Distilling multi-scale neural mechanisms from diverse unlabeled experimental data using deep domain-adaptive inference framework

Kaiwen Sheng[1,2,3,5], Shanghang Zhang[2], Maxime Beau[5], Peng Qu[4], Le Yang[4], Xiaofei Liu[2], Liuyuan He[2], Youhui Zhang[4], Lei Ma[2,3*], Kai Du[1*]

[1]Institute for Artificial Intelligence, Peking University, Beijing, China

[2]School of Computer Science, Peking University, Beijing, China

[3]Beijing Academy of Artificial Intelligence, Beijing, China

[4]Department of Computer Science and Technology, Tsinghua University, Beijing, China

[5]Wolfson Institute for Biomedical Research, University College London, London, United Kingdom

[*]Correspondence: lei.ma@pku.edu.cn (L.M.) and kai.du@pku.edu.cn (K.D.)

Abstract

Behavior and cognition emerge from the complex interplay of neural properties at different scales. However, inferring these multi-scale properties from diverse experimental data remains a classical challenge in computational and systems neuroscience. Advanced machine learning (ML) techniques, such as deep learning and Bayesian inference, have shown promise in addressing this issue. Nonetheless, the performance of ML models trained on synthetic data generated from computational models degrades dramatically on experimental data. To systematically tackle these challenges, we introduce the concept of "out-of-distribution (OOD)" to quantify distributional shift between synthetic and experimental datasets, and propose a deep domain-adaptive inference framework that aligns the distribution of synthetic data with experimental data by minimizing OOD errors. Our framework achieves state-of-the-art performance on a wide range of real experimental data when inferring neural properties at different scales. We demonstrate the efficacy of our framework in two scenarios: inferring detailed biophysical properties at the neuron and microcircuit scales, and inferring monosynaptic connections in hippocampal CA1 networks from *in vivo* multi-electrode extracellular recordings in free-running mice. Our approach represents a pioneering systematic solution to the OOD problem in neuroscience research and can potentially facilitate bottom-up modeling of large-scale network dynamics underlying brain function and dysfunction.

Teaser

Our deep domain-adaptive inference framework addresses the out-of-distribution (OOD) problem in inferring multi-scale neural properties from experimental data, enabling state-of-the-art performance and broad implications for neuroscience research.

Introduction

The brain's computation and function are determined by intricate interplay between neural properties at multiple scales, from individual neurons to circuits and networks. For example, at the neuronal level, the biophysical properties of a neuron (i.e. ion channels and membrane properties) serve as the basis for its own computation (1, 2), while at the system level, the connectivity pattern of a neural network is a critical factor in determining its dynamics (3, 4). Moreover, the strength of individual synaptic connections can further modulate the network function, even if the network connectivity is fixed (5). However, it is not always feasible to direct measure these multi-scale neural properties through experiments. Therefore, inferring the desired properties from available neural data becomes essential for a deeper understanding of the neural processes underlying cognition and behavior — a fundamental goal in computational and systems neuroscience (6-9).

Over recent years, many thrilling experimental techniques have been developed to explore neural functions from different perspectives, such as patch-seq (10), extracellular recordings (7, 9), calcium imaging (11) and optogenetics (12, 13), which provide rich data to infer the underlying neural properties at different scales. However, the diversity and large scale of these data also pose a huge challenge for the inference procedure. Recently, many advanced machine learning (ML) methods, such as deep learning (14, 15) or Bayesian inference (16), have been introduced to address this challenge, owing to their extraordinary abilities in dealing with complex high-dimensional data. A classic paradigm for the supervised ML methods is to leverage labeled synthetic data (i.e., with ground truths) randomly generated from computational models to infer neural properties underlying unlabeled experimental data. However, these ML models trained on synthetic data suffered severe performance degradation on experimental data, even if they achieved near-optimal performance on synthetic data. In the present study, we first identify this phenomenon as the out-of-distribution (OOD) problem, a central challenge in ML (17, 18), which is caused by the difference in statistical distribution between randomly generated synthetic data and experimental data, and then propose a systematic way to tackle the problem.

The OOD problem is crucial to the reliability of ML applications in real world but remains unresolved to date. For example, a self-driving car may lose control in a novel environment that is rarely represented in the training data distribution, such as a rare weather condition (19). Likewise, ML models trained on synthetic data randomly generated from computational models will show inferior performance on experimental data, because of substantial distributional discrepancies (**Fig. S1A**). Furthermore, the diversity of neurotechnology makes it more challenging to approach the OOD problem with a single framework.

Here, we propose the deep domain-adaptive inference framework to systematically handle the OOD issue in inferring neural properties at various observational scales (**Fig. 1A**), which achieved state-of-the-art performance on a wide range of real experimental data. The key solution of our framework is to align the distributions of model-generated synthetic data with experimental data through *self-training* and *domain adaptation* (20, 21) (**Fig. 1A**), which can be elegantly summarized as four backbone equations (equation (1)-(4)). Once the feature distributions of synthetic data and experimental data are similar after alignments, the optimal inference performance of the ML models on synthetic data can be well generalized to experimental data. In principle, our framework is compatible with multiple ML architectures. In this study, our framework is based on a deep neural network (DNN), thus termed deep domain-adaptive inference framework.

We demonstrate the framework's efficacy in two fundamental scenarios of inferring neural properties (**Fig. 1B & C**). Notably, in both scenarios, our single framework outperforms a variety of state-of-the-art (SOTA) methods in terms of efficiency and accuracy. The first scenario is to infer detailed biophysical properties at the neuron and microcircuit scales. At the neuronal scale, our framework simultaneously inferred ion channel and membrane properties of neurons across brain regions of mice from intracellular patch clamp recordings in the Allen Cell Types Database (22), where our framework achieved 1.89 times higher inference score than the popular Evolutionary Search algorithms (23-26). In addition, our framework accurately inferred cellular and synaptic properties of a microcircuit from the extracellular recordings in the stomatogastric ganglion of *Cancer Borealis* (27), where our framework was 2.5 times more accurate in terms of the predicted biophysical patterns than Bayesian Inference methods (16). Surprisingly, in the second scenario, inferring "all-or-nothing" (1 or 0) monosynaptic connections from *in vivo* multi-electrode extracellular recordings from the CA1 region of the free-running mouse hippocampus (4), our method achieved a *near-perfect* 0.988 Mathews correlation coefficient (MCC) and significantly exceeded the two state-of-the-art methods with MCCs of 0.621 and 0.579 (15, 28). The bottom line is that we get all this excellent performance with just a single framework, which no other SOTA approach can achieve.

In conclusion, we have developed a novel computational framework that enables ML techniques to use model-generated synthetic data to accurately and efficiently infer neural properties at different observational levels from various experimental data. Our framework is built upon four unified equations that can be easily applied to many other kinds of experimental data, potentially facilitating the establishment of novel disease or mechanistic models in behaving animals by linking cellular and subcellular mechanisms to large-scale in vivo calcium imaging (8) or electrophysiological recordings (9) through our framework. Furthermore, our framework explores the mechanistic formation of neural

representations underlying behavior and cognition from a bottom-up perspective, providing a valuable complement to current statistical and geometrical analysis on large-scale neural representations from a top-down perspective (6, 29, 30).

Results

Deep domain-adaptive inference framework

Modern ML methods rely heavily on labeled data to achieve high performance, while the lack of ground-truth labels of neural properties of experimental data poses a significant challenge to these methods. Here, we present the deep domain-adaptive inference framework to infer neural properties from various "unlabeled" experimental data based on computational models (**Fig. 1A**). Our framework takes the experimental data as the input and outputs the corresponding neural properties. To tackle the shortage of labels of experimental data, we use labeled synthetic data generated from computational models to train the ML model of our framework (**Fig. 1A**). However, the distributions of experimental data (domain A) and synthetic data (domain B) have large discrepancies (**Fig. 1A**). As a result, ML models trained solely on labeled synthetic data will yield poor performance on experimental data, known as the OOD problem (17, 18). Therefore, we introduce domain adaptation and self-training to automatically assimilate the (statistical) distributions of the two domains, which yields high performance on experimental data (**Fig. 1A**).

The pipeline of our framework consists of four key steps (**Fig. 1A**): (1) to build up a computational model to describe the mechanisms of the neural properties underlying unlabeled experimental data (x_{exp}) via model parameters (y_{syn}); (2) to generate a labeled synthetic dataset (data: x_{syn}; labels: y_{syn}) by simulating the computational model with randomly sampled model parameters (i.e., the labels); (3) to train the ML model to align the discrepant distributions of synthetic data and experimental data at the feature space with self-training and domain adaptation, and hence the OOD problem can be alleviated; (4) to output the inferred neural properties from unlabeled experimental data. The accuracy will be evaluated either by comparing some crucial features of experimental data and responses from the computational model with the inferred neural properties, or by ground-truth labels obtained experimentally.

The diversity of experimental data poses a big obstacle to achieving a unified framework. To deal with various types of experimental data, we further instantiate step (3) of the pipeline, the core of our framework to tackle the OOD problem, to cover two essential inference scenarios: inferring biophysical properties and inferring monosynaptic connectivity.

Inferring biophysical properties is a regression problem where the neural properties are continuous values, meaning that each experimental datum corresponds to a distinct set of neural properties. Therefore, simple alignments between the entire distributions of the two domains will yield poor performance. To achieve precise alignment of each experimental datum, we define each experimental datum as a separate

sub-domain, which consists of the experimental datum and some synthetic data that share similar biophysical features (**Fig. 1B**). Then, self-training and domain adaptation will align the distributions of experimental data and synthetic data within each sub-domain (**Fig. 1B**). By contrast, inferring monosynaptic ("all-or-nothing") connectivity is a binary classification problem, so there are only two discrete sub-domains (i.e., connection or no connection. Similarly, self-training and domain adaptation will align the distributions of experimental data and synthetic data based on the two discrete subdomains (**Fig. 1C**).

We implement our framework based on the ResNet-LSTM-FC architecture because of its outstanding performance on synthetic data (31). In our previous work, we demonstrated that this architecture significantly outperforms the normal convolution neural network (CNN) architecture in many neural inference tasks on synthetic datasets (31). Specifically, this DNN architecture has two key components: temporal feature extractor f (ResNet-LSTM) and regressor or classifier g (FC) (**Fig. 2A**). The temporal feature extractor f extracts both local and long-range temporal features, and then the regressor or classifier outputs the associated neural properties (**Fig. 2B**).

The loss we use to train the DNN consists of two parts: self-training loss (L_{ST}) and domain adaptation loss (L_{DA}), given by equation (1):

$$L = L_{ST} + L_{DA} \#(1)$$

In each training step, we use backpropagation (32) to tune the parameters of the temporal feature extractor (θ_f) and regressor or classifier (θ_g) based on the loss function.

The self-training loss (L_{ST}) enables the DNN to directly learn to extract temporal features from experimental data and to maintain the high accuracy of experimental data during training (**Fig. 2C & E**), which is achieved by applying supervised learning on unlabeled experimental data. Pseudo labels (y_{pseudo}) are assigned to the unlabeled experimental data, which are defined as the output of the DNN from a selected previous training step (**Fig. 2E**; Methods). Then supervised learning can be applied to experimental data to minimize the error between outputs at the current training step and the pseudo labels, given by equation (2):

$$L_{ST} = \frac{1}{N_{exp}} \sum_{i=1}^{N_{exp}} \mathcal{L}\left(g\left(f\left(x_{exp}^{(i)}\right)\right), y_{pseudo}\left(x_{exp}^{(i)}\right)\right) \#(2)$$

where $y_{pseudo}\left(x_{exp}^{(i)}\right)$ is the pseudo label of experimental datum $x_{exp}^{(i)}$, and $\mathcal{L}$ is the loss function. As a result, the temporal feature extractor can capture rich temporal characteristics of experimental data, such

as subthreshold oscillations or spike dynamics of an intracellular patch-clamp recording (**Fig. 2C & Fig. S2;** Methods).

