Domain Adaptive Neural Inference for Neurons, Microcircuits and Networks

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

Computational modeling is an essential approach in neuroscience for linking neural mechanisms to experimental observations. Recent advanced machine learning techniques, such as deep learning, leverage synthetic data generated from computational models to reveal underlying neural mechanisms from experimental data. However, despite significant progress, one unsolved problem in these methods is that the synthetic data differ substantially from experimental data, leading to severely biased results. To this end, we introduce the Domain Adaptive Neural Inference framework to construct synthetic data that closely resemble the distribution of experimental data and use the matching synthetic data to predict the neural mechanisms of experimental data. We demonstrate the accuracy, efficiency, and versatility of our framework in various experimental observations, including inferring single-neuron biophysics across mouse brain regions from intracellular recordings in the Allen Cell Types Database; inferring biophysical properties of a microcircuit of *Cancer Borealis* from extracellular recordings; and inferring monosynaptic connectivity of mouse CA1 networks from *in vivo* multi-electrode extracellular recordings. The framework outperforms state-of-the-art methods in every application, and can potentially be generalized to a wide range of computational modeling approaches in biosciences.

Introduction

Computational modeling is widely used in neuroscience to reveal underlying cellular and circuit mechanisms from experimental observations (Parker & Newsome, 1998; Shine et al., 2021; X. J. Wang & Krystal, 2014). Due to the limitations of experimental techniques, we cannot always directly measure the neural properties of interest. An alternative approach is to construct biophysical models based on the hypothesized mechanisms of neural properties of concern. Then, the neural properties, i.e., parameters of the models, can be inferred from the alignment between the responses of the computational models with the experimental data. Classic examples include the famous Hodgkin-Huxley (HH) equations (Hodgkin & Huxley, 1952), and the central pattern generator (Grillner, 1985), which infer the ionic membrane properties of individual neurons or microcircuits from electrophysiological recordings. Today, there is a growing trend toward the development of various experimental techniques to systematically explore neural circuit function, such as patch-seq (Cadwell et al., 2016), extracellular recordings (Buzsaki, 2004; Steinmetz et al., 2021), calcium imaging (T. W. Chen et al., 2013; Denk et al., 1994; Denk et al., 1990; Grienberger & Konnerth, 2012; Tian et al., 2009) and optogenetics (Boyden et al., 2005; Packer et al., 2012; Prakash et al., 2012; Rickgauer & Tank, 2009). Accordingly, computational models have also increased their complexity to cover the scales and details (Billeh et al., 2020; Markram et al., 2015). However, the diversity of experimental data and the complexity of computational models pose enormous challenges for model fitting and inferences of neural properties.

To tackle this challenge, many advanced machine learning techniques have been introduced and made significant progress (Ben-Shalom et al., 2019; Endo et al., 2021; Gonçalves et al., 2020; Sheng et al., 2021), as a result of their extraordinary abilities in dealing with complex high-dimensional data. The common approach used by them is synthesizing a training dataset with "ground-truth" labels generated by computational models and then applying the trained machine learning model directly to experimental data. However, these methods suffered performance degradation on experimental data, even if they achieved near-optimal performance on synthetic data. In this paper, we identify this phenomenon as the out-of-distribution (OOD) problem, a fundamental challenge in machine learning (Sapoval et al., 2022; Shen et al., 2021), which is caused by the distributional discrepancies between synthetic data and experimental data. The OOD problem can lead to severely biased results and remains unsolved in inferring neural properties from experimental data.

The OOD problem is essential for machine learning techniques in real-world applications. For example, a self-driving car may misidentify the moon as a yellow traffic light and slow down on an expressway because the moon is rarely represented in the training data distribution (Filos et al., 2020), or a machine learning model for clinical diagnoses trained on the dataset from one hospital could suffer over 60% performance loss when applied to patients from another hospital because of the distributional shifts between the two datasets (Cohen et al., 2020). Likewise, machine learning models trained on synthetic data generated from computational models will produce biased results of experimental data, as a result of substantial distributional discrepancies (**Fig. 1a**). For example, although the single-compartment Hodgkin-Huxley (HH) model (Pospischil et al., 2008) can reproduce the dynamics of a wide range of cortical neurons, the distributions of typical biophysical patterns, such as firing rates and action potential amplitudes, still differ dramatically from intracellular recordings in the Allen Cell Types Database (Allen Cell Types Database, 2015) (**Fig. S1a**). Furthermore, the diversity of experimental techniques makes it even harder to deal with the OOD problem in a unified way.

Here, we propose a general domain-adaptive neural inference framework to systematically handle the OOD issue in inferring neural properties from a wide range of experimental data and at various observational scales (**Fig. 1a**). The key solution of our framework is to align the distributions of synthetic data and experimental data through *self-training* and *domain adaptation* (M. Wang & Deng, 2018; Xie et al., 2020) (**Fig. 1a**). Once the distributions of features of synthetic data and experimental data are aligned and similar, the high performance of the machine learning models trained on synthetic data can be well generalized to experimental data. In principle, our framework is compatible with multiple machine learning architectures. In this study, we base our framework on a deep neural network (DNN), while DNN alone shows poor performance on experimental data.

In this study, we demonstrated the framework's applicability in two fundamental scenarios of inferring neural properties (**Fig. 1b & c**). The first scenario is to infer detailed physiological properties at the neuron and microcircuit scales. At the neuron scale, our framework simultaneously inferred ion channel densities and membrane properties of 574 neurons from the intracellular recordings in 14 brain regions of mice in the Allen Cell Types Database (Allen Cell Types Database, 2015). Similarly, our framework accurately inferred 31 cellular and synaptic properties of a microcircuit from the extracellular recordings in the stomatogastric ganglion of *Cancer Borealis* (Haddad & Marder, 2018). The second scenario is to infer “all-or-none” (1-or-0) monosynaptic connections from *in vivo* multi-electrode extracellular recordings in a hippocampal

CA1 region of freely running mice (English et al., 2017). In both scenarios, our framework outperforms a variety of state-of-the-art (SOTA) methods in terms of efficiency and accuracy. All the performance evaluations are made on experimental data instead of synthetic data to guarantee the true applicability of our framework.

To summarize, our DDN-based framework can accurately and efficiently infer a wide range of neural properties from diverse kinds of experimental data with little manual interference. Additionally, our domain-adaptive framework paves the path for OOD generalization from synthetic data to various unlabeled experimental data. Furthermore, our framework could potentially inspire next-generation methods in biosciences that use computational models to study real-world applications (de Melo et al., 2022), such as behavioral analysis (Bolanos et al., 2021), protein function prediction (Wan & Jones, 2020), and medicine and healthcare (R. J. Chen et al., 2021).

Results

Domain adaptive neural inference framework

Modern machine learning 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 general domain adaptive neural inference framework to infer neural properties from various "unlabeled" experimental data (**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 machine learning 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, machine learning models trained solely on labeled synthetic data will yield poor performance on experimental data, known as the OOD problem (Sapoval et al., 2022; Shen et al., 2021). Therefore, we introduce domain adaptation and self-training to automatically assimilate the distributions of the two domains, which guarantees high performance on experimental data (**Fig. 1a**).

The pipeline of our framework is composed of four key steps (**Fig. 1a**): (1) we need to build up a computational model to describe the mechanisms of the neural properties underlying unlabeled experimental data (x_{exp}) via some model parameters (y_{syn}); (2) a labeled synthetic dataset (data: x_{syn}; labels: y_{syn}) will be generated by simulating the computational model with randomly sampled model parameters (i.e., the labels); (3) self-training and domain adaptation will train the machine learning model to align the discrepant distributions of synthetic data and experimental data at the feature space, and hence the OOD problem can be alleviated; (4) after being trained on the two aligned datasets, the machine learning model can accurately output the inferred neural properties from unlabeled experimental data.

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 of our framework, the core of our framework to tackle the OOD problem, to cover two essential inference scenarios: inferring biophysical properties and inferring monosynaptic connectivity. Firstly, 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-none") 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**).

Finally, we implement our framework based on the ResNet-LSTM-FC architecture because of its outstanding performance on synthetic data (Sheng et al., 2021). 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 (Sheng et al., 2021). 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):

$$L = L_{ST} + L_{DA} \tag{1}$$

In each training step, we use backpropagation (Rumelhart et al., 1986) 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):

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

where $y_{pseudo}\left(x_{exp}^{(i)}\right)$ is the pseudo label of experimental datum $x_{exp}^{(i)}$. 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 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):

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

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

where λ_{dis} is the scaling factor of the discrepancy loss, and $\mathcal{L}$ is the loss function (e.g., L2 loss or cross-entropy loss; Methods). 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 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 a unified framework in a set of applications, ranging from inferring single-neuron biophysics to inferring network connectivity, and compare the framework's performance with SOTA machine learning methods in each application.

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

The neocortex exhibits diverse neuronal types and cortical functions are governed by the interactions between different types of neurons (Tremblay et al., 2016; Zeng & Sanes, 2017). Inferring neuronal properties of different neuronal 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 (Gouwens et al., 2018; Pospischil et al., 2008), 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 (Pospischil et al., 2008) (**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, which 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 (Ganin et al., 2016) 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):

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

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

where $SP(D_{syn}, x_{exp})$ is the similar pool of experimental data x_{exp} in the synthetic dataset, and $\left|SP(D_{syn}, x_{exp})\right|$ 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 (Ben-David et al., 2010; Ben-David et al., 2006).

