Signal-to-signal networks for improved spike estimation from calcium imaging data

Jilt Sebastian[1,2,3‡], Mriganka Sur [4], Hema A. Murthy[2], and Mathew Magimai.-Doss[1]

1 Idiap Research Institute, Martigny, Valais, Switzerland
2 Department of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Vaud, Switzerland
3 Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India
4 Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology Cambridge, United States

‡Current Address: Speech and Music Technology Lab, Department of Computer Science and Engineering, Indian Institute of Technology Madras, India
* jiltsebastian@gmail.com

Abstract

Spiking information of individual neurons is essential for functional and behavioural analysis in neuroscience. During electrophysiological experiments in animals, calcium imaging techniques are employed to obtain activities of individual neurons and neuronal populations and they result in slowly-varying fluorescence signals with poor temporal resolution. Estimating the temporal positions of action potentials from these signals is a challenging problem. In the literature, a number of generative model based and data-driven algorithms have been studied with limited success. In this article, we propose a neural network based signal-to-signal (S2S) conversion approach, where the neural network takes as input raw-fluorescence signal and learns to predict spike information signal in an end-to-end manner. Theoretically, the proposed approach formulates the problem of spike estimation from the fluorescence signal as a single channel source separation problem with unknown mixing conditions. Through experimental studies on spikefinder bench-marking dataset, we show that the proposed S2S conversion approach outperforms state-of-the-art-methods. We show that the resulting system: (a) has low complexity with respect to existing approaches and is reproducible; (b) is layer-wise interpretable; and (c) has the capability to generalise across different calcium indicators.

Author summary

Information processing by a population of neurons is studied using two-photon calcium imaging techniques. A neuronal spike results in an increased intra-cellular calcium concentration. Fluorescent calcium indicators change their brightness upon a change in the calcium concentration and this change is captured in the imaging technique. The task of estimating the actual spike positions from the brightness variation is referred to as spike estimation. Several signal processing and machine learning-based algorithms have been proposed to solve this problem. However, the task is still far from being solved. Here we present a novel neural network based data-driven algorithm for this task. Our method takes the fluorescence recording as the input and synthesises the spike information signal which is well-correlated with the actual spike positions. Our method outperforms state-of-the-art methods in spike estimation on standard evaluation framework. We analyse different components of the model and discuss its benefits.

Introduction

Analysis of brain circuitry requires understanding the information encoded in the neuronal spikes at micro and macro levels. Latest scanning methods track the activity of a population of neurons by making use of fluorescence emitting capability of calcium indicator proteins/dye [1–4]. The calcium fluorescence recording of each neuron, however, is only an indirect indicator of the actual spiking process. Presence of fluorescence level fluctuations, slow dynamics of calcium fluorescence signal, and unknown noise-levels makes it hard to identify the exact underlying spike information [5–7]. Hence, technologies capable of solving this are of utmost interest to the neuroscience community.

Spike estimation algorithms proposed in the past can be broadly categorised into generative and data-driven approaches:

1. Generative methods model the fluorescence signal as the response of the calcium indicator caused by the spike occurrence. They rely on several model-specific assumptions. Deconvolution-based approaches consider convolutive assumptions

about the spiking process [8–10, 37] whereas biophysical model-based approaches estimate the most probable spike train which generated the fluorescence output [11]. Other model-based approaches include template matching [3, 12], auto-regressive formulation [31] and approximate Bayesian inference based on deconvolution [5, 13]. These models are limited by the a-priori assumptions about the model which has stringent approximations regarding the shape of the calcium response and the noise statistics. Recently, a non-model based signal processing approach for spike estimation is presented in [16] which uses agnostic nature of group delay to estimate the spike locations. It has a comparable performance with other popular algorithms in the literature.

2. Supervised models predict the spike information from the fluorescence signal either using a set of features derived from the signal or using the raw signal itself. Data-driven methods are gaining traction recently owing to the availability of simultaneous electrophysiological and two-photon scanning-based recordings of neurons. Supervised Spike Triggered Mixture (STM) model [17] is used for learning the λ parameter of a given Poisson model using a neural network in [18]. Recent methods use fluorescence signals with or without a contextual window (supplementary material - [15]) for estimating the spike information. Neural network-based variants such as "convi6", "Deepspike", "Purgatorio', and "Embedding of CNNs" have had varied levels of success, and outperformed data-driven baseline method [17] on a standard evaluation framework (supplementary material - [15]). A gated recurrent unit (GRU)-based approach recently attempted to estimate action potentials directly from the 2-D calcium imaging output, combining regions-of-interest (ROI) estimation and spike estimation tasks [19]. A variant of adversarial variational auto encoder (VAE) is employed in [14] for improved spike inference over factorised posteriors used in standard VAEs.

Spikefinder challenge (`http://spikefinder.codeneuro.org/`) [15] provided a new set of algorithms which performed better than the benchmark [17]. Although these algorithms use different techniques, it was found that their fusion could not improve the results further. The challenge was aimed to standardise the spike estimation evaluation and to provide comparison of state-of-the-art techniques. Most of the top-performing algorithms in the contest used convolutional, recurrent and, deep neural networks and its variants [15]. All the top-performing data-driven algorithms have recurrent layer in the network and hence are computationally complex. The best-performing supervised model used a convolutional neural network architecture with an intermediate LSTM layer to predict the spiking probability from a contextual window of fluorescence signal ("convi6" in supplementary material of [15]). Generative models such as MLspike [11] and auto-regressive model [31] were comparable to the supervised approaches with dataset-specific parameter tuning on the training datasets. MLspike uses a biophysical model and estimates the maximum probable spike information given the fluorescence signal. The second best generative approach in spikefinder is based on autoregressive approximation to the calcium fluorescence signal [31]. Spike information is then estimated by solving a non-negative sparse optimisation problem.