Domain adaptation is performed at the feature space, which is formed by the temporal features of all experimental data and synthetic data ($f(x_{exp}) = \left\{f\left(x_{exp}^{(i)}\right)\right\}_{i=1}^{N_{exp}}$ and $f(x_{syn}) = \left\{f\left(x_{syn}^{(i)}\right)\right\}_{i=1}^{N_{syn}}$) (**Fig. 2A, C, D & F**), termed as experimental features and synthetic features. The domain adaptation loss (L_{DA}) is the sum of two factors: (1) synthetic prediction loss on synthetic data (L_{syn}, equation (3), **Fig. 2F & G**) which guarantees the high accuracy on synthetic data, and (2) discrepancy loss between synthetic features and experimental features ($L_{dis}(f(x_{syn}), f(x_{exp}))$) which aims to shorten the distance between the distributions of synthetic features ($f(x_{syn})$) and experimental features ($f(x_{exp})$), given by equation (4):

$$L_{syn} = \frac{1}{N_{syn}} \sum_{i=1}^{N_{syn}} \mathcal{L}\left(g\left(f\left(x_{syn}^{(i)}\right)\right), y_{syn}^{(i)}\right) \#(3)$$

$$L_{DA} = L_{syn} + \lambda_{dis}\, L_{dis}\left(f(x_{syn}), f(x_{exp})\right) \#(4)$$

where λ_{dis} is the scaling factor of the discrepancy loss. After the high accuracy is achieved on the synthetic dataset and the distance between synthetic feature distribution and experimental feature distribution is shortened, the high accuracy can be transferred to experimental data (**Fig. 2G**).

Overall, equations (1)-(4) form our framework's backbone and are unified across different application scenarios (**Fig. 2F & G**). Critically, the discrepancy loss (L_{dis}) has flexible implementations in inferring biophysical properties and inferring monosynaptic connectivity, which will be described in the following sections. In addition, as the extracted features and the pseudo labels are often noisy and inaccurate at the beginning of the training process, including these data for training will deteriorate the performance. Therefore, experimental data will be gated by a task-dependent similarity metric to judge whether the errors of pseudo-labels are tolerable for self-training and domain adaptation (gate in **Fig. 2A**). We term these two thresholds as the self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}) respectively.

Below, we will elaborate on applying this unified framework in a set of applications, ranging from inferring single-neuron and microcircuit biophysics to inferring network connectivity, and compare the framework's performance with SOTA ML methods in each application.

Inferring biophysical properties of single neurons across different brain regions of mice

The neocortex exhibits diverse neuron types and cortical functions are governed by the interactions between different types of neurons (2, 33). Inferring neuronal properties of different neuron types is fundamental to understanding cell-type-specific biophysics, functions, and development. Here, we use our framework to infer the biophysical properties of different neuron types in 14 brain regions of mice in the Allen Cell Types Database. Following the methodology of previous studies (24, 34), models of different cell types use a unified set of ion channel species. The computational model is a single-compartment HH model, composed of a membrane, a passive leaky channel, and 3 active ion channels, which has been demonstrated to capture a wide range of dynamics of cortical neuron types (34) (**Fig. 3A**; Methods). There are 8 biophysical properties to be inferred including the passive membrane properties, background noise magnitude, the maximal conductance of ion channels, the time constant scaling factor of ion channels, and the spike threshold (**Fig. 3A**; Methods).

Inferring biophysical properties of neurons is a regression task, which aims at fully replicating the input-output patterns of the real neural structure by fitting the model parameters to the experimental data. In this case, each experimental datum is associated with a different set of neural properties, and hence each forms a "sub-domain" (**Fig. 1B & Fig. 3B**). Each sub-domain consists of the experimental datum and its associated similar pool. Each similar pool contains the top-N most similar synthetic data to the experimental datum based on a task-dependent similarity metric (**Fig. 3C & D**). Domain adversarial training (35) is used to assimilate the feature distributions of the experimental datum and its similar pool. A domain classifier h is trained to classify the features as synthetic or experimental features, while the temporal feature extractor is trained to fool the domain classifier by learning domain-invariant features (Methods). The discrepancy loss is expanded to equations (5)-(6):

$$L_{dis}\left(f\left(D_{syn}\right), f\left(D_{exp}\right)\right) = -\frac{1}{N_{exp}}$$

$$\sum_{x_{exp}\in D_{exp}}\left(\mathcal{L}_{CE}\left(h\left(f\left(x_{exp}\right)\right), 1\right) + \frac{1}{\left|SP\left(D_{syn}, x_{exp}\right)\right|}\sum_{x_{syn}\in SP\left(D_{syn}, x_{exp}\right)}\mathcal{L}_{CE}\left(h\left(f\left(x_{syn}\right)\right), 0\right)\right) \#(5)$$

$$\hat{\theta}_h = \arg\min_{\theta_h} -L_{dis}\left(f\left(D_{syn}\right), f\left(D_{exp}\right)\right) \#(6)$$

where $SP\left(D_{syn}, x_{exp}\right)$ is the similar pool of experimental data x_{exp} in the synthetic dataset, and $\left|SP\left(D_{syn}, x_{exp}\right)\right|$ is the size of the similar pool, and θ_h is the weights of the domain classifier, and $\mathcal{L}_{CE}$ is the cross-entropy loss. The domain adversarial loss approximates the $\mathcal{H}$~divergence which describes how the hypothesis space $\mathcal{H}$ can distinguish two distributions (36, 37).

Synthetic data in a similar pool will be ranked based on their similarity to the corresponding experimental datum, and the synthetic label of the one ranked 1st (i.e., the most similar one) will be the pseudo label of the experimental datum (**Fig. 3C**). Then self-training can be applied to the experimental data and their corresponding pseudo labels, given by equation (2). After each training step, similar pools and pseudo labels will be updated by the new inferences of experimental data based on the similarity metric (**Fig. 3B**).

To infer the biophysical properties of single neurons, we define the similarity metric as the z-scored mean absolute error (MAE) of a set of biophysical patterns between synthetic data and experimental data (Methods), and the lower the MAE the more similar the data will be. The biophysical patterns of neuronal responses include spiking statistics and membrane potential characteristics at action potential and resting state (**Fig. 3D**; Methods).

Our framework accurately inferred the biophysical properties of 574 neurons across 14 brain regions of mice simultaneously (**Fig. 4A & B**; Methods). The results show that our framework can deal with large-scale experimental data efficiently and can intrinsically be applied to data recorded from different neuron types and brain regions (**Fig. 4A**). The MAE of the biophysical patterns of all the 574 neurons is calculated and sorted by brain regions (**Fig. 4B**), where there is no preference for a particular brain region regarding inference accuracy. We demonstrate the MAE of each biophysical pattern of all 574 neurons. While closely examining the results, both spike and subthreshold patterns, two of the major distinguishable patterns across different neuron types, are well predicted, indicating that our inference results can potentially facilitate the study of subcellular differences of different neuron types. Among all biophysical patterns, the half-spike width is the most poorly tuned biophysical pattern (**Fig. 4C & D**). There are two possible reasons for this. Firstly, the single-compartment HH model is relatively simple to fit this biophysical pattern, while increasing the number of (sodium) channels could give a better fitting accuracy (24). Secondly, the tolerance of the half-spike width feature is set very low ($\sigma = 0.001$ ms; Methods and **Table S1**), so even an MAE of 20 is a discrepancy of just 0.02 ms, which is still a small error.

Evolutionary search (ES) algorithms are the perhaps most popular methods in the inference of single neuron biophysics, so we compared the inference accuracy and efficiency between our framework and a typical ES method (38) (Methods). The inference efficiency is defined as the number of synthetic data needed to achieve a given number of good inferences, where good inference is defined as the inference whose MAE of biophysical patterns is below a threshold. Our framework can obtain a 1.89 times higher inference score (i.e., the number of good inferences given the same sample size) than the ES algorithm

(our framework: 574, ES: 303, total: 1052) (**Fig. 4E**), and the outstanding performance of the framework is insensitive to the choice of the good inference threshold (**Fig. S3A**). Also, our framework could achieve a larger number of good inferences with much fewer synthetic samples compared to the ES algorithm (**Fig. 4E**). Additionally, our performance is not sensitive to the self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}) (**Fig. S3B**). We further demonstrate that high performance can only be obtained by recruiting domain adaptation and self-training in deep learning. Due to the dramatic differences in the distribution of the synthetic data and experimental data, the vanilla DNN trained only on synthetic data will perform poorly on experimental data (**Fig. S1A**) even if it performs extremely well on the synthetic data (**Fig. S1B**).

Inferring biophysical properties of the stomatogastric ganglion microcircuit of the *Cancer Borealis*

Central pattern generators in invertebrates, such as the stomatogastric nervous system, have helped uncover the generation of rhythmic patterns or oscillations (39, 40), which are the basis of motor behavior. The precise timing of the rhythmic pattern is the result of the interaction of cellular and synaptic properties of the circuit. Therefore, inferring the cellular and synaptic properties of the microcircuit will gain insights into the relationship between (sub-) cellular properties and motor behavior. Here, we apply our framework to simultaneously infer cellular and synaptic properties of the stomatogastric ganglion microcircuit of the *Cancer Borealis* (27). The computational model consists of 3 neurons and 7 synaptic connections (**Fig. 5A**). There are 31 properties to be inferred to reproduce the pyloric patterns of the circuit behavior, including 7 maximal ion channel conductances and 1 leaky channel conductance for each neuron, and 7 synaptic conductances (**Fig. 5A**; Methods). The extracellular recordings of the microcircuit were taken at $11°C$, and the similarity metric is the MAE of a set of biophysical patterns regarding the phase statistics, burst durations, and cycle periods of the firing patterns of the three neurons (**Fig. 5B;** Methods). Inferring biophysical properties of the microcircuit is also a regression problem and hence follows the same inference procedure as inferring single-neuron biophysics (**Fig. 1B & Fig. 3B**).

Sequential neural posterior estimation (SNPE) is currently the SOTA method to infer the properties of the microcircuit (16). So, we compared the inference performance of our framework with the SNPE-C method (16) (**Fig. 5E**; performance given by the pretrained model from Gonçalves et al (16); Methods). Our framework has 2.54 times smaller inference errors for all biophysical patterns than SNPE-C (MAE of our framework: 0.0136, SD=0.0024; MAE of SNPE-C: 0.0346, SD=0.0048) (**Fig. 5E**). Furthermore, our outstanding performance is not sensitive to the choices of hyperparameters, including the self-training threshold (S_{ST}), domain-adaptation threshold (S_{DA}) and discrepancy loss scaling factor (λ_{dis}) (**Fig. S4**).

These biophysical patterns, especially the rhythmic firing patterns, are crucial to the circuit mechanisms of motor generation. For example, some of the rhythmic patterns, such as the time cycle, can be largely modulated by temperature and neuromodulators (27). Thus, inferences that can accurately describe these patterns may reveal the modulation effect on the subcellular properties of a microcircuit. It is worth mentioning that our deep domain-adaptive framework is fully compatible with SNPE methods, further strengthening SNPE's ability to infer multiple sets of neural properties from unlabeled experimental data (see Discussion).

Inferring monosynaptic connectivity of networks in the CA1 region of mice

It is currently feasible to simultaneously record the activity of a large number of neurons *in vivo*, which sheds light on uncovering the mechanisms of brain functions at the network level (7, 9, 41-43). One crucial step in studying network functionality is reconstructing their connectomic map since connectivity patterns underlie functions (44). Here, we apply our framework to infer ("all-or-nothing") monosynaptic connectivity between neurons in mouse hippocampal networks from *in vivo* multi-electrode extracellular recordings (**Fig. 6A**). The connectivity between neuron pairs can be reflected by their cross-correlograms (CCGs) (45). CCGs are a series of time-lagged correlations of two spike trains, which are widely used to characterize the patterns of neural connectivity (4, 15, 28, 46). Therefore, CCG is the input to our framework, while the presence or absence of connections between neuron pairs is the output (**Fig. 6A**). The computational model is an MAT neural network of 1000 neurons with randomly initiated synaptic connections (15) (**Fig. 6A**; Methods), which can reproduce spike timing features similar to the real biological network in cortex and hippocampus (47, 48).