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 (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, the similarity metric is defined 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 neuronal 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 patterns and subthreshold patterns, the two major distinguishable patterns across different cell types, are well predicted, which indicates that our inference results can potentially facilitate the study of subcellular differences of different cell 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 (Gouwens et al., 2018). 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 have been widely used in the inference of single neuronal properties, so we compared the inference accuracy and efficiency between our framework and a typical ES method (Gonçalves et al., 2020) (Methods). The inference efficiency is defined as the number of synthetic data needed to achieve a given number of good inferences, and a 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 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 the same number of good inferences with around 35 times 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 (Marder & Bucher, 2007; Schulz et al., 2006), 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* (Haddad & Marder, 2018). 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 based on simulation and outperforms ES methods (Gonçalves et al., 2020). So, we compared the inference performance of our framework with the SNPE-C method (Gonçalves et al., 2020) (**Fig. 5e**; Methods). Our framework has 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 (Haddad & Marder, 2018). Thus, inferences that can accurately describe these patterns may reveal the modulation effect on the subcellular properties of a microcircuit.

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 (Briggman et al., 2011; Buzsaki, 2004; Ding et al., 2016; Jarrell et al., 2012; Steinmetz et al., 2021; Stringer et al., 2019). One crucial step in studying the functionality of neural networks is to reconstruct their connectomic map since connectivity patterns underlie function (Brunel, 2016). Here, we apply our framework to infer ('all-or-none') 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) (Ostojic et al., 2009). CCGs are a series of time-lagged correlations of two spike trains, which are widely

used to characterize the patterns of neural connectivity (Aertsen & Gerstein, 1985; Endo et al., 2021; English et al., 2017; Kobayashi et al., 2019). 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 (Endo et al., 2021) (**Fig. 6a**), which can reproduce spike timing features similar to the real biological network in cortex and hippocampus (Kobayashi et al., 2009; Omura et al., 2015) (Methods).

Unlike previous tasks of inferring biophysical properties of single neurons or microcircuits, inferring "all-or-none" 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 guarantee 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 (Kang et al., 2019), given by equation (7)-(8):

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

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

where, $\widehat{D}^{cc'}\left(f(D_{syn}^{c}), f(D_{exp}^{c})\right)$ $(c, c' \in \{0,1\})$ 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 data and experimental data 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 (Van der Maaten & Hinton, 2008) 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 (Endo et al., 2021) and GLMCC (Kobayashi et al., 2019) (Methods). The experimental data were collected by simultaneous extra- and juxta-cellular recordings in the CA1 region of freely running mice (English et al., 2017). Some of them have experimentally validated ground-truth labels, which are used to evaluate the inference accuracy (Methods). Our framework outperforms the other two methods quantified with the Matthews correlation coefficient (MCC) (our framework: 0.988, SD=0.023; GLMCC: 0.621, CoNNECT: 0.579) and AUC (our framework: 1.0, SD=0; GLMCC: 0.852; CoNNECT: 0.897) (**Fig. 6d, e & Fig. S5a**). 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 (English et al., 2017). Secondly, we tested if the size of the synthetic network affects the inference accuracy of our framework. 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

Building computational models of hidden neural properties underlying the observed neural activities is crucial to uncovering the functional mechanisms of the neural system (Gidon et al., 2020; Grillner, 2003; Harris & Mrsic-Flogel, 2013; Hodgkin & Huxley, 1952; Litwin-Kumar et al., 2017). A key step is to infer the underlying neural properties so that the model is consistent with experimental observations. Some machine learning methods try to solve this problem by training the models on labeled synthetic data and are then directly applied to experimental data because there is a lack of ground-truth labels of experimental data. In this paper, 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 tackle this challenge, we proposed the general domain adaptive neural inference framework, two key components of which are domain adaptation and self-training. Our framework successfully inferred the biophysical properties of 574 neurons in 14 brain regions of mice simultaneously, 31 biophysical properties of stomatogastric ganglion circuit of *Cancer Borealis*, and monosynaptic connectivity of CA1 networks of mice.

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. This kind of inference is sometimes termed as simulation-based inference. Typical methods developed for the simulation-based inference range from evolutionary search (ES) algorithms (Druckmann et al., 2007; Gouwens et al., 2018; Gurkiewicz & Korngreen, 2007; Prinz et al., 2003), probabilistic modelling (Gonçalves et al., 2020; Oesterle et al., 2020; Schröder et al., 2019) to deep learning approaches (Ben-Shalom et al., 2019; Sheng et al., 2021). 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 are different 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 (Abbott, 2008). One future improvement of our framework is to infer multiple feasible sets of neural properties that are aligned with the experimental data (Alonso & Marder, 2019; Gonçalves et al., 2020; Gutierrez et al., 2013; Marder & Goaillard, 2006; Marder et al., 2015; Metzen & Chacron, 2017; Prinz et al., 2004). This

could potentially be achieved by combing the temporal feature extractor, domain adaptation and self-training of our framework with SNPE methods (Gonçalves et al., 2020).

Inferring monosynaptic connectivity is a binary classification task, so each class forms a sub-domain. Experimental and synthetic data are clustered in one sub-domain based on their distance between each other. We evaluated our performance on experimental data with experimentally validated ground-truth labels (English et al., 2017), which have relatively low noise levels (**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 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 convincing results on some CCGs as its predictions may be unstable in different repetitions (the right example in **Fig. S6b**). It is hard to tell by eye which method is correct, and one simple way to test and validate these methods is to increase the recording time so the CCGs will become less noisy, and the predictions will be more reliable (Bartho et al., 2004; Fujisawa et al., 2008) (**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 (He et al., 2016). Residual Blocks (ResBlocks) are the basic components in ResNet-50 and other variants of ResNet. Each ResBlock is composed of three units, where each unit contains 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 (Wu & He, 2018). 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 (Allen Cell Types Database, 2015). 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 (Elkan, 2003) 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).

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

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

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) (Gonçalves et al., 2020; Prinz et al., 2003; Prinz et al., 2004).

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

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

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

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

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

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: $\bar{g}_{glutamate}, \bar{g}_{cholinergic}$. 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. (Gonçalves et al., 2020).

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 (Endo et al., 2021). 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.

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

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

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

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

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. (Endo et al., 2021).

Training configurations to infer biophysical properties

The batch size is 50. The initial learning rate is 0.0001 and adapted by Adam optimizer (Kingma & Ba, 2014) 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. (Gouwens et al., 2018). 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):

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

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. (Gouwens et al., 2018), and the smallest tolerance of the existence of spontaneous firing is set to be 100 (**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 (Allen Cell Types Database, 2015). 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. 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. (Gonçalves et al., 2020; Prinz et al., 2004): 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 S1**, 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. (Haddad & Marder, 2018) 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. The configurations of SNPE-C are set to be the same as in Gonçalves et al. (Gonçalves et al., 2020). The density estimator is masked autoregressive flow (Papamakarios et al., 2017) with 5 masked autoencoders for distribution estimation (Germain et al., 2015) with [200, 400] hidden units each. The implementation of SNPE-C can be found at https://github.com/mackelab/delfi/.

Ground-truth of synaptic connections *in vivo*

The experimental data are juxtacellular-extracellular recordings from the hippocampal CA1 region in awake mice (English et al., 2017). 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. For all CCGs in both datasets, they are calculated from spontaneous spikes, which meant 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. For the analysis of the presence of the synaptic connections, two criteria are considered.

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.

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

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

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 (Kingma & Ba, 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We implemented domain adaptation as the contrastive adaptation network (Ganin et al., 2016; Kang et al., 2019). The empirical kernel mean embeddings of CDD are given by equations (23)-(31):

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

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

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

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

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

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

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

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

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

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).

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

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

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. (Kobayashi et al., 2019) and we used the published code at https://github.com/NII-Kobayashi. We set the membrane time constant as 5 ms, synaptic delay as 1 ms, and γ as 5×10^{-4}, which are the same as in the original study (Kobayashi et al., 2019). 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. We use the pre-trained CoNNECT model obtained from https://github.com/shigerushinomoto/CoNNECT.git 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 machine learning algorithms (Bradley, 1997). 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) (Matthews, 1975). 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 (Chicco & Jurman, 2020).

Code Availability

The source code of the domain adaptive neural inference framework will be available upon acceptance.