Approaches which model sequences in an end-to-end fashion are being developed recently for classification and regression tasks in audio signal processing and natural language understanding with commendable success. One of the first efforts in that direction was done in the context of machine translation [21]. This approach was later extended to speech recognition [22, 23]. Sequence-to-sequence models have been used for text summarization task using RNNs in [20]. End-to-end methods have been also employed to predict the speech of the target speaker from the overlapped speech mixture

for speech separation task [24, 25] and clean speech for enhancement tasks [26, 27]. In this case, as the models work on raw-waveform, they do not necessitate hand crafted features, rather they learn the relevant information in a task-dependent manner from the signal. They also provide the flexibility to choose task-specific objective functions for training which implicitly consider the temporal context. It is possible to analyse the filters to understand the learning trends in the temporal and frequency domains.

Motivated by the success of sequence-to-sequence models, we present a signal-to-signal neural network (hereafter referred to as S2S) for spike estimation task. The proposed method can be visualised as an analysis-synthesis method, where, as illustrated in Fig 1, the input calcium fluorescence signal is analysed by an input convolution layer; filtered by time distributed dense layer(s) (also called as hidden layers);[1] and finally the spike signal is synthesised by an output convolution layer. This synthesis layer makes our architecture distinct from other spike estimation methods. All the parameters of the network are learned in an end-to-end manner with a cost function based on Pearson correlation between the estimated spike signal and the ground truth spike signal. Contrary to *all* the machine learning approaches to spike inference which predict one output value for an input feature/signal (with optional window), the S2S network synthesises the spike information signal from the calcium fluorescence signals. We hypothesise that such a spike estimation network can outperform existing approaches as it reconstructs the spike information for each sample of the input. This network differs from the above discussed sequence-to-sequence models in several aspects. First, the nature or characteristics of the output signal (discrete spike estimates) is very different from the input signal (calcium fluorescence trace). Second, it performs automatic short-time processing by means of shared weights across the temporal axis. Third, as we will see later, the output of each layer can be visualised to gain insight into the information captured by each layer. The frequency responses of both analysis and synthesis layers in the network can also be analysed. Finally, this method shows the applicability of signal-to-signal neural network for processing neuronal signals.

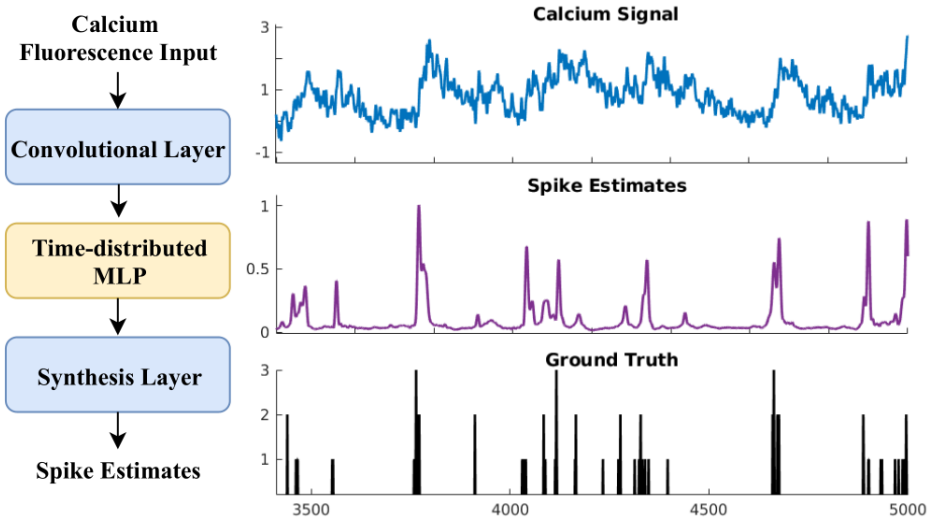

Fig 1. Block diagram of the S2S network.

The proposed S2S method is compared with competitive algorithms in the spikefinder contest with the same dataset and evaluation procedure. The S2S method

[1] In this article, we use the terms dense layer and hidden layer interchangeably.

provides state-of-the-art results for spike estimation. A detailed comparison is made with the best neural network-based model in the spikefinder contest. The research questions in terms of reliability, generalisation ability, dependency on training targets and, design concerns of S2S are studied. A layer-wise analysis of the network is also performed to provide an intuitive explanation for the learning.

Results

We used Spikefinder challenge dataset [15] for evaluations. It includes five bench-marking datasets consisting of 92 recordings from 73 neurons. One part of this dataset was given for training the supervised models and the other part for testing as a part of the competition. Five datasets [29] from GENIE project [40] were also available for training the models. These additional datasets make sure that the supervised models do not over-fit to the training data. It further tests the generalisation capability of the supervised methods. For further details about the spike finder dataset, the reader is referred to [15]. As per the protocol of spike finder challenge, the computations are done at 25 Hz (40 ms bin width). Evaluation is carried out in terms of Pearson correlation coefficient, Spearman's rank correlation coefficient and area under the receiver operating characteristics (denoted as AUC). We use the Pearson correlation coefficient as the primary evaluation measure, as done in the Spikefinder challenge. Whilst, Spearman's rank correlation coefficient and AUC serves as secondary and tertiary evaluation measures, respectively. This enables us to benchmark the proposed S2S method against the Spikefinder challenge submissions.