Unlike previous tasks of inferring biophysical properties of single neurons or microcircuits, inferring "all-or-nothing" connectivity is a binary classification task where each class represents a sub-domain (**Fig. 1C**). In each training iteration, labeled synthetic data form the two original sub-domains, and then experimental data are clustered with synthetic data based on the distance between experimental features ($f(x_{exp})$) and synthetic features ($f(x_{syn})$) (**Fig. 6B**; Methods). Notably, labels of synthetic data are indispensable, which guarantees that synthetic data with different labels will not be clustered in the same group. After clustering, some experimental data will be assigned with a pseudo label depending on the label of synthetic data in the same cluster. Similarly, to avoid noisy pseudo-labels, experimental data whose distance to the cluster center exceeds the self-training threshold (S_{ST}) or the domain adaptation threshold (S_{DA}) will be excluded from the self-training process or the domain adaptation process, respectively.

For domain adaptation, the discrepancy loss (L_{dis}) is the estimated contrastive domain discrepancy (CDD) at the feature space (49), given by equation (7)-(8):

$$L_{dis}\left(f\left(D_{syn}\right), f\left(D_{exp}\right)\right) \triangleq \frac{1}{2}\left(\underbrace{\sum_{c=0}^{1} \widehat{D}^{cc}\left(f\left(D_{syn}^{c}\right), f\left(D_{exp}^{c}\right)\right)}_{intra\ class\ distance} - \underbrace{\sum_{c=0}^{1} \sum_{\substack{c'=0 \\ c' \neq c}}^{1} \widehat{D}^{cc'}\left(f\left(D_{syn}^{c}\right), f\left(D_{exp}^{c'}\right)\right)}_{inter\ class\ distance}\right) \#(7)$$

$$D_{syn}^{c} = \left\{x_{syn} | c = y_{syn}, \left(x_{syn}, y_{syn}\right) \in D_{syn}\right\}, D_{exp}^{c} = \left\{x_{exp} | c = y_{pseudo}\left(x_{exp}\right), x_{exp} \in D_{exp}\right\} \#(8)$$

where, $\widehat{D}^{cc'}\left(f\left(D_{syn}^{c}\right), f\left(D_{exp}^{c}\right)\right)$ $(c, c' \in \{0,1\})$ is the empirical kernel mean embedding which describes the distributional distance at the feature space between synthetic data labeled as class c and experimental data pseudo labeled as class c' (Methods). Therefore, domain adaptation will minimize the distance between synthetic features and experimental features in the same class and maximize the distance between the two classes (**Fig. 6B**; equation (7)). Besides, self-training is applied to train the DNN on experimental data with pseudo labels (**Fig. 6B**).

To illustrate the mechanism of domain adaptation and self-training, we visualize the feature space reduced to a 2-dimension space by t-SNE (50) of our framework and the vanilla DNN without domain adaptation and self-training techniques (**Fig. 6C**; Methods). Before training, all samples are twisted together in the feature space. After training, the connection and no-connection samples of the synthetic data and experimental data are separately clustered in the feature space of our framework, whereas for the vanilla DNN, samples from different classes are still twisted. This indicates that the distributional discrepancies between synthetic data and experimental data largely affect the inference performance (i.e., the OOD problem).

We compared the inference accuracy of our framework with two SOTA methods that also use CCGs to infer monosynaptic connectivity, CoNNECT (15) and GLMCC (28) (Methods). The experimental data were collected by simultaneous extra- and juxta-cellular recordings in CA1 of freely running mice (4). Some of them have experimentally validated ground-truth labels, which are used to evaluate the inference accuracy (Methods). For CoNNECT, we used the pretrained model from Endo et al. (15). For GLMCC, we used the same set of hyperparameters as reported in Kobayashi et al. (28) (Methods). Strikingly, our framework achieves *near-perfect* results quantified with either the Matthews correlation coefficient (MCC) (our framework: 0.988, SD=0.023; GLMCC: 0.621, CoNNECT: 0.579) or AUC (our framework: 1.0, SD=0; GLMCC: 0.852; CoNNECT: 0.897), significantly outperforming the other two methods (**Fig. 6D, E & Fig. S5A**; Methods). Furthermore, we tested the sensitivity of hyperparameters of our

framework. Firstly, the longer the experimental recordings, the more accurate and stable the results will be (**Fig. 6F & Fig. S5B**). This is because more information about the interaction between two neurons can be represented in contrast to confounding factors, such as common inputs to the two neurons or random simultaneous firing (4). Secondly, we tested if the synthetic network's size affects our framework's inference accuracy. The results stayed overall stable across different sizes of the synthetic network (**Fig. 6F & Fig. S5C**). Finally, we also examined how the self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}) would affect the inference results. The two thresholds will only slightly affect the inference accuracy (**Fig. 6G & Fig. S5D**). Therefore, our framework is robust in terms of choices of hyperparameters, and it has a great potential to be directly applied to new experimental data from different sources.

Discussion

Inferring neural properties from experimental data is crucial for revealing the underlying functional mechanisms (3, 51-54). Many ML methods have been proposed to tackle this problem by training a ML model on labeled synthetic data randomly generated from computational models and then directly apply the ML model to experimental data (14, 16, 31). In this study, we pointed out that these methods will encounter the OOD problem and hence produce inaccurate neural properties which can further result in a biased understanding of the mechanisms. To address this challenge, we proposed the deep domain-adaptive inference framework with two key components: *domain adaptation* and *self-training*. Our framework demonstrated state-of-the-art performance in two fundamental inference cases: (1) inferring biophysical properties, which is a regression task; (2) inferring monosynaptic connectivity, which is a classification task. Specifically, for the first case, our framework inferred 8 biophysical properties of 574 neurons in 14 brain regions of mice simultaneously based on a single-compartment HH model, with 1.89 times higher inference score than the ES algorithm (38). And, our framework inferred 31 cellular and synaptic properties of stomatogastric ganglion circuit of *Cancer Borealis* based on a biophysical circuit model, with 2.5 times high accuracy than the SNPE-C method in terms of biophysical patterns (16). For the second case, our framework near-perfectly inferred monosynaptic connectivity of CA1 networks of mice from *in vivo* extracellular recordings based on an MAT network model, significantly outperforming CoNNECT (15) and GLMCC (28).

Our framework uses two instantiations to infer biophysical neural properties and monosynaptic connectivity. Inferring biophysical properties is a continuous regression task and its objective is to replicate the input-output patterns of the real neural structure by fitting the model parameters to experimental data. Typical methods range from ES algorithms (23-26), probabilistic modelling (16) to deep learning approaches (14, 31). Most of the methods apply iterative simulations to find the model parameters that can produce the most similar model responses to experimental data. The goodness of the fitting is often evaluated by a set of hand-crafted feature summaries (i.e., similarity metrics). These features differ from tasks, depending on the characteristics of the neural structure to be studied. These hand-crafted features are indispensable, as it is infeasible to fit experimental data in all aspects because computational models only identify the determinant subset of features of the neural system to account for the experimental data while ignoring other redundancies (55). One future improvement of our framework is to infer multiple feasible sets of neural properties that are aligned with the experimental data (16, 56-61). This could be potentially achieved by combing the temporal feature extractor, domain adaptation and self-training of our framework with SNPE methods (16).

As for Inferring monosynaptic connectivity, we evaluated our performance on experimental data with experimentally validated ground-truth labels (4), which have relatively low noise scales (**Fig. 6D**) because the two ground-truth labeling methods are only validated on clear data. To test how our framework will behave on noisy CCGs, we constructed a larger dataset where the two ground-truth labeling methods have discrepancies in their labels (**Fig. S6B**; Methods). In general, our framework and the two ground-truth labeling methods have some different predictions of these noisy CCGs (**Fig. S6A**). Specifically, our framework is more correlated with the prediction of the second ground-truth labeling method (**Fig. S6A**), possibly because the first ground-truth labeling method is overstrict to give connection labels that may ignore bidirectional connections or connections that coexist with network modulations (the left two examples in **Fig. S6B**; Methods). However, the second ground-truth dataset may also sometimes be overstrict for connection data due to its statistical test on consecutive time bins, which may miss connections that have fast and reliable spike transmission (the middle right example in **Fig. S6B**; Methods). Our framework sometimes cannot give out affirmative results on some CCGs as its predictions may be unstable in different repetitions (the right example in **Fig. S6B**). It is worth further investigation as to which method is better. One possible way to verify these methods is to increase the recording time so that the CCGs will become less noisy and the predictions will be more reliable (62, 63) (**Fig. 6F**). Overall, it's crucial and urgent to develop large-scale datasets with ground truths to evaluate these methods.

In summary, our domain adaptive neural inference framework largely alleviates the OOD problem in inferring biophysical properties and monosynaptic connectivity from experimental data. This indicates that our framework can potentially facilitate the interaction between experimental and modeling studies to uncover the relationship between low-level neural properties and high-level functions.

Methods

Deep neural network architecture

The temporal feature extractor first extracts local temporal features by ResNet and then extracts global temporal dependencies of the local features using LSTM cells (**Fig. 2B**). ResNet is one of the most powerful convolutional architectures to extract features nowadays (64). Residual Blocks (ResBlocks) are the basic components in ResNet-50 and other variants of ResNet. Each ResBlock comprises three units, each containing a convolutional layer (whose kernel size is 1, 3, 1 correspondingly), a batch normalization layer, and a leaky ReLU activation layer. A skip connection associates the input with the output of the ResBlock through a summation operation. Compared with the original ResNet-50 design, we modified the first input channel from 3 to the number of model response sequences, replaced the ReLU activation function with the Leaky ReLU activation function with slope, and replaced the Batch Normalization layers with the Group Normalization layers with 32 groups (65). We also eliminated the average pooling layer at the bottom of the original ResNet-50. The number of hidden units of LSTM is 2048. The regressor or classifier consists of 5 fully-connected layers, whose dimensions are 512, 256, 256, 128, and 128 respectively.

Dimension reduction and clustering in feature space with t-SNE and K-Means

The specimen ID of the experimental data used to visualize features in **Fig. 2C** is 509881736 from the Allen Institute Cell Types Database (22). The number of features is 512 and each feature is a 219-dimension vector. **Fig. S2** shows all 512 features. The distance metric for t-SNE dimensional reduction is cosine distance and the perplexity of t-SNE is set as 60. The t-SNE algorithm reduces the feature space dimension to 2. Elkan K-Means clustering (66) is then performed in the reduced feature space with a maximum of 10000 iterations and 1000 initializations.

For the visualization of the feature space in **Fig. 6C**, we use all experimental data and 50 synthetic data labeled with connection and no-connection each. The temporal feature vectors are first processed by principal component analysis (PCA) to a dimension of 50. The compressed feature vectors are then processed by t-SNE with the perplexity of 30 and with the distance metric of Euclidean distance.

HH single-compartment neuron model

The HH single-compartment neuron model is given by equations (9)-(10).

$$C_m \frac{dV}{dt} = g_L(E_L - V) + \bar{g}_{\mathrm{Na}} m^3 h(E_{\mathrm{Na}} - V) + \bar{g}_{\mathrm{K}} n^4 (E_K - V) + \bar{g}_M p(E_K - V) + I(t) + \sigma\eta(t) \#(9)$$

$$\frac{dq}{dt} = \frac{q_\infty(V) - q}{\tau_q(V)}, \quad q \in \{m, h, n, p\} \#(10)$$

where C_m is the membrane capacity, $E_{\mathrm{Na}} = 53$ mV, $E_{\mathrm{K}} = -107$ mV and E_{L} are reversal potentials, V_0 is the initial membrane potential, V_T is the spike threshold, $\bar{g}_{Na}, \bar{g}_K, \bar{g}_M$ are the maximal conductance of each ion channel, g_L is the leaky ion channel conductance, τ_{max} is the time constant scaling factor, σ is the noise factor, and the intrinsic noise $\eta(t)$ is the standard Gaussian noise. There are 8 biophysical properties to infer, which are $g_{\mathrm{Na}}, g_{\mathrm{K}}, g_{\mathrm{L}}, g_M, \tau_{max}, V_T, \sigma, E_L$. The lower bound is $[0.5, 10^{-4}, 10^{-4}, 10^{-4}, 50, -90, 10^{-4}, -100]$ and the upper bound is $[80, 15, 0.6, 0.6, 3000, -40, 0.15, -30]$. The stimulus protocol is a step current stimulus, where the amplitude is $\frac{0.13}{S_{soma}}$ pA/cm^2, where we set $S_{soma} = 8.32 \times 10^{-5}$ cm^2. The duration of the step current is 1 s and the stimulus onset is 0.4 s. The stimulus time step is set as 0.2 ms in the final synthetic data set. Synthetic model responses are discarded from training if their firing rates are 0, or they show spontaneous firing before stimulus onset. The discarding ratio is around 7.08.