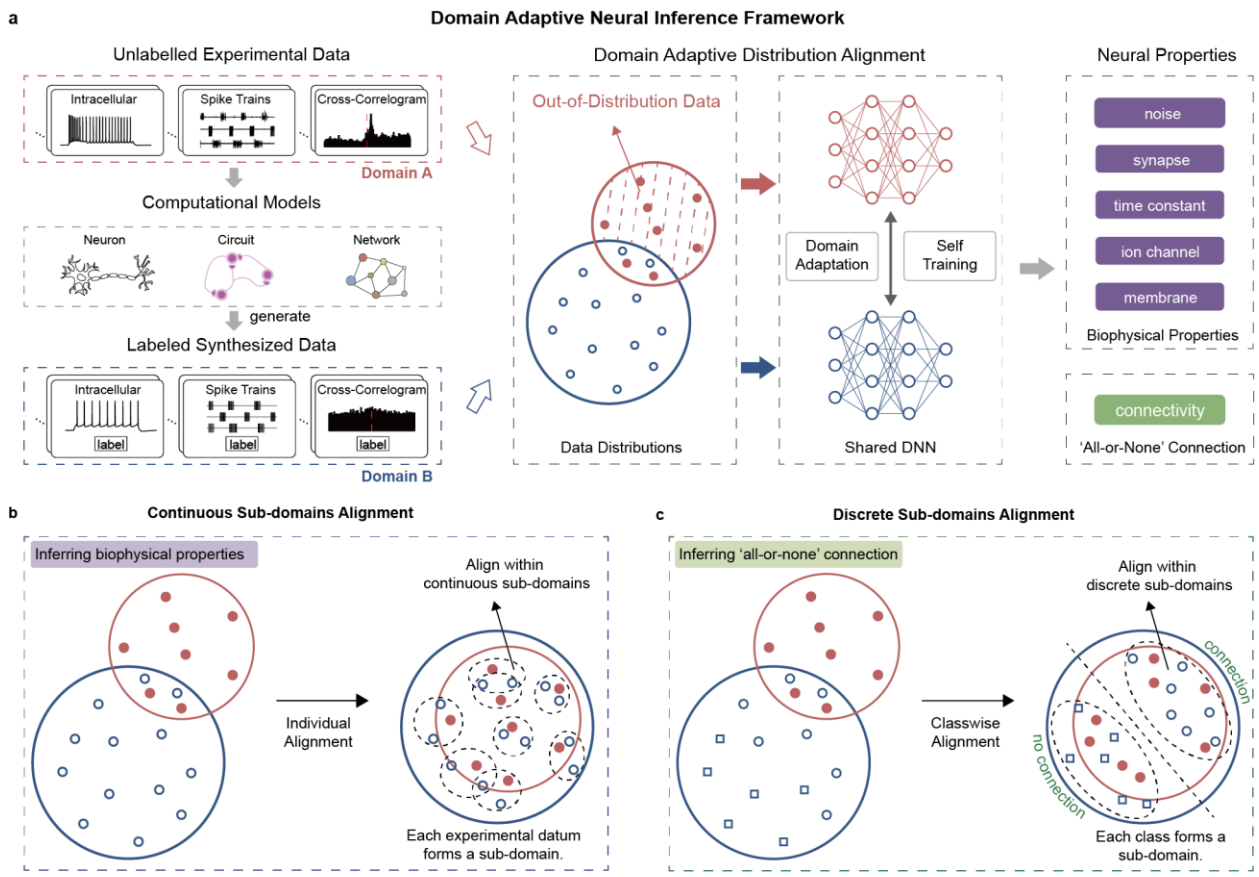
Acknowledgment

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

Competing Interests

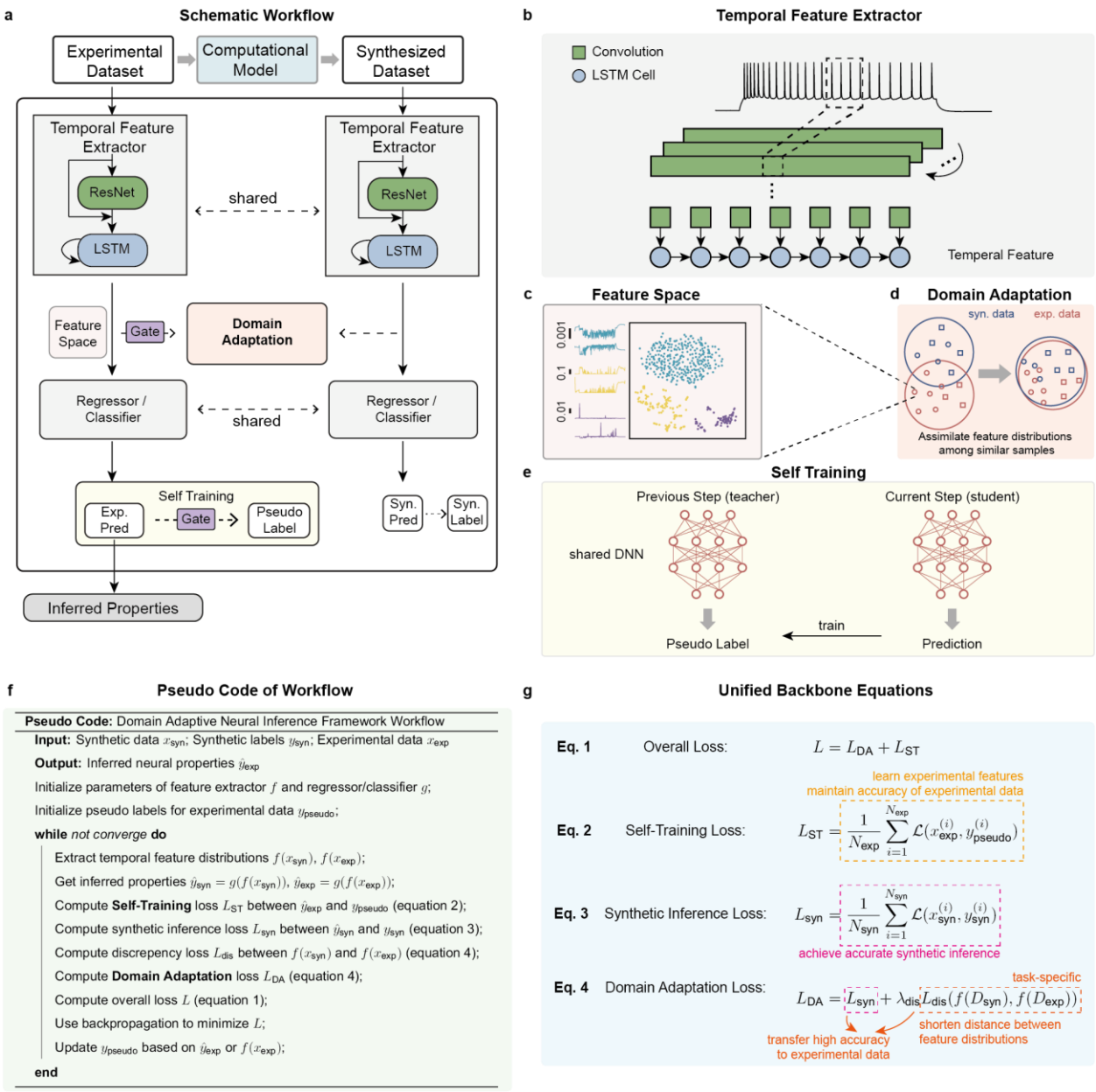
The authors declare no competing interests.

Fig. 1: Neural Properties Inference Methodology.



a, Overview of the inference framework to infer neural properties from various experimental data based on computational models. **b,** The instantiation of our framework on inferring biophysical properties. **c,** The instantiation of our framework on inferring the existence of connectivity.

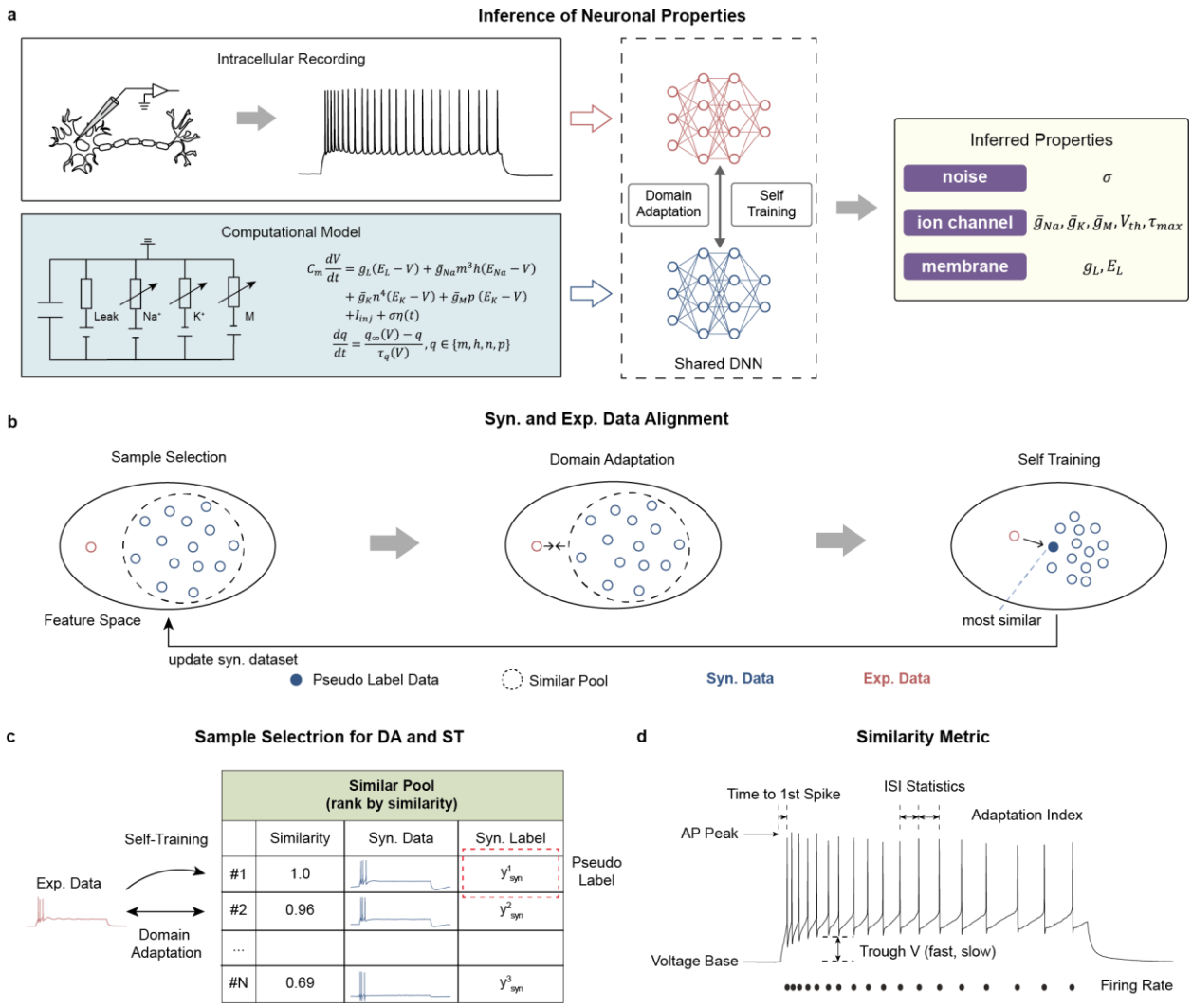
Fig. 2: DNN-embedded framework workflow.