The performance of the S2S method is compared to top six algorithms in the spike finder contest. They are either based on generative [11, 31] or supervised [15, 17, 32, 38] approaches. 7 out of top-10 algorithms are deep learning-based supervised algorithms. Table 1 provide an overview of the baseline methods.

Table 1. Overview of top-performing algorithms in spikefinder challenge.

Team	Contributer(s)	new	type	Model/Architecture
Team1	T. Deneux	No [11]	Generative	Biophysical model
Team2	N. Chenkov, T. McColgan	Yes	Supervised	conv / lstm
Team3	A. Speiser, J. Macke, S. Turaga	Yes	Supervised	RNN/CNN
Team4	P. Mineault	Yes	Supervised	residual / lstm
Team5	P. Rupprecht, S. Gerhard, R. W. Friedrich	Yes	Supervised	conv / max
Team6	J. Friedrich, L. Paninski	No [31]	Generative	Autoregressive model
Baseline: STM	L. Theis	No [17]	Supervised	DNN +Poisson model

Table 2 compares the performance of the S2S method with the methods reported in the spikefinder challenge [15]. It can observed that the S2S network outperformed all the state-of-the-art methods; both generative and supervised. It improved the test correlation by 46% compared to the best performing algorithm in the spikefinder contest. Change in correlation with respect to the challenge baseline system STM (denoted as Δ) is significantly high for the proposed approach (3.5 times compared to the best reference algorithm). S2S provided an improvement of 56% for the secondary evaluation measure over the baseline with the best rank. It achieved comparable AUC with respect to the best method (in terms of AUC measure). The deviation between the correlation coefficient on the train and test sets for the S2S method is minimum (0.0079). Interestingly, S2S is the only method for which the correlation coefficient on the test set is more than the training set, indicating that the proposed method generalises well. Furthermore, as presented in Table 3, the performance of the proposed approach over

the five different test datasets showed a commendable correlation with respect to the shape and monotonicity of the predicted spike information and ground truth.

Table 2. Performance comparison of S2S with spikefinder baselines.

Team Name	Train correlation	Test correlation	Δ correlation	Rank	AUC
Team 1 MLspike new	0.4823	0.4382	0.0810	0.2878	0.846
Team 2 convi6	0.4727	0.4378	0.0806	0.3319	0.846
Team 3 DeepSpike	0.4730	0.4347	0.0775	0.3338	0.851
Team 4 Purgatorio	0.5370	0.4325	0.0753	0.3258	0.815
Team 5 Embedding of CNNs	0.4900	0.4291	0.0719	0.2822	0.821
Team 6 Suite2p	0.4752	0.4188	0.0617	0.3071	0.821
Baseline STM	0.4024	0.3572	-	0.2664	0.821
Proposed S2S	0.6325	**0.6404**	**0.2832**	**0.5208**	0.847

Table 3. Dataset-wise performance of S2S on the test set.

Measure	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Average
Correlation	0.657	0.910	0.585	0.440	0.611	0.640
Rank	0.420	0.816	0.525	0.261	0.581	0.521
AUC	0.901	0.901	0.874	0.682	0.879	0.847

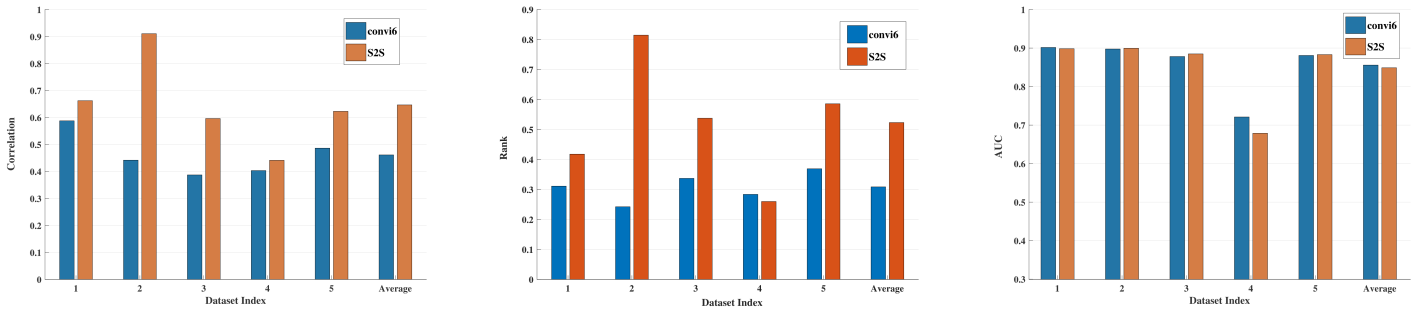

Fig 2. Comparisons with state-of-the-art supervised model "convi6" (Team 2).