The stomatogastric ganglion microcircuit model of the *Cancer Borealis*

The model consists of 3 single-compartment neurons: AB/PD neuron, LP neuron, and PD neuron, where the electronically coupled AB and PD neurons are modeled as a single neuron. Each neuron has 8 kinds of currents: a Na^+ current I_{Na}, a fast and a slow transient Ca^{2+} current I_{CaT} and I_{CaS}, a transient K^+ current I_A, a Ca^{2+}-dependent K^+ current I_{KCa}, a delayed rectifier K^+ current I_{Kd}, a hyperpolarization-activated inward current I_H, a leak current I_L, and synaptic currents I_s. The membrane potential is modeled as equation (11), and the synaptic current is modeled as equations (12)-(15) (16, 26, 56).

$$c_m \frac{dV}{dt} = -\sum_{j=1}^{N_{ion}} g_j(V - E_j) - \sum_{j=1}^{N_{synapse}} \bar{g}_j s_j(V - E_j) - g_L(V - E_L) + I_{noise} \#(11)$$

$$I_s = \bar{g}_s s(V_{post} - E_s) \#(12)$$

$$\frac{ds}{dt} = \frac{\bar{s}(V_{pre}) - s}{\tau_s} \#(13)$$

$$\bar{s}(V_{pre}) = \frac{1}{1 + \exp\left(\frac{V_{th} - V_{pre}}{\sigma}\right)} \#(14)$$

$$\tau_s = \frac{1 - \bar{s}(V_{pre})}{k_-} \#(15)$$

where V_{post} and V_{pre} are the membrane potential of post- and pre-synaptic neurons respectively, $V_{th} = -35$ mV is the half-activation voltage of the synapse, $\sigma = 5$ mV sets the slope of the activation curve, k_- is the rate constant for transmitter receptor dissociation rate which is $1/40$ ms for all glutamate synapses and $1/100$ ms for all cholinergic synapses, and $E_s = -70$ mV is the synaptic reversal potential for all glutamate synapses and $E_s = -80$ mV for all cholinergic synapses. At each time step, each neuron receives Gaussian noise with a mean of zero and a standard deviation of 0.001. The biophysical properties to be inferred are the maximal conductance of ion channels of each neuron: $\bar{g}_{Na}, \bar{g}_{CaT}, \bar{g}_{CaS}, \bar{g}_{A}, \bar{g}_{KCa}, \bar{g}_{Kd}, \bar{g}_{H}, g_L$ and the maximal conductances of 7 synapses: $\bar{g}_{glutamate}, \bar{g}_{cholinergic}$. The maximal conductance of ion channels is uniformly sampled. The lower bounds for the AB/PD neuron are: $[0, 0, 0, 0, 0, 25, 0, 0]$ ms/cm^2, and upper bounds are: $[500, 7.5, 8, 60, 15, 150, 0.2, 0.01]$ ms/cm^2. The lower bounds for the LP neuron are: $[0, 0, 2, 10, 0, 0, 0, 0.01]$ ms/cm^2, and upper bounds are: $[200, 2.5, 12, 60, 10, 120, 0.06, 0.04]$ ms/cm^2. The lower bounds for the PY neuron are: $[0, 0, 0, 30, 0, 50, 0, 0]$ ms/cm^2, and upper bounds are: $[600, 12.5, 4, 60, 5, 150, 0.06, 0.04]$ ms/cm^2. Maximal conductances of synapses are uniformly sampled in the logarithmic domain. The lower bound and upper bound of maximal conductance of the synapse from the AB/PD neuron to the LP neuron are 0.01 and 10000 nS, respectively. The lower bound and upper bound of the maximal conductance of other synapses are 0.01 and 1000 nS, respectively. All these settings are set the same as Gonçalves et al. (16).

Large-scale MAT neural network model

We use the MAT neural network as the computational model, which is more biologically realistic than the HH neural network in terms of fluctuated inputs (15). The membrane potential of neurons is driven by synaptic inputs and background noises. The synaptic conductance evolves with time, given by equations (16)-(17). Excitatory synaptic weights are drawn from a log-normal distribution, given by equation (18). Inhibitory synaptic weights are drawn from a normal distribution, given by equation (19). Only excitatory synaptic connections with $G \geq 0.2$ are included in the training. When scaled to different network sizes, the ratio between excitatory and inhibitory neurons is set fixed to be 4:1.

$$\tau_m \frac{dv_m}{dt} = -(v_m - V_L) - \tau_m[g_e(v_m - V_E) + g_i(v_m - V_I)] - RI_{bg} \#(16)$$

$$\frac{dg_x}{dx} = -\frac{g_x}{\tau_{s,x}} + \sum_j \sum_k G_j \delta(t - t_{jk} - d_j) \#(17)$$

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \#(18)$$

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \#(19)$$

where τ_m is the membrane time constant, v_m is the membrane potential, V_L is the leaky reversal potential, g_e and g_i are excitatory and inhibitory conductances, V_E and V_I are excitatory and inhibitory synaptic reversal potentials, RI_{bg} is the background noise, $\tau_{s,x}$ is the excitatory or inhibitory synaptic time constant, δ is the Dirac delta function, t_{jk} is the kth spike of the jth neuron, d_j is the synaptic delay, and G_j is the synaptic weight. For full model details and parameter settings see Endo et al. (15).

Training configurations to infer biophysical properties

The batch size is 50. The initial learning rate is 0.0001 and adapted by Adam optimizer (67) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The size of the similar pool is 50. At each training step, the similar pool will be updated by a new batch of synthetic data from the synthetic dataset and the synthetic data generated from the new inferences on the experimental data at the previous training step. The domain classifier consists of 5 fully connected layers, and the dimensions of the layers are 512, 256, 256, 128, and 128.

Similarity metric to infer biophysical properties of single neurons

We used 13 biophysical patterns in the similarity metric, which are similar to those used in Gouwens et al. (24). They are the average firing rate, action potential peak, fast trough depth, slow trough depth, time of slow trough, action potential width at half height, the latency to the first spike, the duration of the first interspike interval (ISI), the coefficient of variation of the ISIs, the average ISI, the adaptation index and the resting potential and whether there is spontaneous firing without stimulus. The similarity is calculated as the negative mean absolute error of all biophysical patterns, given by equation (20):

$$sim(x_{syn}, x_{exp}) = -\sum_i \frac{|p_{syn}^i - p_{exp}^i|}{\sigma^i} \#(20)$$

where x_{syn} and x_{exp} are the synthetic data and experimental data, p_{syn}^{i} and p_{exp}^{i} are the ith biophysical pattern, σ^{i} is the smallest tolerance of the ith pattern. The smallest tolerances of the first 12 patterns are set to be the same as in Gouwens et al. (24), and the smallest tolerance of the existence of spontaneous firing is set to be 0.001 (**Table S1**).

Comparison with the evolutionary search method on the inference of biophysical properties of single neurons

The single neuronal response data of mice at a step current of amplitude 130 pA are obtained from the Allen Cell Types Database (22). Data that don't have spikes during the stimulus window are discarded and 1052 data are used in total. For the ES method, we choose $(\mu + \lambda)$ as the evolutionary strategy (38). The probability of mating two individuals is 0.7 and the probability of mutating an individual is 0.3. The population size and offspring size are both 1000. The size of synthetic simulations of our framework is calculated as $(7.08 \times 50 + 1052)t$, where 7.08 is the discarding ratio when sampling the synthetic data, 50 is the batch size, and t is the number of training iterations. For the comparison in **Fig. 4**, the scaling factor of discrepancy loss is 0.0001, and the self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}) are both 3, and the discrepancy loss factor (λ_{dis}) is 1. After each training iteration, inferences will be made on all experimental data and then similarity values will be calculated between each new inference and all experimental datum.

Similarity metric to infer biophysical properties of the microcircuit

The 18 biophysical properties for the similarity calculation are adopted from Gonçalves et al. and Prinz et al. (16, 56): cycle period T (s), AB/PD burst duration d_{AB}^{b} (s), LP burst duration d_{LP}^{b} (s), PY burst duration d_{PY}^{b} (s), gap AB/PD end to LP start Δt_{AB-LP}^{es} (s), gap LP end to PY start Δt_{LP-PY}^{es} (s), delay AB/PD start to LP start Δt_{AB-LP}^{ss} (s), delay LP start to PY start Δt_{LP-PY}^{ss} (s), AB/PD duty cycle d_{AB}, LP duty cycle d_{LP}, PY duty cycle d_{PY}, phase gap AB/PD end to LP start $\Delta\emptyset_{AB-LP}$, phase gap LP end to PY start $\Delta\emptyset_{LP-PY}$, LP start phase $\emptyset_{LP}$, and PY start phase $\emptyset_{PY}$, and the maximal duration of membrane potential above -30 mV for each neuron (ΔD_{AB}, ΔD_{LP}, ΔD_{PY}). If the maximal duration is below 5 ms, it is set to 5 ms. Some features can only be calculated if there is a rhythmic burst firing and are set to infinity if absent. The similarity metric is the same as described in equation (20) and is calculated as the average of 10 simulations with different random seeds. The smallest tolerances of all features are

summarized in **Table S2**, which are the standard deviations of features from a set of model responses generated from uniformly sampled parameters.

Comparison with the SNPE-C on the inference of biophysical properties of the microcircuit

We used extracellular recordings of the microcircuit collected at $11\ °C$, see Haddad and Marder et al. (27) for experimental details. The size of the synthetic dataset is 174000, all of which contain bursts for all neurons. For comparison in **Fig. 5**, the scaling factor α for discrepancy loss is 0.05, the similarity threshold for self-training is 0.1, and the domain adaptation loss factor and self-training loss factor are both 1. 6 randomly sampled traces with a length of 12s from the extracellular recordings are used as the experimental data. We used pretrained SNPE-C from Gonçalves et al. (16), which can be found at https://github.com/mackelab/delfi/. The density estimator is masked autoregressive flow (68) with 5 masked autoencoders for distribution estimation (69) with [200, 400] hidden units each. There are 5 repetitions. For each repetition, we drew 100000 samples from the estimated posterior distribution and selected the one with the highest probability to calculate the MAE of biophysical patterns. The MAE reported in **Fig. 5** was the average of the 5 repetitions.

Ground truth of synaptic connections *in vivo*

The experimental data are juxtacellular-extracellular recordings from the hippocampal CA1 region in awake mice (4). We only considered excitatory synaptic connections. Neurons whose firing rate is less than 3 Hz and spikes during ripple events are eliminated from the analysis. All CCGs in both datasets are calculated from spontaneous spikes, which means all juxta-cellularly evoked spikes of pyramidal neurons are eliminated. The raw experimental data can be found at https://buzsakilab.nyumc.org/datasets/McKenzieS.

For the first ground-truth labeling method, the time bin for CCGs is 0.4 ms and the length was 200 ms. Two criteria are considered for the analysis of the presence of the synaptic connections. The peak of the CCG in the interval $[0.8, 2.8]$ ms should firstly be compared to the slowly co-modulated baseline (equation (21)), and then compared to the largest peak in the anti-causal direction (equation (22)). The lower frequency baseline λ_{slow} is generated from the CCG which is convolved with a partially hollow Gaussian kernel, with a standard deviation of 10 ms and a hollow fraction of 60%. The probability to get an observed (or higher) synchrony count in the mth time lag of the cross-correlogram given λ_{slow} in the same bin is estimated using Poisson Distribution with a continuity correction.