a, Workflow of the DNN implementation of the domain adaptive neural inference framework. The temporal feature extractor and regressor or classifier are shared between synthetic data and experimental data. Firstly, temporal features of synthetic data ($f(D_{syn}) = \{f(x_{syn}) | (x_{syn}, y_{syn}) \in D_{syn}\}$) and experimental data ($f(D_{exp}) = \{f(x_{exp}) | x_{exp} \in D_{exp}\}$) 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

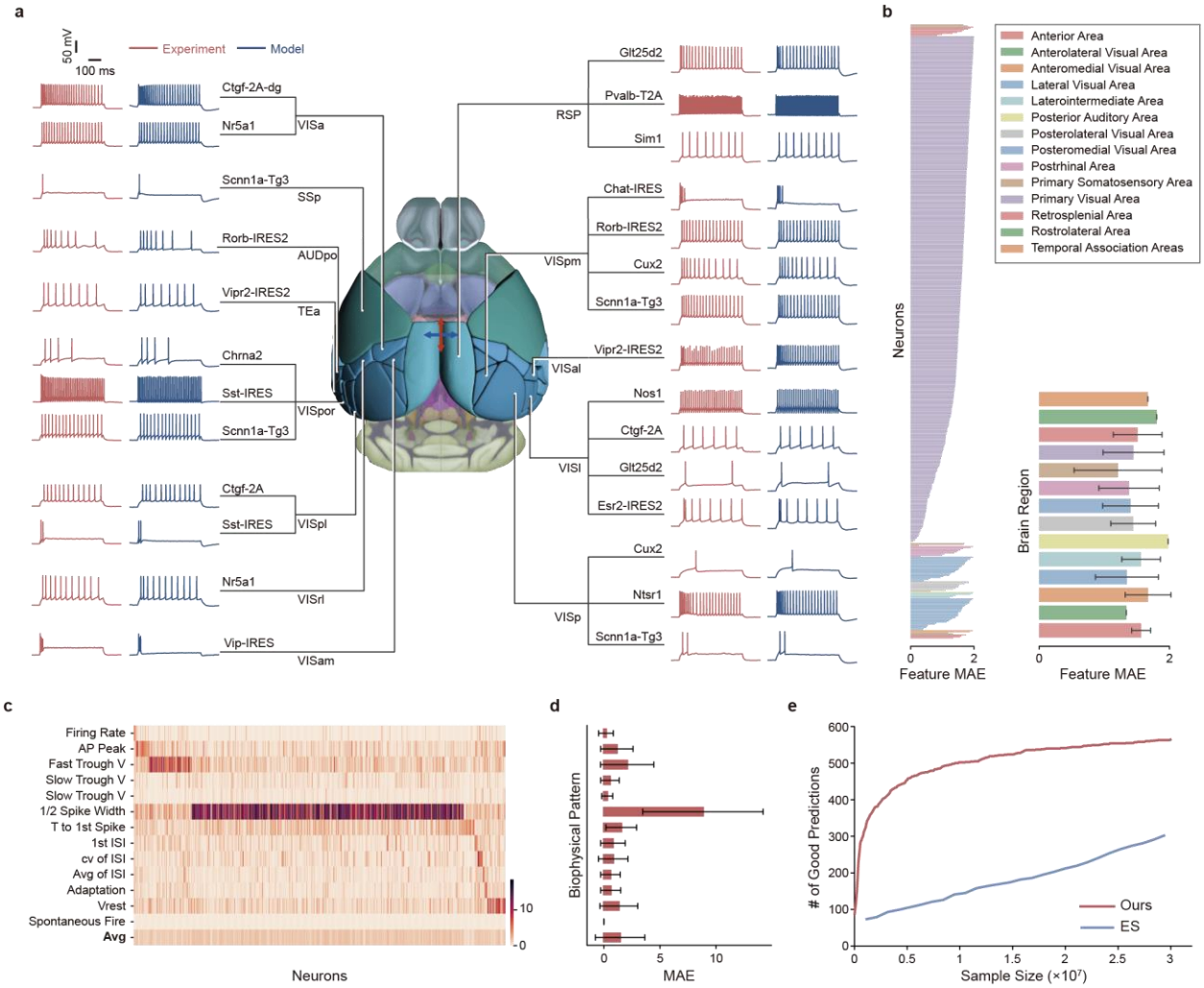
and an LSTM layer (Sheng et al., 2021). **c,** Feature visualization of the experimental datum in **b**. Each dot is a dimension of the features. Each color denotes a cluster and representatives of each cluster are shown on the left (for visualizations of all features, see **Fig. S2**). **d,** Schematic mechanisms of domain adaptation. **e,** Illustration of self-training. The pseudo label is the prediction from the DNN in a previous training step. **f,** Pseudo code of the domain adaptive neural inference framework. **g,** Unified backbone equations used in **f** describing the definition of the loss functions of the framework. 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.



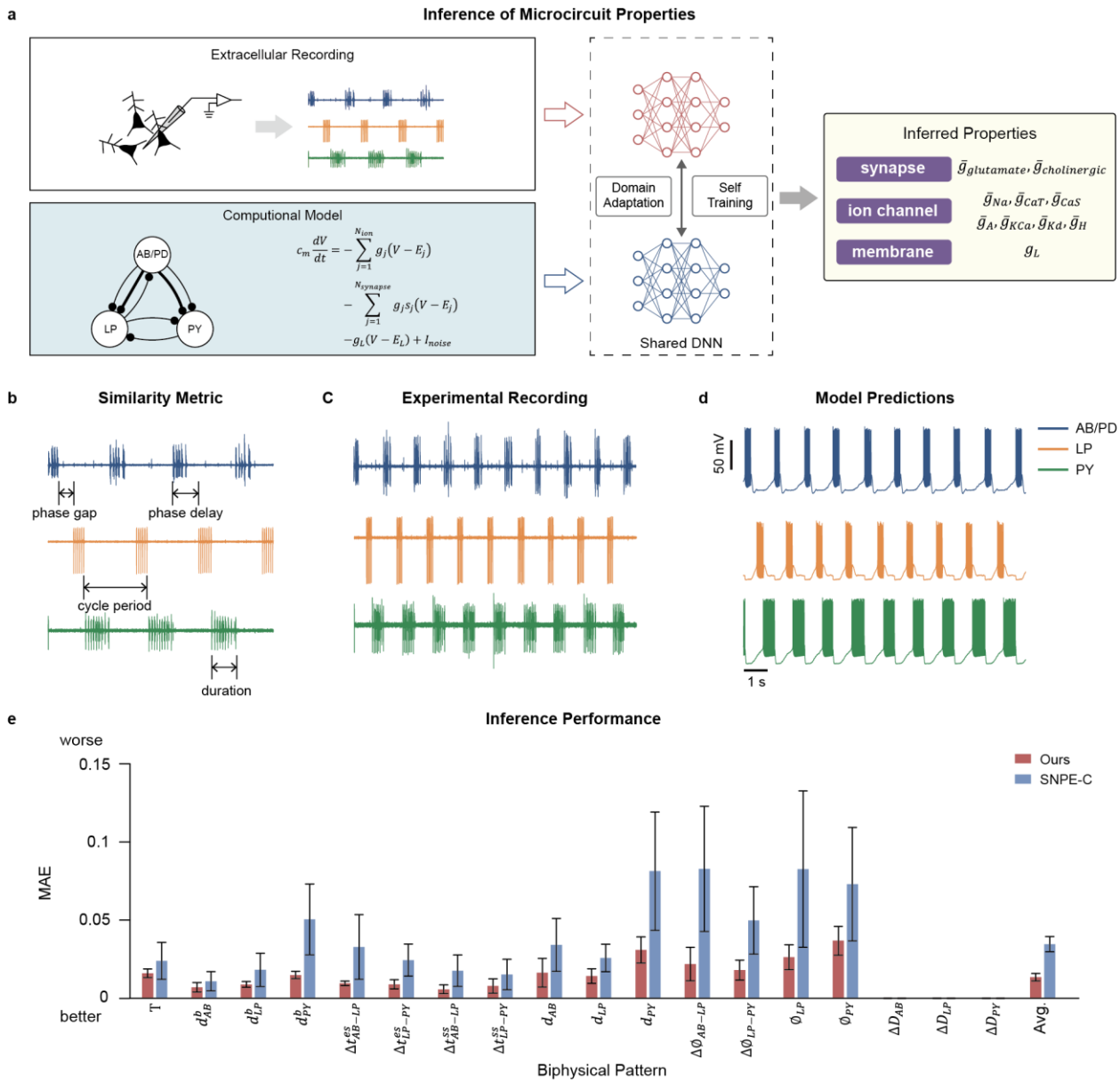
a, The experimental data, computational model, and the biophysical properties to be inferred at the neuron scale. **b,** Illustration of the instantiation of the domain adaptive neural inference framework to infer biophysical properties. Red symbols represent features of experimental data (exp. data) 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**.