Fig 2 compares the performance of S2S with the best supervised approach "convi6" (Team 2 in Table 2 has the highest Pearson correlation coefficient). Across all the five different test sets, S2S achieved better or comparable Pearson and Spearman correlation coefficients and achieved comparable AUC. We also investigated retraining of the neural networks in the S2S method and the convi6 method. We found that the S2S method was insensitive to re-initialization, while convi6 method performance varied on average. These results show that the proposed S2S method yield reliable estimates of the spike signal.

Signal conversion helped in reconstruction of spike information that closely resembles discrete ground truth. S2S had reliable output performance. i,e. it provided similar output after a fresh training process, independent of the initialisation of parameters. Best baseline performance (for primary evaluation measure) is achieved when using the weights provided by the authors. The performance varies by 6% on average with respect to the baseline correlation when the model is trained from the scratch with the same data and model configuration, but over 2 different training sessions. S2S was significantly faster than the baseline. On Sun GPU clusters S2S is 50 times faster than the baseline model.

Analysis

In this part of the article, we provide analysis of the proposed S2S method in terms of architecture, target for training, generalisation capabilities and layer-wise interpretation of the trained network.

Training target

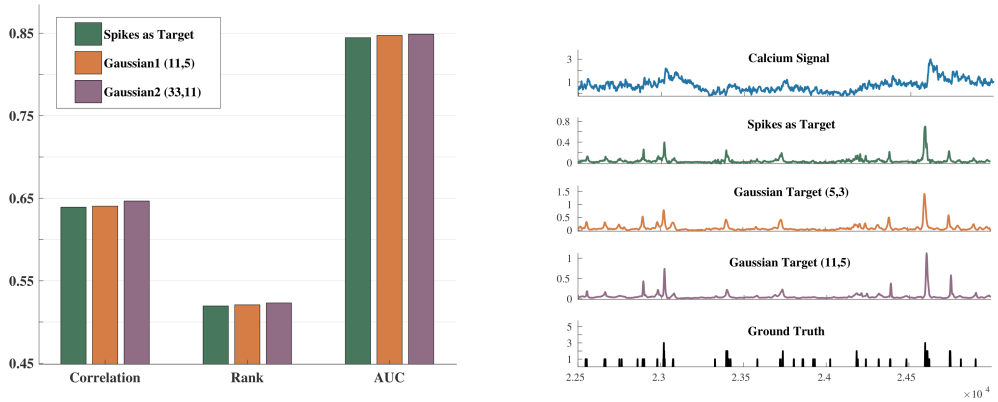

Fig 3. Performance with various training targets.

As discussed later in the Methods section, the S2S network was trained with less-sparse ground truth or target by convolving the sparse discrete spike signal with Gaussian function. Although the Gaussian window hyper-parameters $(11, 5)$, where (x, y) denotes Gaussian window width x and standard deviation y in number of samples, were obtained through cross validation during the network training, a question that arises is: what is the impact of less-sparse targets? So, we investigated training of the S2S network by convolving the target spike signal with Gaussian window of variable sizes. For the sake of clarity, Fig 3 compares the performance obtained for $(11, 5)$ with the performance obtained for $(33, 11)$ and the performance with discrete spike signal as the target signal. It can be observed that convolving with Gaussian function helps. The performance in terms of all the three evaluation measures improves from discrete spike signal to convolution with Gaussian window (5, 3) to convolution with Gaussian window (33, 11). Increasing the width of the Gaussian window beyond $(33, 11)$ resulted in a reduced performance, as the shape of the training target became very different from the discrete target (actual ground truth). This hyperparameter might be related to the firing rate of the indicators used in the training dataset. We observe that a window of $(11, 5)$ suits best for the spikefinder challenge dataset.

Architecture

The number of hidden layers in the network architecture (explained in Methods section) was chosen based on cross validation. We examined the role of the hidden layers by varying the number of hidden layers from three (used in the experiments reported in the Results part) to zero. The target for training was obtained by convolving the Gaussian window (11, 5) with the discrete spike signal. Fig 4 presents the results in terms of the three evaluation measures. It can be observed that, even without an hidden layer, the proposed S2S method yields system competitive to state-of-the-art systems. With the hidden layers the performance improves. Beyond three hidden layers, we found that the performance saturated. As illustrated in the subplot with an example of the input

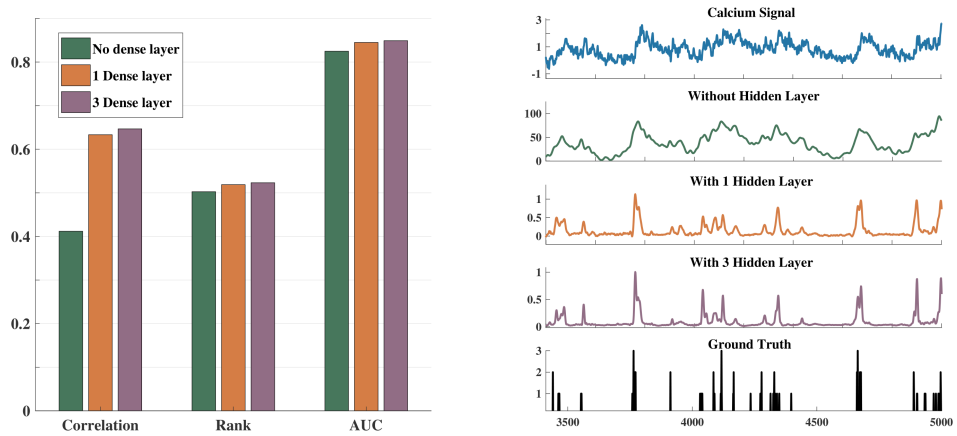

Fig 4. Performance with various number of hidden layers.

signal, in the spike signals estimated with different hidden layers, hidden layer(s) help in improving denoising and spike-resolving capabilities, when compared to the case of without a hidden layer.