Pyramidal-inhibitory pairs that are considered as synaptic connections should have $P_{fast} < 0.001$ and $P_{causal} < 0.0026$, that are considered as absent connections should have $P_{fast} > 0.15$ and $P_{causal} > 0.15$ for the small dataset used for comparison of inference accuracy (**Fig. 6**), while $P_{fast} > 0.1$ and $P_{causal} > 0.1$ for the larger dataset used in **Fig. S6**. The script to perform the ground-truth generation can be found at https://github.com/buzsakilab/buzcode.

$$P_{fast}\left(n\ or\ more \middle| \lambda_{slow}(m)\right) = 1 - \sum_{x=0}^{n-1} \frac{e^{-\lambda_{slow}(m)}\lambda_{slow}(m)^x}{x!} - \frac{e^{-\lambda_{slow}(m)}\lambda_{slow}(m)^n}{2n!} \quad \#(21)$$

$$sP_{causal}\left(n\ or\ more \middle| \lambda_{anticausal}(-m)\right) = 1 - \sum_{x=0}^{n-1} \frac{e^{-\lambda_{anticausal}(-m)}\lambda_{anticausal}(-m)^x}{x!} - \frac{e^{-\lambda_{anticausal}(-m)}\lambda_{anticausal}(-m)^n}{2n!} \quad \#(22)$$

This first ground-truth labeling approach is quite conservative and can miss some valid CCGs. Typically, it filters out CCGs with a monosynaptic connection profile ($[0.5, 2]$ ms peak) superimposed over a common input profile (0 ms peak) because it does not allow any significant modulation in the anti-causal direction (between -2 and 0 ms). Therefore, we complement this approach with a second more inclusive test. As above, the peak of the CCG in the interval $[0.8, 2.8]$ ms should be significantly different from the slowly co-modulated baseline λ_{slow} (5 ms wide hollow Gaussian kernel) but using a P_{fast} threshold of 0.01 rather than 0.001 for 3 consecutive 0.3 ms bins, and without any consideration for time bins in the anti-causal direction.

We constructed two datasets of experimental CCGs. For the first small dataset used for comparison of inference accuracy, the two ground-truth labeling methods give out the same labels for all CCGs. There are 9 connection data and 66 no-connection data. For the second large dataset, there are discrepancies between the results of the two methods. There are 105 data in total. The first method gives out 9 connection data and 96 no-connection data, while the second method gives out 13 connection data and 92 no-connection data.

Training configurations to infer monosynaptic connectivity

The batch size is 50. The initial learning rate is 0.0001 and adapted by Adam optimizer (67) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We implemented domain adaptation as the contrastive adaptation network (35, 49). The empirical kernel mean embeddings of CDD are given by equations (23)-(31):

$$D^{pq}\left(f\left(D_{syn}\right), f\left(D_{exp}\right)\right) = e_1 + e_2 - 2e_3, p, q \in \{0, 1\} \#(23)$$

$$e_1 = \frac{\sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{syn}} \mu_{pq}\left(y_{syn}^i, y_{syn}^j\right) \sum_{l=1}^{5} k^l \left(f\left(x_{syn}^i\right), f\left(x_{syn}^j\right)\right)}{\sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{syn}} \mu_{pq}\left(y_{syn}^i, y_{syn}^j\right)} \#(24)$$

$$e_2 = \frac{\sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{pseudo}^i, y_{pseudo}^j\right) \sum_{l=1}^{5} k^l \left(f\left(x_{exp}^i\right), f\left(x_{exp}^j\right)\right)}{\sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{pseudo}^i, y_{pseudo}^j\right)} \#(25)$$

$$e_3 = \frac{\sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{syn}^i, y_{pseudo}^j\right) \sum_{l=1}^{5} k^l \left(f\left(x_{syn}^i\right), f\left(x_{exp}^j\right)\right)}{\sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{syn}^i, y_{pseudo}^j\right)} \#(26)$$

$$\mu_{pq}\left(y^i, y^j\right) = \begin{cases} 1, \text{if } y^i = p \text{ and } y^j = q \\ 0, \text{otherwise} \end{cases} \#(27)$$

$$\sigma^i = \frac{\sigma_{pq} 2^i}{2^{2.5}\left(N_{syn,p}^2 + N_{exp,q}^2 + 2N_{syn,p}N_{exp,q} - N_{syn,p} - N_{exp,q}\right)} \#(28)$$

$$\begin{aligned} \sigma_{pq} = & \sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{syn}} \mu_{pq}\left(y_{syn}^i, y_{syn}^j\right) l_2^2 \left(f\left(x_{syn}^i\right), f\left(x_{syn}^j\right)\right) \\ & + \sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{pseudo}^i, y_{pseudo}^j\right) l_2^2 \left(f\left(x_{exp}^i\right), f\left(x_{exp}^j\right)\right) \\ & + 2 \sum_{i=1}^{N_{syn}} \sum_{j=1}^{N_{exp}} \mu_{pq}\left(y_{syn}^i, y_{pseudo}^j\right) l_2^2 \left(f\left(x_{syn}^i\right), f\left(x_{exp}^j\right)\right) \#(29) \end{aligned}$$

$$N_{syn,p} = \left|\left\{x_{syn} | y_{syn} = p, \left(x_{syn}, y_{syn}\right) \in D_{syn}\right\}\right| \#(30)$$

$$N_{exp,q} = \left|\left\{x_{exp} | y_{pseudo}\left(x_{exp}\right) = p, x_{exp} \in D_{exp}\right\}\right| \#(31)$$

where l_2 is the Euclidean distance function, k^i is a set of 5 Gaussian kernels with the variance of σ^i.

Pseudo labels of experimental data are obtained by applying spherical K-means at the feature space among synthetic data and experimental data. The initial cluster center of class c is the synthetic cluster center, given by equation (32). Then the algorithm repeats the following two steps until convergence or reaching the maximum iterations: 1) assigning each experimental datum a pseudo label based on the nearest cluster center; 2) updating the cluster center based on experimental data, given by equation (33).

$$O_0^c = \sum_{(x_{syn}, y_{syn}) \in D_{syn}} I(y_{syn} = c) \frac{f(x_{syn})}{\|f(x_{syn})\|}, c \in \{0, 1\} \#(32)$$

$$O_t^c = \sum_{x_{exp} \in D_{exp}} I(y_{pseudo}(x_{exp}) = c) \frac{f(x_{exp})}{\|f(x_{exp})\|}, c \in \{0, 1\} \#(33)$$

where O_0^c is the initial cluster center for class c, and O_t^c is the cluster center for class c at training step t.

Comparison with CoNNECT and GLMCC on the inference of neural connectivity

GLMCC method is proposed by Kobayashi et al. (28) and we used the published code at https://github.com/NII-Kobayashi. We set the membrane time constant as 5 ms, and γ as 5×10^{-4}, which are identical to the original study (28). The time length for a cross-correlogram is 100 ms and the time bin is 1 ms. Other hyperparameters remain unchanged as in the original study. Synaptic delay is tested from 1-5 ms at the step of 1 ms and selected as the one that gives the highest likelihood. As for CoNNECT, we used the pre-trained model obtained from https://github.com/shigerushinomoto/CoNNECT and set the classification threshold as 0.5.

We chose two metrics to measure the inference performance. The area under ROC curve (AUC) is a standard metric used in the evaluation of classification algorithms (70). AUC is scale-invariant and classification-threshold-invariant, which means the value of AUC depends on the rank of the results rather than the absolute values or hard classification threshold. However, AUC is not suitable when it comes to problems that need a hard threshold, so we chose another metric Matthews correlation coefficient (MCC) (71). It needs a hard classification threshold to calculate its value, so it can measure how well a threshold is set. Also, MCC is more informative than F1-score or accuracy in binary classification when the labels of two classes are imbalanced (72).

References

1. S. Lodato, P. Arlotta, Generating neuronal diversity in the mammalian cerebral cortex. *Annu Rev Cell Dev Biol* **31**, 699-720 (2015).

2. H. Zeng, J. R. Sanes, Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat Rev Neurosci* **18**, 530-546 (2017).

3. A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, L. F. Abbott, Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153-1164 e1157 (2017).

4. D. F. English, S. McKenzie, T. Evans, K. Kim, E. Yoon, G. Buzsaki, Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks. *Neuron* **96**, 505-520 e507 (2017).

5. S. Gorur-Shandilya, E. M. Cronin, A. C. Schneider, S. A. Haddad, P. Rosenbaum, D. Bucher, F. Nadim, E. Marder, Mapping circuit dynamics during function and dysfunction. *Elife* **11**, (2022).

6. A. E. Urai, B. Doiron, A. M. Leifer, A. K. Churchland, Large-scale neural recordings call for new insights to link brain and behavior. *Nat Neurosci* **25**, 11-19 (2022).

7. G. Buzsaki, Large-scale recording of neuronal ensembles. *Nat Neurosci* **7**, 446-451 (2004).

8. C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, K. D. Harris, Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, 255 (2019).

9. N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Bohm, M. Broux, S. Chen, J. Colonell, R. J. Gardner, B. Karsh, F. Kloosterman, D. Kostadinov, C. Mora-Lopez, J. O'Callaghan, J. Park, J. Putzeys, B. Sauerbrei, R. J. J. van Daal, A. Z. Vollan, S. Wang, M. Welkenhuysen, Z. Ye, J. T. Dudman, B. Dutta, A. W. Hantman, K. D. Harris, A. K. Lee, E. I. Moser, J. O'Keefe, A. Renart, K. Svoboda, M. Hausser, S. Haesler, M. Carandini, T. D. Harris, Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* **372**, (2021).

10. C. R. Cadwell, A. Palasantza, X. Jiang, P. Berens, Q. Deng, M. Yilmaz, J. Reimer, S. Shen, M. Bethge, K. F. Tolias, R. Sandberg, A. S. Tolias, Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* **34**, 199-203 (2016).

11. C. Grienberger, A. Konnerth, Imaging calcium in neurons. *Neuron* **73**, 862-885 (2012).

12. E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* **8**, 1263-1268 (2005).

13. A. M. Packer, D. S. Peterka, J. J. Hirtz, R. Prakash, K. Deisseroth, R. Yuste, Two-photon optogenetics of dendritic spines and neural circuits. *Nat Methods* **9**, 1202-1205 (2012).

14. R. Ben-Shalom, J. Balewski, A. Siththaranjan, V. Baratham, H. Kyoung, K. G. Kim, K. J. Bender, K. E. Bouchard, Inferring neuronal ionic conductances from membrane potentials using CNNs. *bioRxiv*, 727974 (2019).

15. D. Endo, R. Kobayashi, R. Bartolo, B. B. Averbeck, Y. Sugase-Miyamoto, K. Hayashi, K. Kawano, B. J. Richmond, S. Shinomoto, A convolutional neural network for estimating synaptic connectivity from spike trains. *Sci Rep* **11**, 12087 (2021).

16. P. J. Gonçalves, J. M. Lueckmann, M. Deistler, M. Nonnenmacher, K. Ocal, G. Bassetto, C. Chintaluri, W. F. Podlaski, S. A. Haddad, T. P. Vogels, D. S. Greenberg, J. H. Macke, Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife* **9**, (2020).

17. Z. Shen, J. Liu, Y. He, X. Zhang, R. Xu, H. Yu, P. Cui, Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, (2021).

18. N. Sapoval, A. Aghazadeh, M. G. Nute, D. A. Antunes, A. Balaji, R. Baraniuk, C. J. Barberan, R. Dannenfelser, C. Dun, M. Edrisi, R. A. L. Elworth, B. Kille, A. Kyrillidis, L. Nakhleh, C. R. Wolfe, Z. Yan, V. Yao, T. J. Treangen, Current progress and open challenges for applying deep learning across the biosciences. *Nat Commun* **13**, 1728 (2022).

19. Y. Almalioglu, M. Turan, N. Trigoni, A. Markham, Deep learning-based robust positioning for all-weather autonomous driving. *Nat Mach Intell* **4**, 749-760 (2022).

20. Q. Xie, M.-T. Luong, E. Hovy, Q. V. Le, Self-training with noisy student improves imagenet classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10687-10698 (2020).

21. M. Wang, W. Deng, Deep visual domain adaptation: A survey. *Neurocomputing* **312**, 135-153 (2018).

22. Allen Cell Types Database. (2015).

23. S. Druckmann, Y. Banitt, A. Gidon, F. Schurmann, H. Markram, I. Segev, A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front Neurosci* **1**, 7-18 (2007).