Fig. 4: Inferring biophysical properties of neurons of mice.



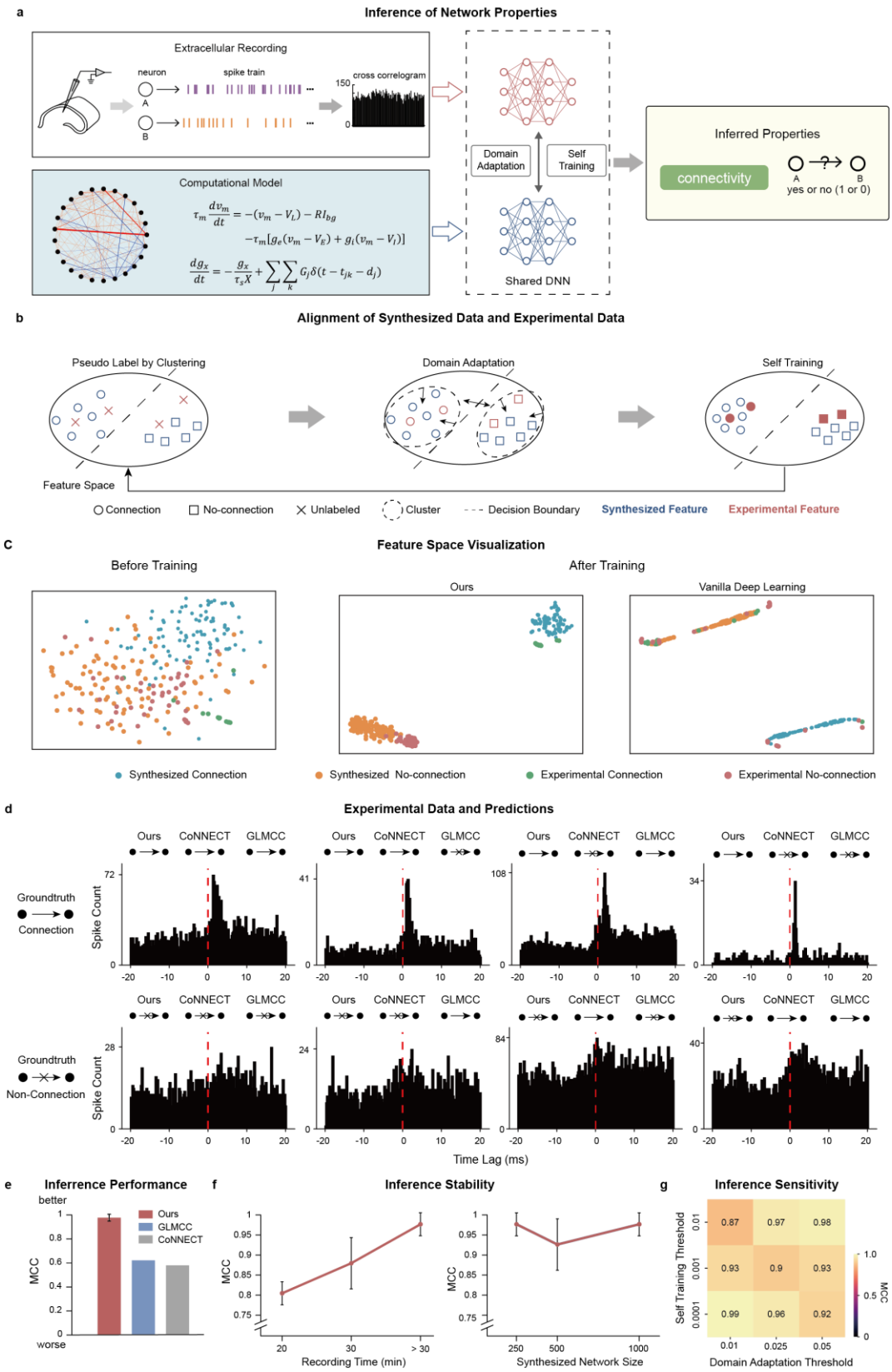
a, Examples of inferred results grouped by brain regions and neuron types. The brain region atlas is directly obtained from Lau et al. (Lau et al., 2008). Blue: the model responses are generated by applying the same stimulus protocol with the inferred properties. Red: the experimental data. **b,** Mean 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**. **e,** Sample efficiency compared to the evolutionary search (ES) method. 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. 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 (Gonçalves et al., 2020) of MAE of biophysical patterns. The discrepancy loss factor is 0.001 and the self-training threshold is 0.1.

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 domain adaptive neural 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,** Examples of cross-correlogram of the experimental data and their ground-truth labels and results of inferred connectivity of our framework, CoNNECT (Endo et al., 2021), and GLMCC (Kobayashi et al., 2019) are presented. The time bin is 0.2 ms. **e,** The inference accuracy compared among the three methods. **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.

References

Abbott, L. F. (2008). Theoretical neuroscience rising. *Neuron, 60*(3), 489-495. doi:10.1016/j.neuron.2008.10.019

Aertsen, A. M., & Gerstein, G. L. (1985). Evaluation of neuronal connectivity: sensitivity of cross-correlation. *Brain Res, 340*(2), 341-354. doi:10.1016/0006-8993(85)90931-x

Allen Cell Types Database. (2015).

Alonso, L. M., & Marder, E. (2019). Visualization of currents in neural models with similar behavior and different conductance densities. *Elife, 8*. doi:10.7554/eLife.42722

Bartho, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., & Buzsaki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol, 92*(1), 600-608. doi:10.1152/jn.01170.2003

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Mach Learn, 79*(1), 151-175. doi:10.1007/s10994-009-5152-4

Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of representations for domain adaptation. *Advances in neural information processing systems, 19*, 137–144.

Ben-Shalom, R., Balewski, J., Siththaranjan, A., Baratham, V., Kyoung, H., Kim, K. G., Bender, K. J., & Bouchard, K. E. (2019). Inferring neuronal ionic conductances from membrane potentials using CNNs. *bioRxiv*, 727974. doi:10.1101/727974

Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., Abbasi-Asl, R., Jia, X., Siegle, J. H., Olsen, S. R., Koch, C., Mihalas, S., & Arkhipov, A. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron, 106*(3), 388-403 e318. doi:10.1016/j.neuron.2020.01.040

Bolanos, L. A., Xiao, D., Ford, N. L., LeDue, J. M., Gupta, P. K., Doebeli, C., Hu, H., Rhodin, H., & Murphy, T. H. (2021). A three-dimensional virtual mouse generates synthetic training data for behavioral analysis. *Nat Methods, 18*(4), 378-381. doi:10.1038/s41592-021-01103-9

Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., & Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci, 8*(9), 1263-1268. doi:10.1038/nn1525

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit, 30*(7), 1145-1159. doi:10.1016/S0031-3203(96)00142-2

Briggman, K. L., Helmstaedter, M., & Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature, 471*(7337), 183-188. doi:10.1038/nature09818

Brunel, N. (2016). Is cortical connectivity optimized for storing information? *Nat Neurosci, 19*(5), 749-755. doi:10.1038/nn.4286

Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. *Nat Neurosci, 7*(5), 446-451. doi:10.1038/nn1233

Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K. F., Sandberg, R., & Tolias, A. S. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol, 34*(2), 199-203. doi:10.1038/nbt.3445

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng, 5*(6), 493-497. doi:10.1038/s41551-021-00751-8

Chen, T. W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., & Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature, 499*(7458), 295-300. doi:10.1038/nature12354

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

Cohen, J. P., Hashir, M., Brooks, R., & Bertrand, H. (2020). On the limits of cross-domain generalization in automated X-ray prediction. *Proceedings of the Third Conference on Medical Imaging with Deep Learning, 121*, 136--155.