Generalisation ability

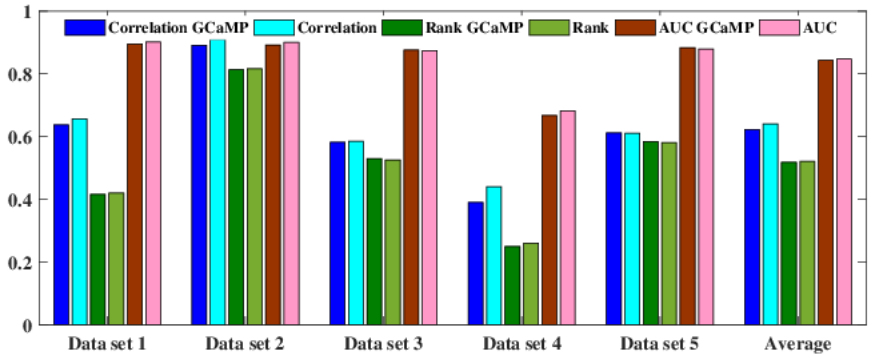

Fig 5. Dataset-wise performance showing the generalisation ability of the three hidden layer S2S network. "GCaMP" indicates that the training was done only with GCaMP indicator dataset.

The spikefinder challenge consists of data obtained with two different indicators, namely, GCaMP indicator and OGB indicator. In the challenge protocol, both the training and test conditions contain signals from both the indicators. We examined the generalisation ability of the proposed approach across different indicators by training the S2S network only on the GCaMP indicator data in the training set and testing on test set of spikefinder contest containing data from both indicators. We trained S2S networks with one hidden layer and three hidden layers. The training was carried out with targets obtained by convolving the ground truth discrete spike signals with Gaussian window $(11, 5)$. Table 4 presents the results for both one and three hidden layer S2S networks. It can be observed that S2S network trained only with GCaMP indicator yields competitive performance with that trained using all the training data. Furthermore, it can be observed in Fig 5 that the evaluation measures on OGB datasets

only gracefully degraded for the GCaMP-only trained model in comparison with the combined-model for *all* the datasets.

Table 4. Generalisation across indicators.

Configuration	Training Indicator(s)	Correlation	Rank	AUC
1 hidden layer	GCaMP	0.608	0.510	0.832
1 hidden layer	GCaMP + OGB	0.639	0.519	0.845
3 hidden layers	GCaMP	0.622	0.518	0.843
3 hidden layers	GCaMP +OGB	0.640	0.521	0.847

Layer-wise output

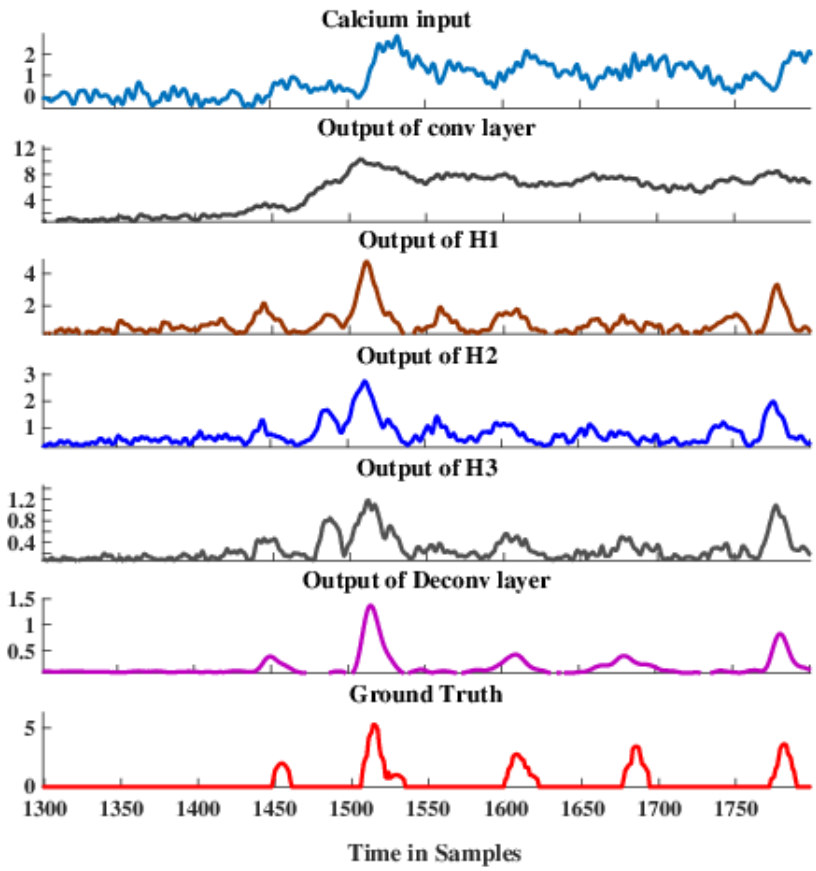

Fig 6. Layer-wise outputs of a 3-dense layer S2S network.