24. N. W. Gouwens, J. Berg, D. Feng, S. A. Sorensen, H. Zeng, M. J. Hawrylycz, C. Koch, A. Arkhipov, Systematic generation of biophysically detailed models for diverse cortical neuron types. *Nat Commun* **9**, 710 (2018).

25. M. Gurkiewicz, A. Korngreen, A numerical approach to ion channel modelling using whole-cell voltage-clamp recordings and a genetic algorithm. *PLoS Comput Biol* **3**, e169 (2007).

26. A. A. Prinz, C. P. Billimoria, E. Marder, Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. *J Neurophysiol* **90**, 3998-4015 (2003).

27. S. A. Haddad, E. Marder, Circuit robustness to temperature perturbation is altered by neuromodulators. *Neuron* **100**, 609-623 e603 (2018).

28. R. Kobayashi, S. Kurita, A. Kurth, K. Kitano, K. Mizuseki, M. Diesmann, B. J. Richmond, S. Shinomoto, Reconstructing neuronal circuitry from parallel spike trains. *Nat Commun* **10**, 4468 (2019).

29. M. Jazayeri, S. Ostojic, Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr Opin Neurobiol* **70**, 113-120 (2021).

30. S. Chung, L. F. Abbott, Neural population geometry: An approach for understanding biological and artificial neural networks. *Curr Opin Neurobiol* **70**, 137-144 (2021).

31. K. Sheng, P. Qu, L. Yang, X. Liu, L. He, Y. Zhang, L. Ma, K. Du, A general LSTM-based deep learning method for estimating neuronal models and inferring neural circuitry. *bioRxiv*, 2021.2003.2014.434027 (2021).

32. D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Nature* **323**, 533-536 (1986).

33. R. Tremblay, S. Lee, B. Rudy, GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260-292 (2016).

34. M. Pospischil, M. Toledo-Rodriguez, C. Monier, Z. Piwkowska, T. Bal, Y. Fregnac, H. Markram, A. Destexhe, Minimal Hodgkin-Huxley type models for different classes of cortical and thalamic neurons. *Biol Cybern* **99**, 427-441 (2008).

35. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**, 2096–2030 (2016).

36. S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, Analysis of representations for domain adaptation. *Advances in Neural Information Processing Systems* **19**, 137–144 (2006).

37. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J. W. Vaughan, A theory of learning from different domains. *Mach Learn* **79**, 151-175 (2010).

38. F.-A. Fortin, F.-M. De Rainville, M.-A. G. Gardner, M. Parizeau, C. Gagné, DEAP: Evolutionary algorithms made easy. *The Journal of Machine Learning Research* **13**, 2171-2175 (2012).

39. E. Marder, D. Bucher, Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. *Annu Rev Physiol* **69**, 291-316 (2007).

40. D. J. Schulz, J. M. Goaillard, E. Marder, Variable channel expression in identified single and electrically coupled neurons in different animals. *Nat Neurosci* **9**, 356-362 (2006).

41. C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, K. D. Harris, High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361-365 (2019).

42. T. A. Jarrell, Y. Wang, A. E. Bloniarz, C. A. Brittin, M. Xu, J. N. Thomson, D. G. Albertson, D. H. Hall, S. W. Emmons, The connectome of a decision-making neural network. *Science* **337**, 437-444 (2012).

43. H. Ding, R. G. Smith, A. Poleg-Polsky, J. S. Diamond, K. L. Briggman, Species-specific wiring for direction selectivity in the mammalian retina. *Nature* **535**, 105-110 (2016).

44. N. Brunel, Is cortical connectivity optimized for storing information? *Nat Neurosci* **19**, 749-755 (2016).

45. S. Ostojic, N. Brunel, V. Hakim, How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J Neurosci* **29**, 10234-10253 (2009).

46. A. M. Aertsen, G. L. Gerstein, Evaluation of neuronal connectivity: sensitivity of cross-correlation. *Brain Res* **340**, 341-354 (1985).

47. R. Kobayashi, Y. Tsubo, S. Shinomoto, Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Front Comput Neurosci* **3**, 9 (2009).

48. Y. Omura, M. M. Carvalho, K. Inokuchi, T. Fukai, A lognormal recurrent network model for burst generation during hippocampal sharp waves. *J Neurosci* **35**, 14585-14601 (2015).

49. G. Kang, L. Jiang, Y. Yang, A. G. Hauptmann, Contrastive adaptation network for unsupervised domain adaptation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4888-4897 (2019).

50. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of machine learning research* **9**, (2008).

51. A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117**, 500-544 (1952).

52. A. Gidon, T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, M. E. Larkum, Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* **367**, 83-87 (2020).

53. S. Grillner, The motor infrastructure: from ion channels to neuronal networks. *Nat Rev Neurosci* **4**, 573-586 (2003).

54. K. D. Harris, T. D. Mrsic-Flogel, Cortical connectivity and sensory coding. *Nature* **503**, 51-58 (2013).

55. L. F. Abbott, Theoretical neuroscience rising. *Neuron* **60**, 489-495 (2008).

56. A. A. Prinz, D. Bucher, E. Marder, Similar network activity from disparate circuit parameters. *Nat Neurosci* **7**, 1345-1352 (2004).

57. M. G. Metzen, M. J. Chacron, Stimulus background influences phase invariant coding by correlated neural activity. *Elife* **6**, (2017).

58. E. Marder, M. L. Goeritz, A. G. Otopalik, Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Curr Opin Neurobiol* **31**, 156-163 (2015).

59. G. J. Gutierrez, T. O'Leary, E. Marder, Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron* **77**, 845-858 (2013).

60. L. M. Alonso, E. Marder, Visualization of currents in neural models with similar behavior and different conductance densities. *Elife* **8**, (2019).

61. E. Marder, J. M. Goaillard, Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci* **7**, 563-574 (2006).

62. P. Bartho, H. Hirase, L. Monconduit, M. Zugaro, K. D. Harris, G. Buzsaki, Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol* **92**, 600-608 (2004).

63. S. Fujisawa, A. Amarasingham, M. T. Harrison, G. Buzsaki, Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci* **11**, 823-833 (2008).

64. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778 (2016).

65. Y. Wu, K. He, Group normalization. *Proceedings of the European conference on computer vision (ECCV)*, 3-19 (2018).

66. C. Elkan, Using the triangle inequality to accelerate k-means. *Proceedings of the 20th International Conference on Machine Learning*, 147–153 (2003).

67. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, (2014).

68. G. Papamakarios, T. Pavlakou, I. Murray, Masked autoregressive flow for density estimation. *Advances in Neural Information Processing Systems* **30**, 2335–2344 (2017).

69. M. Germain, K. Gregor, I. Murray, H. Larochelle, MADE: masked autoencoder for distribution estimation. *Proceedings of the 32nd International Conference on Machine Learning* **37**, 881--889 (2015).

70. A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* **30**, 1145-1159 (1997).

71. B. W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* **405**, 442-451 (1975).

72. D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* **21**, 1-13 (2020).

73. C. Lau, L. Ng, C. Thompson, S. Pathak, L. Kuan, A. Jones, M. Hawrylycz, Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics* **9**, 153 (2008).

Acknowledgements

Funding

This work is supported by National Key R&D Program of China (N0. 2020AAA0130400).

Author contributions

Conceptualization and methodology: KS, KD

Investigation and analysis: KS, SZ, MB, PQ, LY

Discussion: KS, SZ, MB, PQ, LY, XL, LH, YZ, LM, KD

Supervision: KD, LM

Writing—original draft: KS, SZ, MB, LM, KD

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

The source code of our framework in this work will be available upon acceptance. All data are available in the main text or the supplementary materials.

Figures

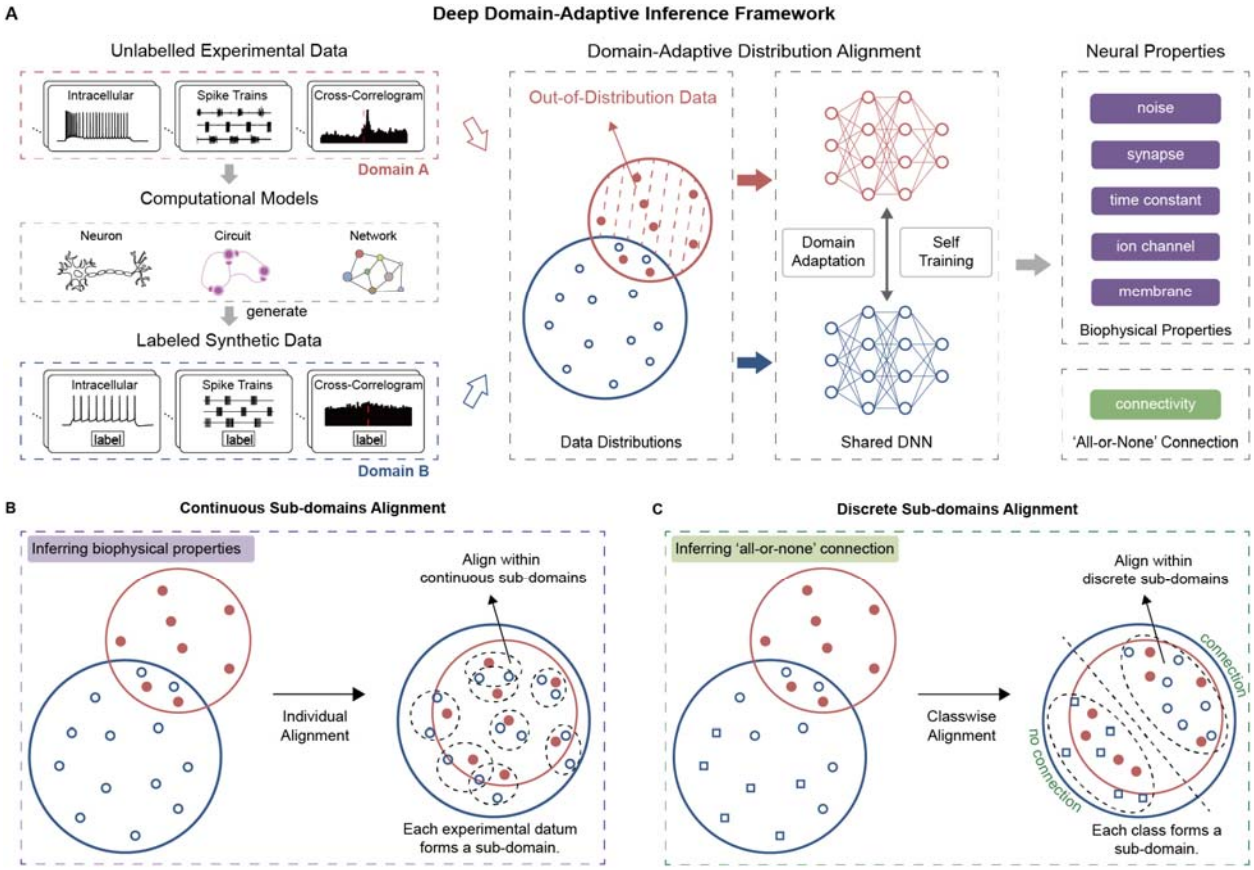

Fig. 1. Neural properties inference methodology. (**A**) Overview of our framework to infer neural properties from various experimental data based on computational models. (**B**) The instantiation of our framework on inferring biophysical properties. In this case, each experimental datum is associated with a different set of continuous neural properties (i.e., regression problem). (**C**) The instantiation of our framework on inferring the existence of connectivity. In this case, the neural properties are discrete (classification problem).