de Melo, C. M., Torralba, A., Guibas, L., DiCarlo, J., Chellappa, R., & Hodgins, J. (2022). Next-generation deep learning based on simulators and synthetic data. *Trends Cogn Sci, 26*(2), 174-187. doi:10.1016/j.tics.2021.11.008

Denk, W., Delaney, K. R., Gelperin, A., Kleinfeld, D., Strowbridge, B. W., Tank, D. W., & Yuste, R. (1994). Anatomical and functional imaging of neurons using 2-photon laser scanning microscopy. *J Neurosci Methods, 54*(2), 151-162. doi:10.1016/0165-0270(94)90189-9

Denk, W., Strickler, J. H., & Webb, W. W. (1990). Two-photon laser scanning fluorescence microscopy. *Science, 248*(4951), 73-76. doi:10.1126/science.2321027

Ding, H., Smith, R. G., Poleg-Polsky, A., Diamond, J. S., & Briggman, K. L. (2016). Species-specific wiring for direction selectivity in the mammalian retina. *Nature, 535*(7610), 105-110. doi:10.1038/nature18609

Druckmann, S., Banitt, Y., Gidon, A., Schurmann, F., Markram, H., & Segev, I. (2007). A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front Neurosci, 1*(1), 7-18. doi:10.3389/neuro.01.1.1.001.2007

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

Endo, D., Kobayashi, R., Bartolo, R., Averbeck, B. B., Sugase-Miyamoto, Y., Hayashi, K., Kawano, K., Richmond, B. J., & Shinomoto, S. (2021). A convolutional neural network for estimating synaptic connectivity from spike trains. *Sci Rep, 11*(1), 12087. doi:10.1038/s41598-021-91244-w

English, D. F., McKenzie, S., Evans, T., Kim, K., Yoon, E., & Buzsaki, G. (2017). Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks. *Neuron, 96*(2), 505-520 e507. doi:10.1016/j.neuron.2017.09.033

Filos, A., Tigkas, P., McAllister, R., Rhinehart, N., Levine, S., & Gal, Y. (2020). Can autonomous vehicles identify, recover from, and adapt to distribution shifts? *International Conference on Machine Learning*, 3145-3153.

Fujisawa, S., Amarasingham, A., Harrison, M. T., & Buzsaki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci, 11*(7), 823-833. doi:10.1038/nn.2134

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *J. Mach. Learn. Res., 17*(1), 2096–2030.

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

Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., Holtkamp, M., Vida, I., & Larkum, M. E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science, 367*(6473), 83-87. doi:10.1126/science.aax6239

Gonçalves, P. J., Lueckmann, J. M., Deistler, M., Nonnenmacher, M., Ocal, K., Bassetto, G., Chintaluri, C., Podlaski, W. F., Haddad, S. A., Vogels, T. P., Greenberg, D. S., & Macke, J. H. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife, 9*. doi:10.7554/eLife.56261

Gouwens, N. W., Berg, J., Feng, D., Sorensen, S. A., Zeng, H., Hawrylycz, M. J., Koch, C., & Arkhipov, A. (2018). Systematic generation of biophysically detailed models for diverse cortical neuron types. *Nat Commun, 9*(1), 710. doi:10.1038/s41467-017-02718-3

Grienberger, C., & Konnerth, A. (2012). Imaging calcium in neurons. *Neuron, 73*(5), 862-885. doi:10.1016/j.neuron.2012.02.011

Grillner, S. (1985). Neurobiological bases of rhythmic motor acts in vertebrates. *Science, 228*(4696), 143-149. doi:10.1126/science.3975635

Grillner, S. (2003). The motor infrastructure: from ion channels to neuronal networks. *Nat Rev Neurosci, 4*(7), 573-586. doi:10.1038/nrn1137

Gurkiewicz, M., & Korngreen, A. (2007). A numerical approach to ion channel modelling using whole-cell voltage-clamp recordings and a genetic algorithm. *PLoS Comput Biol, 3*(8), e169. doi:10.1371/journal.pcbi.0030169

Gutierrez, G. J., O'Leary, T., & Marder, E. (2013). Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron, 77*(5), 845-858. doi:10.1016/j.neuron.2013.01.016

Haddad, S. A., & Marder, E. (2018). Circuit robustness to temperature perturbation is altered by neuromodulators. *Neuron, 100*(3), 609-623 e603. doi:10.1016/j.neuron.2018.08.035

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature, 503*(7474), 51-58. doi:10.1038/nature12654

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. doi:10.1109/CVPR.2016.90

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol, 117*(4), 500-544. doi:10.1113/jphysiol.1952.sp004764

Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., & Emmons, S. W. (2012). The connectome of a decision-making neural network. *Science, 337*(6093), 437-444. doi:10.1126/science.1221762

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. doi:10.48550/arXiv.1412.6980

Kobayashi, R., Kurita, S., Kurth, A., Kitano, K., Mizuseki, K., Diesmann, M., Richmond, B. J., & Shinomoto, S. (2019). Reconstructing neuronal circuitry from parallel spike trains. *Nat Commun, 10*(1), 4468. doi:10.1038/s41467-019-12225-2

Kobayashi, R., Tsubo, Y., & Shinomoto, S. (2009). Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Front Comput Neurosci, 3*, 9. doi:10.3389/neuro.10.009.2009

Lau, C., Ng, L., Thompson, C., Pathak, S., Kuan, L., Jones, A., & Hawrylycz, M. (2008). Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics, 9*, 153. doi:10.1186/1471-2105-9-153

Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolinsky, H., & Abbott, L. F. (2017). Optimal degrees of synaptic connectivity. *Neuron, 93*(5), 1153-1164 e1157. doi:10.1016/j.neuron.2017.01.030

Marder, E., & Bucher, D. (2007). Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. *Annu Rev Physiol, 69*, 291-316. doi:10.1146/annurev.physiol.69.031905.161516

Marder, E., & Goaillard, J. M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci, 7*(7), 563-574. doi:10.1038/nrn1949

Marder, E., Goeritz, M. L., & Otopalik, A. G. (2015). Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Curr Opin Neurobiol, 31*, 156-163. doi:10.1016/j.conb.2014.10.012

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Kahou, G. A., Berger, T. K., Bilgili, A., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, J. D., Delalondre, F., Delattre, V., Druckmann, S., Dumusc, R., Dynes, J., Eilemann, S., Gal, E., Gevaert, M. E., Ghobril, J. P., Gidon, A., Graham, J. W., Gupta, A., Haenel, V., Hay, E., Heinis, T., Hernando, J. B., Hines, M., Kanari, L., Keller, D., Kenyon, J., Khazen, G., Kim, Y., King, J. G., Kisvarday, Z., Kumbhar, P., Lasserre, S., Le Be, J. V., Magalhaes, B. R., Merchan-Perez, A., Meystre, J., Morrice, B. R., Muller, J., Munoz-Cespedes, A., Muralidhar, S., Muthurasa, K., Nachbaur, D., Newton, T. H., Nolte, M., Ovcharenko, A., Palacios, J., Pastor, L., Perin, R., Ranjan, R., Riachi, I., Rodriguez, J. R., Riquelme, J. L., Rossert, C., Sfyrakis, K., Shi, Y., Shillcock, J. C., Silberberg, G., Silva, R., Tauheed, F., Telefont, M., Toledo-Rodriguez, M., Trankler, T., Van Geit, W., Diaz, J. V., Walker, R., Wang, Y., Zaninetta, S. M., DeFelipe, J., Hill, S. L., Segev, I., & Schurmann, F. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell, 163*(2), 456-492. doi:10.1016/j.cell.2015.09.029

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure, 405*(2), 442-451.

Metzen, M. G., & Chacron, M. J. (2017). Stimulus background influences phase invariant coding by correlated neural activity. *Elife, 6*. doi:10.7554/eLife.24482

Oesterle, J., Behrens, C., Schroder, C., Hermann, T., Euler, T., Franke, K., Smith, R. G., Zeck, G., & Berens, P. (2020). Bayesian inference for biophysical neuron models enables stimulus optimization for retinal neuroprosthetics. *Elife, 9*. doi:10.7554/eLife.54997

Omura, Y., Carvalho, M. M., Inokuchi, K., & Fukai, T. (2015). A lognormal recurrent network model for burst generation during hippocampal sharp waves. *J Neurosci, 35*(43), 14585-14601. doi:10.1523/JNEUROSCI.4944-14.2015

Ostojic, S., Brunel, N., & Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J Neurosci, 29*(33), 10234-10253. doi:10.1523/jneurosci.1275-09.2009

Packer, A. M., Peterka, D. S., Hirtz, J. J., Prakash, R., Deisseroth, K., & Yuste, R. (2012). Two-photon optogenetics of dendritic spines and neural circuits. *Nat Methods, 9*(12), 1202-1205. doi:10.1038/nmeth.2249

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked autoregressive flow for density estimation. *Advances in neural information processing systems, 30*, 2335–2344.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci, 21*, 227-277. doi:10.1146/annurev.neuro.21.1.227

Pospischil, M., Toledo-Rodriguez, M., Monier, C., Piwkowska, Z., Bal, T., Fregnac, Y., Markram, H., & Destexhe, A. (2008). Minimal Hodgkin-Huxley type models for different classes of cortical and thalamic neurons. *Biol Cybern, 99*(4-5), 427-441. doi:10.1007/s00422-008-0263-8