The proposed method is interpretable, in the sense that the processing carried out by each layer in the network can be visualised to gain insights. Fig 6 shows the layer-wise outputs of the three hidden layer S2S network. These are generated by feeding the Calcium input signal and calculating the total response per layer with the trained weights at each layer. For more details regarding the estimation of layer-wise output, the reader is advised to refer to the Methods section. For the given Calcium fluorescence input, the analysis convolution layer seems to make it less-noisy (or smooth out fine

variation in the signal), while maintaining possibly the information "relevant" for spike estimation. Observe that the shape of the input signal is preserved. The first hidden layer output after ReLU non linearity seems to resolve potential spike positions. This output is then refined or filtered by the subsequent two hidden layers and the output synthesis layer by enhancing the signal at the true spike locations and suppressing the spurious spike locations. This yields an output spike signal estimate that correlates well with the ground truth. Each layer is contributing towards the goal of maximising the similarity or correlation between the spike signal estimate and the ground truth. It is worth noting that output of hidden layers H2 and H3 do not differ too much when compared to H1. The output synthesis convolution layer seems to carry out the major refinements. This observation is inline with the earlier observation that one hidden layer S2S network yields performance comparable to the three hidden layer S2S network.

Filter responses

The S2S network learns to synthesise the spike information from the Calcium fluorescence signals owing to the presence of filters in the analysis and synthesis convolutional layers. Hence, we analysed the frequency response of the filters. Fig 7 shows the cumulative frequency response of the filters in the analysis convolutional layer and the synthesis convolutional layer. It can be observed that the analysis layer gives emphasis to low frequency information (between 0-5 Hz). Whilst in the case of synthesis layer, we observe a harmonic structure, which could be attributed to the fact that that the layer is predicting spike signal. The harmonic structure is present in the time-domain analysis as well. This is probably because the network is learned to predict spike trains with different time-scales and frequency components. This special architecture makes our approach distinct and more efficient in comparison to other neural network-based spike estimation methods.

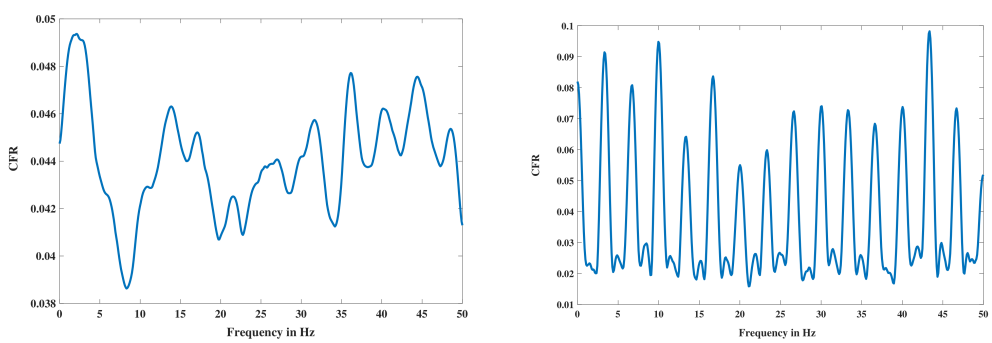

Fig 7. Cumulative frequency responses of filters: (**left**) analysis and (**right**) synthesis filters.

Discussion

We showed that the proposed S2S network method can provide state-of-the-art results in spike estimation task. The S2S network resulted in significant improvement of both primary and secondary evaluation measures. This architecture seems to be appropriate for the spike estimation problem. In particular, the use of synthesis layer, which reconstructs the spike signal, and the cost function based on correlation coefficient resulted in faster convergence and better performance.

Design aspects of the S2S network

We examined various design aspects of the S2S network. Our studies revealed that: (a) making the targets less-sparse by convolving the ground truth discrete spike signal with Gaussian window helps, (b) at least one hidden layer is needed to resolve well the spike locations, (c) the method is not sensitive to initialisation; converges within 50 epochs; and yields similar performances when retrained (the results are reproducible), and (d) the method is capable of generalising across unseen indicators.

One of the major concerns in using a supervised approach for spike inference is its computational complexity. Generative approaches, on the other hand, require mainly novel parametric settings for new datasets. S2S method, although being a supervised method, is computationally efficient. All the top performing supervised algorithms in the spikefinder contest had recurrent units in the architecture, which resulted in increased training time. Proposed S2S network is 50 times faster than the best performing supervised model when trained with GPU using Sun Grid Engine. This is potentially due to the fact that the network has very few trainable parameters and the use of time-distributed hidden layers. As presented in the Methods section, the three hidden layer S2S network has only 8790 trainable parameters. This has potential implications in terms of generalisation.