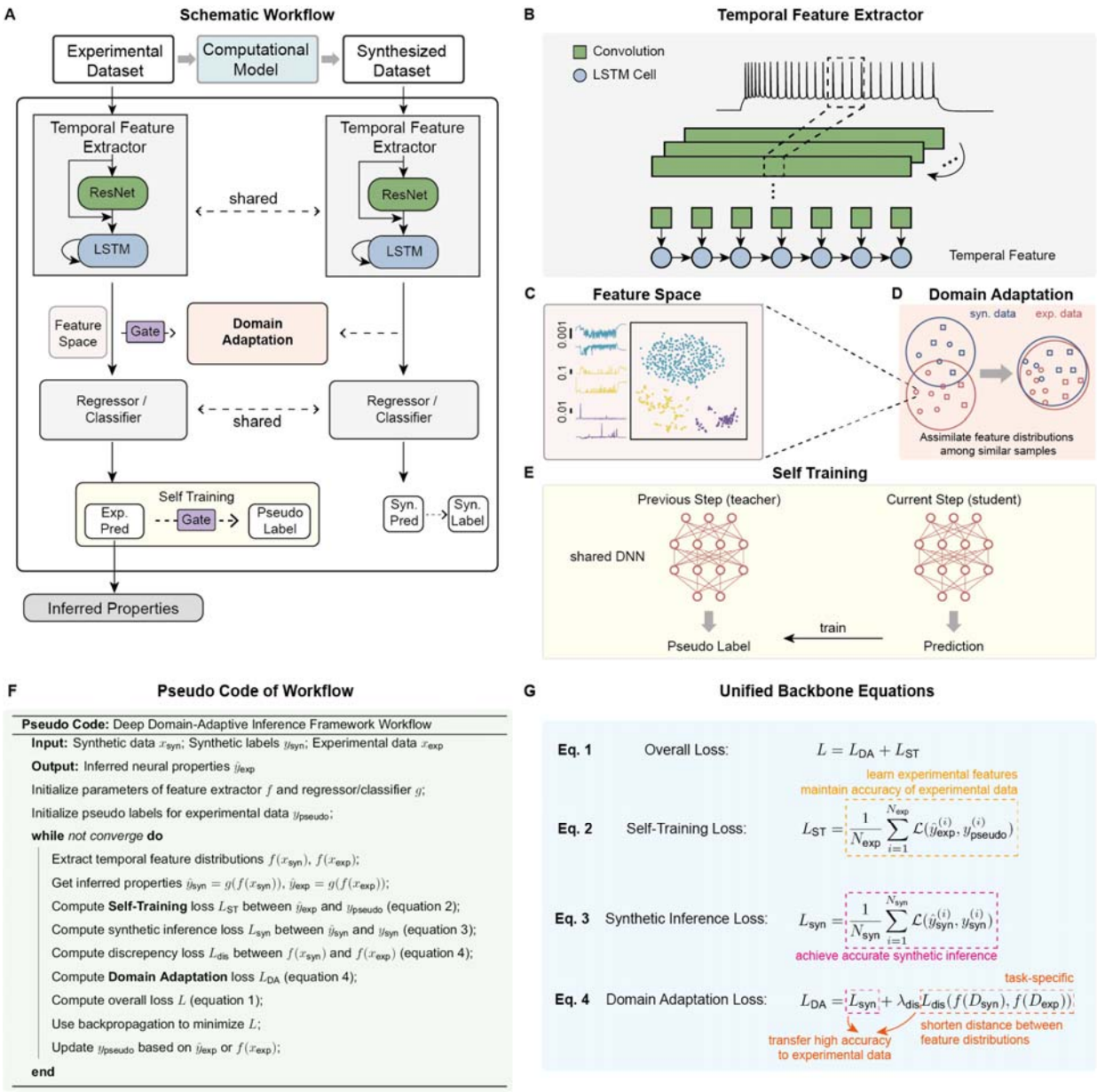

Fig. 2. Deep domain-adaptive inference framework workflow. (**A**) Workflow of the deep domain-adaptive neural inference framework. The temporal feature extractor and regressor/classifier are shared between synthetic data and experimental data. Firstly, temporal features of synthetic data and experimental data are extracted by the shared temporal feature extractor. The output space from the temporal feature extractor is the feature space. Secondly, the shared regressor or classifier processes the temporal features and outputs the inferred neural properties. (**B**) Architecture of temporal feature extractor. It consists of a ResNet and an LSTM layer (31). (**C**) Feature visualization of the experimental datum in **B**. Each dot is a dimension of the features. Each color denotes a cluster and representatives of each cluster are shown on the left (for visualizations of all features, see **Fig. S2**). (**D**) Schematic

mechanisms of domain adaptation. **(E)** Illustration of self-training. We use the prediction from the DNN in a previous training step as the pseudo label for the current step. (**F**) Pseudo code of the deep domain-adaptive inference framework. (**G**) Unified backbone equations of the framework as used in **F**. The only task-specific equation is the discrepancy loss for the two scenarios shown in **Fig. 1B & C**.

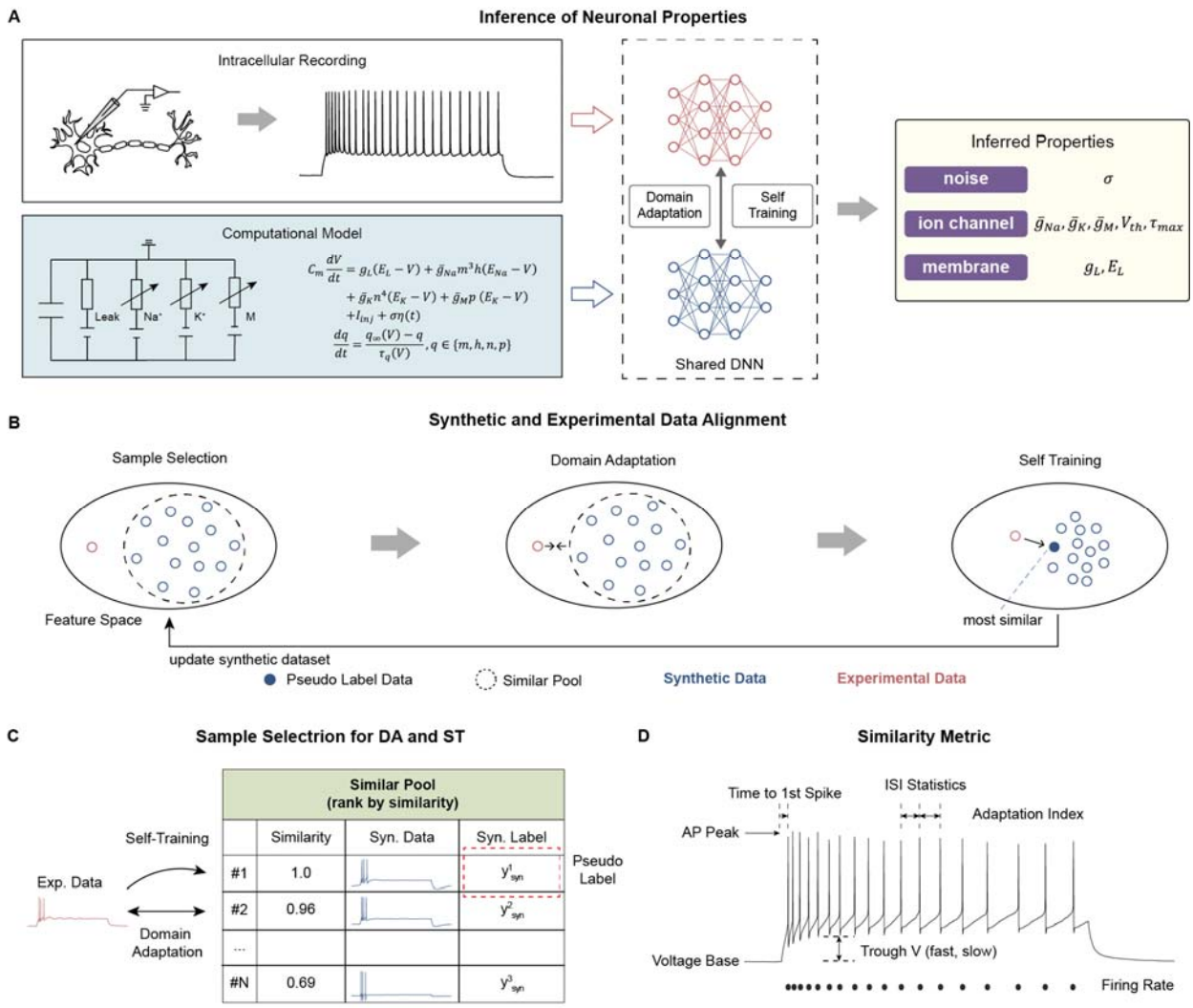

Fig. 3. Framework instantiation to infer biophysical properties. (**A**) The experimental data, computational model, and the biophysical properties to be inferred at the neuron scale. (**B**) Illustration of the instantiation of the deep domain-adaptive inference framework to infer biophysical properties. Red symbols represent features of experimental data (exp. data) and blue ones for synthetic data (syn. data). (**C**) An illustration of the similar pool for a piece of intracellular experimental data. Each similar pool contains the top-N synthetic data that are most similar to the piece of experimental data. The synthetic label of the piece of synthetic data that is most similar to the respective piece of experimental data is the pseudo label. The similarity is calculated between the synthetic model responses and their corresponding experimental data based on the similarity metric in **D**. The similarity value is normalized by the maximum similarity value. (**D**) The similarity metric for the calculation of similarities as in **C** (Methods).

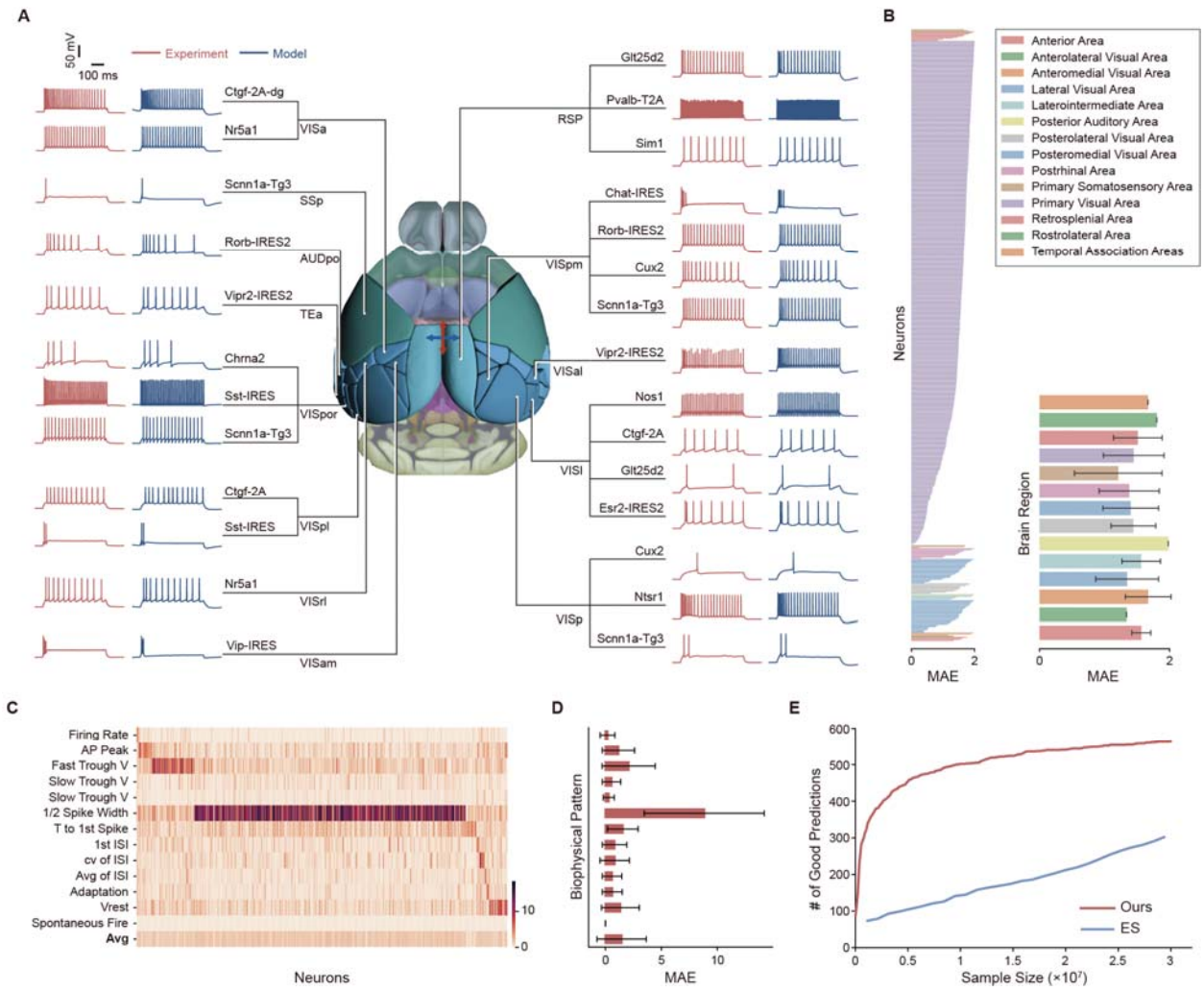

Fig. 4. Inferring biophysical properties of neurons of mice. (**A**) Examples of inferred results grouped by brain regions and neuron types. The brain region atlas is modified from Lau et al. (73). Blue curves: the model responses are generated by applying the same stimulus protocol with the inferred properties. Red curves: experimental data taken from Allen Cell Types Database (22). (**B**) Mean absolute error (MAE) of biophysical patterns between the model responses and the experimental data of 574 neurons sorted by brain regions. Insect: the average MAE of all neurons in the brain region. Error bar: standard deviation. (**C**) MAE of each biophysical pattern of 574 neuron models, sorted by the biophysical patterns with the maximum error. (**D**) MAE of each feature across all neuron models in **C**. The y-axis has the same order of labels as in c. Error bar: standard deviation across 574 models. (**E**) Sample efficiency compared to the evolutionary search (ES) method (16). The y-axis is the number of good predictions from the two methods, where good is defined as predictions whose MAE of biophysical patterns is smaller than 2.0.