Prakash, R., Yizhar, O., Grewe, B., Ramakrishnan, C., Wang, N., Goshen, I., Packer, A. M., Peterka, D. S., Yuste, R., Schnitzer, M. J., & Deisseroth, K. (2012). Two-photon optogenetic toolbox for fast inhibition, excitation and bistable modulation. *Nat Methods, 9*(12), 1171-1179. doi:10.1038/nmeth.2215

Prinz, A. A., Billimoria, C. P., & Marder, E. (2003). Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. *J Neurophysiol, 90*(6), 3998-4015. doi:10.1152/jn.00641.2003

Prinz, A. A., Bucher, D., & Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat Neurosci, 7*(12), 1345-1352. doi:10.1038/nn1352

Rickgauer, J. P., & Tank, D. W. (2009). Two-photon excitation of channelrhodopsin-2 at saturation. *Proc Natl Acad Sci U S A, 106*(35), 15025-15030. doi:10.1073/pnas.0907084106

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. doi:10.1038/323533a0

Sapoval, N., Aghazadeh, A., Nute, M. G., Antunes, D. A., Balaji, A., Baraniuk, R., Barberan, C. J., Dannenfelser, R., Dun, C., Edrisi, M., Elworth, R. A. L., Kille, B., Kyrillidis, A., Nakhleh, L., Wolfe, C. R., Yan, Z., Yao, V., & Treangen, T. J. (2022). Current progress and open challenges for applying deep learning across the biosciences. *Nat Commun, 13*(1), 1728. doi:10.1038/s41467-022-29268-7

Schröder, C., James, B., Lagnado, L., & Berens, P. (2019). Approximate bayesian inference for a mechanistic model of vesicle release at a ribbon synapse. *Advances in neural information processing systems, 32*, 7070-7080.

Schulz, D. J., Goaillard, J. M., & Marder, E. (2006). Variable channel expression in identified single and electrically coupled neurons in different animals. *Nat Neurosci, 9*(3), 356-362. doi:10.1038/nn1639

Shen, Z., Liu, J., He, Y., Zhang, X., Xu, R., Yu, H., & Cui, P. (2021). Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*. doi:10.48550/arXiv.2108.13624

Sheng, K., Qu, P., Yang, L., Liu, X., He, L., Zhang, Y., Ma, L., & Du, K. (2021). A general LSTM-based deep learning method for estimating neuronal models and inferring neural circuitry. *bioRxiv*, 2021.2003.2014.434027. doi:10.1101/2021.03.14.434027

Shine, J. M., Muller, E. J., Munn, B., Cabral, J., Moran, R. J., & Breakspear, M. (2021). Computational models link cellular mechanisms of neuromodulation to large-

scale neural dynamics. *Nat Neurosci, 24*(6), 765-776. doi:10.1038/s41593-021-00824-6

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Bohm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., Putzeys, J., Sauerbrei, B., van Daal, R. J. J., Vollan, A. Z., Wang, S., Welkenhuysen, M., Ye, Z., Dudman, J. T., Dutta, B., Hantman, A. W., Harris, K. D., Lee, A. K., Moser, E. I., O'Keefe, J., Renart, A., Svoboda, K., Hausser, M., Haesler, S., Carandini, M., & Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science, 372*(6539). doi:10.1126/science.abf4588

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature, 571*(7765), 361-365. doi:10.1038/s41586-019-1346-5

Tian, L., Hires, S. A., Mao, T., Huber, D., Chiappe, M. E., Chalasani, S. H., Petreanu, L., Akerboom, J., McKinney, S. A., Schreiter, E. R., Bargmann, C. I., Jayaraman, V., Svoboda, K., & Looger, L. L. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods, 6*(12), 875-881. doi:10.1038/nmeth.1398

Tremblay, R., Lee, S., & Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron, 91*(2), 260-292. doi:10.1016/j.neuron.2016.06.033

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*(11).

Wan, C., & Jones, D. T. (2020). Protein function prediction is improved by creating synthetic feature samples with generative adversarial networks. *Nat Mach Intell, 2*(9), 540-550. doi:10.1038/s42256-020-0222-1

Wang, M., & Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing, 312*, 135-153. doi:10.1016/j.neucom.2018.05.083

Wang, X. J., & Krystal, J. H. (2014). Computational psychiatry. *Neuron, 84*(3), 638-654. doi:10.1016/j.neuron.2014.10.018

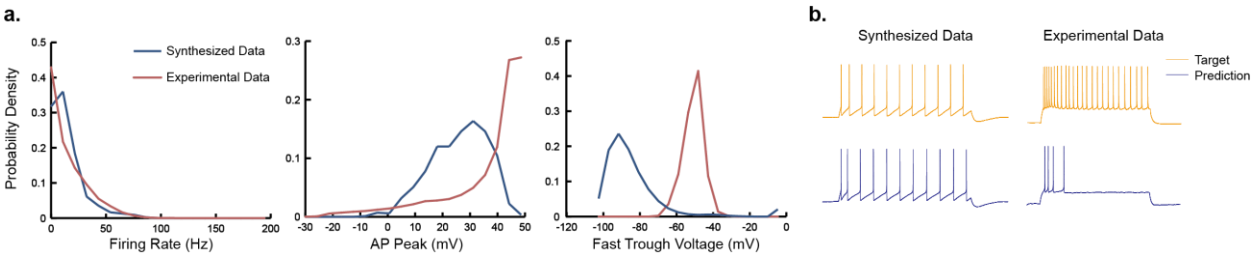
Wu, Y., & He, K. (2018). Group normalization. *Proceedings of the European conference on computer vision (ECCV)*, 3-19. doi:10.1007/978-3-030-01261-8_1

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

Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat Rev Neurosci, 18*(9), 530-546. doi:10.1038/nrn.2017.85

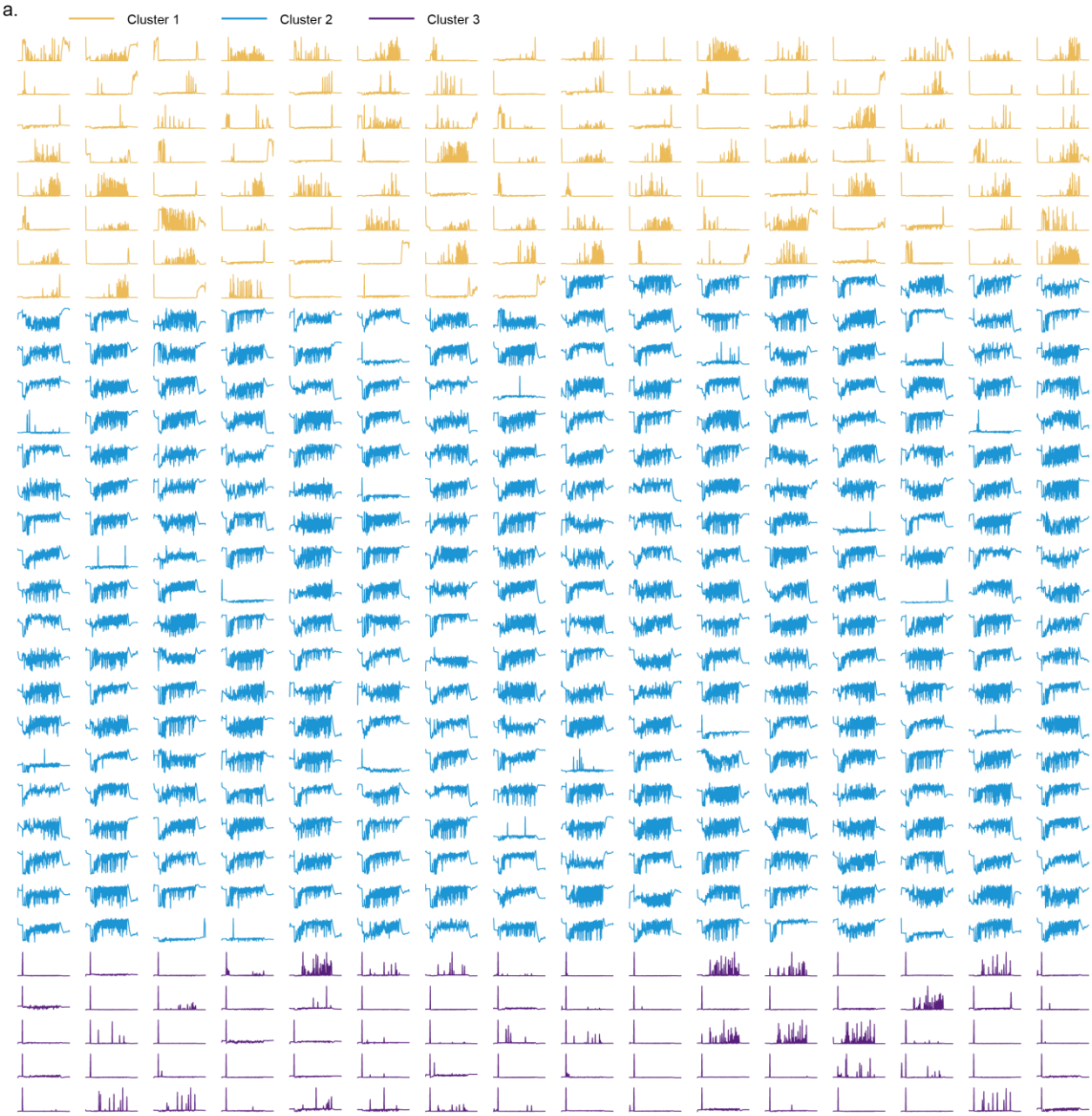
Supplementary Information

Fig. S1: OOD problem in the inference of biophysical properties of single neurons.