Multiple evaluation measures

Performance evaluation of a spike estimation algorithm should consider the overall shape of the spike information signal, monotonic relationship with the ground truth and, consider the accuracy of the temporal estimates of spikes with various thresholds. There is no single measure that encompasses all these information [33]. For instance, Pearson correlation coefficient mainly considers the shape of spike density function into account. However, this measure is invariant under affine transformations. Thus, it is difficult to interpret the outputs as spike counts or rates. Spearman's rank order correlation coefficient measures both the strength and the direction of association between the two ranked variables. The ranking is based on the dynamic ranges of the spike information and discrete spike train. Unlike Pearson correlation coefficient, this correlation measure also considers the possible non linear relationship between the variables. Finally, area under the receiver operating characteristics (AUC) measures how well the spikes have been detected. AUC is not sensitive to changes in the relative height of different parts of the spike information. Thus, this measure alone is not adequate. The proposed S2S method outperforms all the other systems in terms of Pearson correlation coefficient and Spearman's rank correlation coefficient, and achieves comparable performance in terms of AUC (only DeepSpike method has marginally high AUC). This indicates that the S2S method is yielding better estimate of spike signal than existing generative and supervised methods. Specifically, the S2S network resulted in an improvement of 46% in primary evaluation measure which is two times that achieved by the Spikefinder contest (23%). This performance gain could help in further directions in spike estimation based on signal reconstruction strategies.

Methods

Signal-to-signal network for spike signal estimation

As discussed earlier in the Introduction section, in various fields related to sequence processing such as speech processing and natural language processing, approaches are emerging to convert one sequence into another sequence in an end-to-end manner. For instance, converting sequence of acoustic features to sequence of letters or words.

Translating sequence of words from one language to sequence of words in another language. We cast the problem of spike estimation from the Calcium fluorescence signal as a sequence-to-sequence conversion problem, as a signal is a time-ordered measurement. Towards that, we take inspiration from a recent work on end-to-end speech source separation [24] to a develop signal-to-signal network which, as illustrated in Figure S1, learns to predict/estimate spike signal given the Calcium fluorescence signal as input. Intuitively, the S2S network can be seen as a single channel signal enhancement system where the desired spike signal present in the Calcium fluorescence signal is enhanced, while the undesirable signals or variabilities are suppressed.

The S2S network consists of a convolution layer which is followed by fully connected dense layer(s), and finally an output convolution layer. The hidden layers are implemented in a time-distributed manner, i.e. the output of the first convolution layer for every segment of Calcium fluorescence signal is independently processed by the hidden layer(s) and then synthesised by the output convolution layer. In the network, the output of each hidden node or filter is fed to a rectified linear unit (ReLU) non-linearity, except for the output convolution layer. ReLU units in the neural network do not saturate unlike *sigmoid* and *hyperbolic tangent* activations. They also help in learning a non-negative representation in the successive layers.

The hyper-parameters of the S2S network are: the length of signal input w_{seq}, kernel width kW_{in} and kernel shift dW_{in} of the input convolution layer, number of filters in the input convolution layer $nFilt_{in}$, number of hidden layers I and the number of nodes nhu_i in each hidden layer $i \in \{1, \ldots I\}$, kernel width kW_{out} and kernel shift dW_{out} of the output convolution layer and the number of filters in the output convolution layer $nFilt_{out}$. Based on the systems reported in the spikefinder challenge, we set $w_{seq} = 100$ samples (i.e., 1 sec). We set $kW_{in} = w_{seq}$ and $dW_{in} = 1$ sample. In the output convolution layer, $kW_{out} = nhu_I$ (i.e. number of nodes in the last hidden layer), $dW_{out} = 1$ sample and $nFilt_{out} = w_{seq}$. In other words, for every frame of 100 sample input, the output convolution layer synthesises 100 samples of spike signal, which are then overlapped and added to get a spike signal of same length as that of the input Calcium fluorescence recording. The number of frames is determined by dW_{in}, which is one sample. The remainder of the hyper-parameters $nFilt_{in} = 30$, $I = 3$ and $nhu_1 = nhu_2 = nhu_3 = 30$ were determined by cross validation on the training set through a coarse grid search. Thus, the three hidden layer S2S network compared against the state-of-the-art methods in the Results section has $(100 \times 30 + 30) + (30 \times 30 + 30) + (30 \times 30 + 30) + (30 \times 30 + 30) + (30 \times 100) = 8790$ parameters. In the case of the one hidden layer study presented in the Analysis section, the network has $(100 \times 30 + 30) + (30 \times 30 + 30) + (30 \times 100) = 6960$ parameters, and in the case of no hidden layer $(100 \times 30 + 30) + (30 \times 30) = 6030$ parameters.

The S2S network is trained by optimising a cost function based on the Pearson correlation coefficient between the spike signal estimated by the network and the ground truth spike signal. It is worth mentioning that we also investigated other cost functions such as, mean square error, cross correlation with sigmoid non-linearity after the output convolution layer. The Pearson correlation coefficient yielded the best system on both the cross validation experiments conducted on the training set and the experiments on the test set. For training the S2S network, we split the training data in the spikefinder challenge into two parts: training set (80%) and validation set (20%). At each training epoch, the training set was used for training the network parameters and the validation set was used for cross validating the network. We used the Adam optimiser with a starting learning rate of 0.001 and a batch size of 20. Early stopping with a patience factor of 6 is used to stop the training whenever the validation cost increases with respect to the previous epoch. As the ground truth is a sparse discrete signal, for efficient training of the network, the target spike signal was convolved with a Gaussian

window to make the targets less sparse. The Gaussian window $(11, 5)$ was obtained through cross validation, i.e. the window that yielded the best validation cost. The implementation was carried out on Keras [42] with TensorFlow [43] backend. Upon acceptance of this paper, the software will be made available at `https://github.com/Jiltseb/S2S_for_spike_inference`.