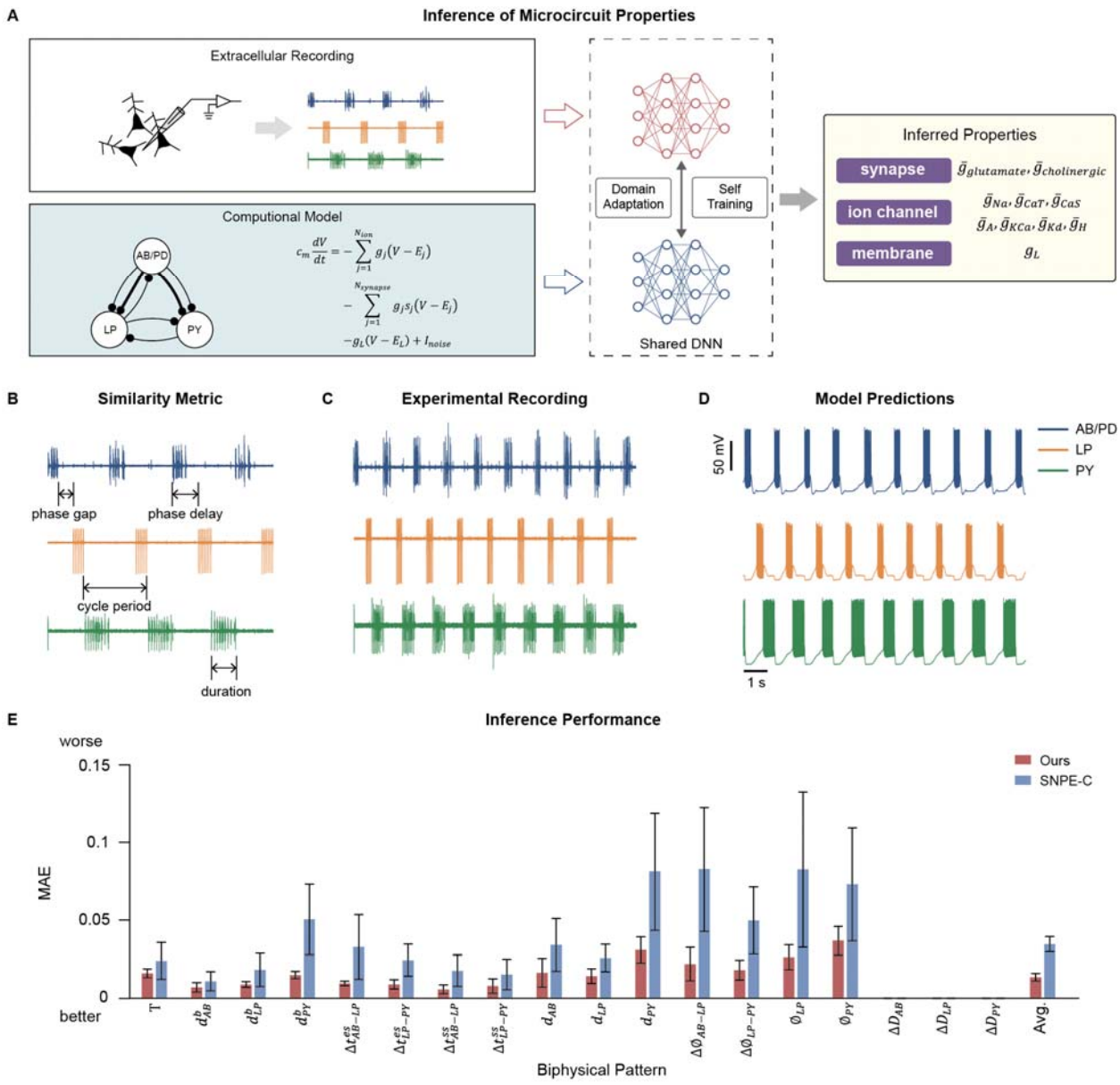

Fig. 5. Inferring biophysical properties of a microcircuit of the *Cancer Borealis*. (**A**) The experimental data, computational model, and the neural properties to be inferred at the microcircuit scale. In the illustration of the computational model panel, the thin connections are fast glutamatergic synapses, and the thick connections are slow cholinergic synapses. (**B**) The similarity metric for this task. Different pyloric statistics are compared for the calculation of similarity. (**C**) The experimental extracellular recordings of the microcircuit. (**D**) The predicted model responses by applying a similar stimulus protocol as the experiment based on a set of inferred properties of the model. (**E**) Performance comparison with SNPE-C (16) of MAE of biophysical patterns. The discrepancy loss factor is 0.001 and the self-training

threshold is 0.1. The performance of SNPE-C was given by the pretrained model from Gonçalves et al (16) (Methods). Error bar: standard deviation across 5 random repetitions.

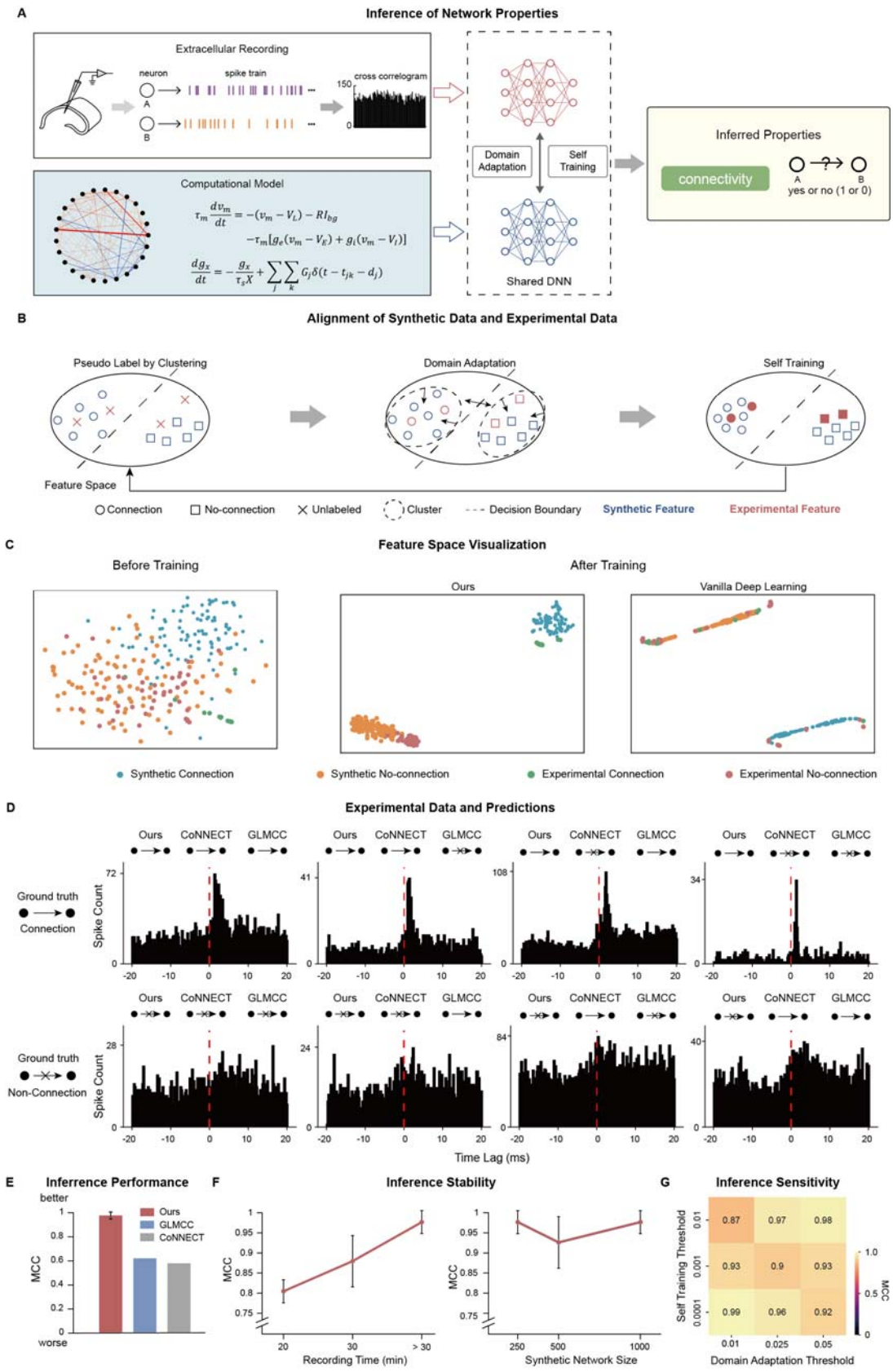
A
Inference of Network Properties
Extracellular Recording
neuron
spike train
cross correlogram
Computational Model
Domain Adaptation
Self Training
Shared DNN
Inferred Properties
connectivity
yes or no (1 or 0)
B
Alignment of Synthetic Data and Experimental Data
Pseudo Label by Clustering
Domain Adaptation
Self Training
Feature Space
Connection
No-connection
Unlabeled
Cluster
Decision Boundary
Synthetic Feature
Experimental Feature
C
Feature Space Visualization
Before Training
After Training
Ours
Vanilla Deep Learning
Synthetic Connection
Synthetic No-connection
Experimental Connection
Experimental No-connection
D
Experimental Data and Predictions
Ours
CoNNECT
GLMCC
Ground truth
Connection
Non-Connection
Spike Count
Time Lag (ms)
E
Inferrence Performance
better
worse
MCC
F
Inference Stability
Recording Time (min)
Synthetic Network Size
G
Inference Sensitivity
Self Training Threshold
Domain Adaptation Threshold

Fig. 6. Framework instantiation to infer monosynaptic connectivity. (**A**) The experimental data, computational model, and the neural properties to be inferred at the network scale. (**B**) Illustration of the instantiation of the deep domain-adaptive inference framework to infer monosynaptic connectivity. Red symbols represent features of experimental data and blue ones for synthetic data. (**C**) The visualization of the feature space of both synthetic data and experimental data our framework and vanilla deep learning without domain adaptation and self-training. (**D**) Example cross-correlograms of the experimental data and their ground-truth labels and results of inferred connectivity of our framework, CoNNECT (15), and GLMCC (28) are presented. The time bin is 0.2 ms. (**E**) The inference accuracy compared among the three methods. Error bar for our framework: standard deviation across 5 random repetitions. (**F**) The effect of the length of experimental recording time (left) and the effect of the network size of the synthetic data (right) on the inference accuracy. (**G**) The sensitivity of domain adaptation threshold (S_{DA}) and self-training threshold (S_{ST}) on the inference accuracy.