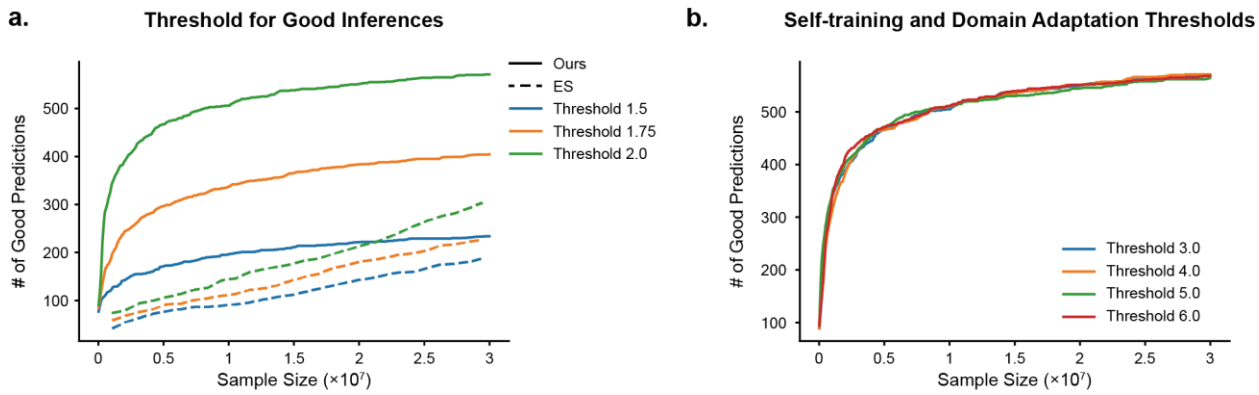
a, Distributions of three biophysical patterns of the synthetic data and experimental data. **b,** The model predictions of a piece of synthetic data and experimental data by the vanilla DNN without domain adaptation and self-training.

Fig. S2: Feature space visualization.



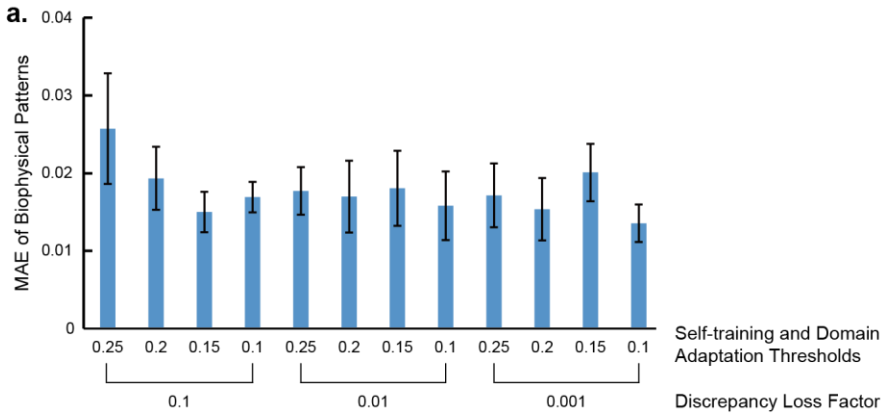
a, Feature visualization of all 512 features as shown in **Fig. 2c.**

Fig. S3: Hyperparameter analysis for inference of biophysical properties of single neurons.



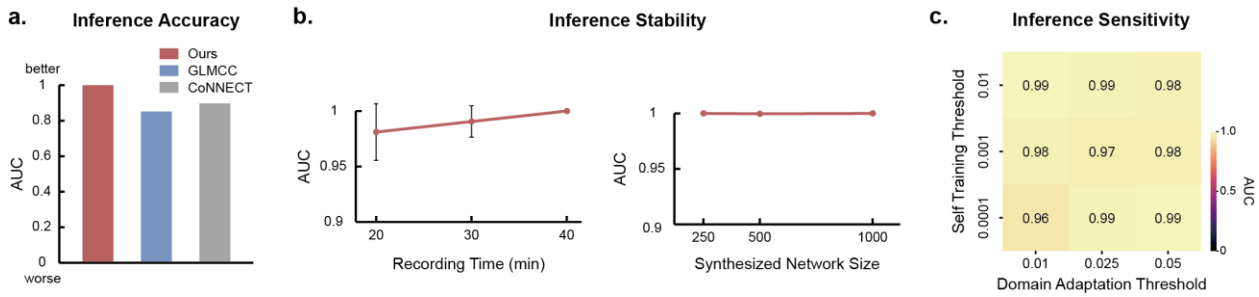
a, Comparison between our framework and the ES method with different thresholds to define good inferences. **b,** Comparison between different self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}). The two thresholds are identical in this task.

Fig. S4: Hyperparameter analysis for inference of properties of the microcircuit.



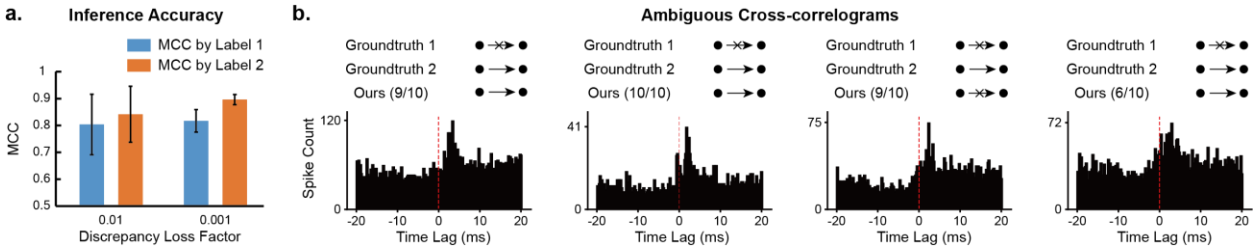
a, Comparison between different self-training threshold (S_{ST}) and domain adaptation threshold (S_{DA}) and discrepancy loss factors (λ_{dis}). Domain adaptation and self-training thresholds are identical in this task. Error bar: standard deviation.

Fig. S5: Performances quantified by the AUC metric for inference of monosynaptic connectivity



a, The inference accuracy comparison between the three methods, quantified by the AUC metric. **b**, The effect of the length of experimental recording time (left) and the effect of the network size on the inference accuracy of our framework. Error bar: standard deviation. **c**, The sensitivity of domain adaptation threshold (S_{DA}) and self-training threshold (S_{ST}) on the inference accuracy of our framework.

Fig. S6: Ambiguous and noisy cross-correlograms.



a. MCC between predictions by our framework and two ground-truth methods. The domain adaptation threshold (S_{DA}) and self-training threshold (S_{ST}) are 0.01 and 0.0001 respectively. Each is repeated 5 times. Error bar: standard deviation. **b**. Examples of ambiguous CCGs and predictions of two ground-truth methods and our framework. The bracket denotes the number of predictions of our framework to be connection or non-connection in 10 repetitions of two sets of hyperparameters in **a**.

Table S1 The smallest tolerances used in the similarity calculation of biophysical patterns of single neurons (adapted from (Gouwens et al., 2018)).

Biophysical Pattern	Smallest Tolerance (σ)
Average firing frequency	0.5 Hz
Latency to the first spike	5 ms
Average interspike interval (ISI)	0.5 ms
Duration of first ISI	1 ms
Coefficient of variation of ISIs	0.1
Adaptation index	0.001
Action potential peak	2 mV
Fast trough depth	2 mV
Slow trough depth	2 mV
Time of slow trough (as a fraction of ISI)	0.05
Action potential width at half-height	0.1 ms
Resting potential	2 mV
Average firing frequency	0.5 Hz
Existence of Spontaneous Firing	100

Table S2 The smallest tolerances used in the similarity calculation of biophysical patterns of the microcircuit.

Biophysical Pattern	Smallest Tolerance (σ)
T	1.06316870e+03
d_{AB}^{b}	7.33209261e+02
d_{LP}^{b}	6.65553032e+02
d_{PY}^{b}	4.59874694e+02
Δt_{AB-LP}^{es}	4.33098114e+02
Δt_{LP-PY}^{es}	7.05637617e+02
Δt_{AB-LP}^{ss}	8.59753455e+02
Δt_{LP-PY}^{ss}	9.09731752e+02
d_{AB}	2.39840167e-01
d_{LP}	3.37006001e-01
d_{PY}	1.95884647e-01
$\Delta\emptyset_{AB-LP}$	1.64340799e-01
$\Delta\emptyset_{LP-PY}$	2.95808090e-01
$\emptyset_{LP}$	1.90751003e-01
$\emptyset_{PY}$	2.05308162e-01
ΔD_{AB}	2.93970444e+01
ΔD_{LP}	5.81104516e+01

ΔD_{PY}	3.55649109e+01