^4 \to \mathbb{R}^{512}$ by a multilayer perceptron (MLP). Since the four-dimensional CANN feature space and the 512-dimensional StyleGAN $\mathcal{W}$ latent space have different dimensionalities, a global isomorphic correspondence between the two spaces is mathematically not possible. We therefore fit the readout R_θ only to ensure accurate correspondence at the key states which are most relevant for visualization and approximate local flatness in their vicinity. Specifically, the 36 leaf anchors are mapped accurately to their inverted $\mathcal{W}$ codes, while the 12 middle-category centers and four major-category centers are mapped to the corresponding mean $\mathcal{W}$ codes of their constituent leaves. We additionally sample intermediate points along the major-to-middle and middle-to-leaf connections and constrain their outputs toward the corresponding linear interpolations in $\mathcal{W}$ space. A curvature regularization is applied along these paths to encourage the readout to vary smoothly and approximately affinely in hierarchy-relevant directions. These constraints are intended to provide a continuous and interpretable image readout around the hierarchical states traversed by the CANN.

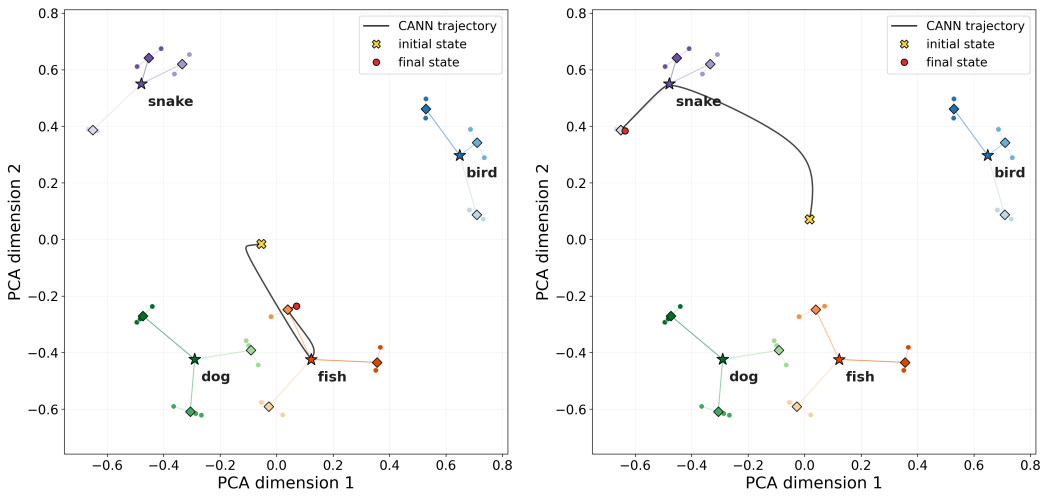

Figure 7: Representative CANN trajectories in the 4D feature space. Specifically, the two panels show the inner trajectories corresponding to two retrieval examples shown in Fig. 5C.

CANN retrieval trajectories. After fitting R_θ, both the readout and StyleGAN-XL generator are fixed. The CANN then evolves independently in the 4D feature space under the 36 Gaussian leaf inputs described in Section 4.3. During retrieval, the bump width is monotonically decreased from a broad initial state to a narrow final state. At any time λ, the bump center $\mathbf{x}(\lambda)$ can be decoded as

$$I(\lambda) = G(R_\theta(\mathbf{x}(\lambda))) \,, \tag{49}$$

allowing the continuous retrieval trajectory to be visualized directly in image space.

Figure 7 further shows representative trajectories in the four-dimensional feature space. Although only the leaf anchors act as memory inputs, trajectories initialized away from the stored memories first approach the neighborhood of a major-category center, subsequently move toward a corresponding middle-category center as the bump contracts, and finally converge to an individual leaf. The decoded images therefore provide a direct visualization of the coarse-to-fine evolution generated by the CANN dynamics.