Data

We used the spikefinder challenge dataset to validate the proposed S2S network method. We followed spikefinder evaluation [15] owing to the following reasons: First, it is one of the largest publicly available dataset containing different scanning rates and methods and Calcium indicators. Second, spikefinder challenge provides a bench-marking framework to compare different spike estimation methods using different evaluation measures. Finally, the challenge provides state-of-the-art baseline systems to which the proposed method can be compared to. As mentioned earlier, spikefinder dataset has five bench-marking datasets consisting of 92 recordings from 73 neurons. Five additional datasets from GENIE project [40] have been provided to ensure that the supervised models do not over-fit to the training data and also to test the generalisation ability of the supervised methods. The zoom factor of all the recordings are at use-case resolutions for the calcium imaging experiments. For further information, the reader is referred to [15]. The dataset is available at the open source public repository (`https://github.com/codeneuro/spikefinder`).

Comparison framework

Performance of the proposed method was compared to two set of baselines: The results are first compared to the top-five algorithms in the spike finder contest. They are either generative [11, 31] or supervised [15, 17, 32, 38] approaches. Generative approach, which was originally published in [11] uses a biophysical model and estimates the maximum probable spike information from the fluorescence signals. An efficient Bayesian inference is performed on the model including a drifting baseline and nonlinear modelling of calcium to fluorescence conversion. This was the winner of spikefinder contest. The performance of this algorithm is boosted (w.r.to the original work) owing to the parameter tuning with respect to the spikefinder challenge dataset. The second best generative approach [31] was based on auto-regressive approximation to the calcium fluorescence signal. Spike information is then estimated by solving a non-negative sparse optimisation problem. Based on the advances in neural network-based models for various applications, it is not surprising that 7 out of top-10 algorithms are deep learning-based supervised algorithms (see Table 1). All the top performing algorithms used recurrent layer in the network. For the best-performing algorithms, we have taken the results in terms of correlation and AUC from the spikefinder challenge [15], whereas the rank measure is obtained by averaging the cell-wise results across the datasets in the evaluation framework (`https://github.com/berenslab/spikefinder_analysis`).

We compared the performance and efficiency of the proposed network with the best performing supervised baseline, available as an open source Python software (`https://github.com/kleskjr/spikefinder-solution`). We trained the network using the open source software. However, the neural network model weights provided by the authors yielded better linear correlation values than that obtained by training the network. Convolutional layers take a large temporal context to predict a single output corresponding to the spike information. Dataset index provided as an auxiliary input to the network observed to improve the performance of the system. Convolutional layers are followed by an intermediate LSTM-recurrent layer and further by a set of

convolutional layers of smaller width. The learned filters collectively capture the spike-related information.

Evaluation

All the evaluations were done at 25 Hz (40 ms bin width), following the protocol used in spikefinder evaluations (`https://github.com/berenslab/spikefinder_analysis`). As per the protocol, we used Pearson correlation coefficient as the primary evaluation measure. Δ correlation is computed as the average difference in correlation coefficient compared to the STM algorithm. Secondary evaluation measure was Spearman's rank correlation coefficient, which considers the direction and non-linear relationship between the estimates and actual spikes. Area under the ROC curve (AUC) curve was the final measure which evaluates detection of spikes in a 40 ms bin. We used the function *roc_curve* from *scikit-learn* [44] Python package for computing the AUC and *spearmanr* function from *scipy* [45] Python package for computing the rank. These evaluation measures collectively denote the similarity with the ground truth. We consider them in the order of their preference as in the spikefinder challenge. The rank measure on test set was not included in the challenge paper, although it was the secondary measure on the training sets.

Conclusion

This paper presented signal-to-signal conversion networks (S2S) and its use in spike estimation from imaging data. S2S filters synthesise the output which has maximum similarity with the discrete ground truth. The network performs automatic short-time processing and outperforms state-of-the-art spike estimation methods on standard evaluation framework. Each layer of the network has a specific functionality and the synthesis filters have a harmonic frequency response covering a rich basis set.

Acknowledgements

This work was partially supported by Swiss Government Excellence Scholarship Project with ESKAS No: 2017.0575. The authors would like to acknowledge the contribution of Mr. Pavan Kumar Dubagunta for his fruitful discussions towards the work and Mr. Mari Ganesh Kumar for his review of the draft and discussions about the spikefinder challenge.

References

1. Kerr JN, Denk W. Imaging in vivo: watching the brain in action Nature Reviews Neuroscience. 2008; 9(3):195–205. https://doi.org/10.1038/nrn2338 PMID: 18270513

2. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS. Three-dimensional mapping of microcircuit correlation structure. Frontiers in neural circuits. 2013; 7:151. https://doi.org/10.3389/fncir.2013.00151 PMID: 24133414

3. Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. Nature methods. 2010; 7(5):399. https://doi.org/10.1038/nmeth.1453 PMID: 20400966

4. Pachitariu M, Stringer C, Harris KD. Robustness of spike deconvolution for calcium imaging of neural spiking. bioRxiv. 2017; 156786. https://doi.org/10.1101/156786

5. Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. Biophysical journal. 2009; 97(2):636–55. https://doi.org/10.1016/j.bpj.2008.08.005 PMID: 19619479

6. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature. 2013; 499(7458):295–300. https://doi.org/10.1038/nature12354 PMID: 23868258

7. Wilt B, Fitzgerald JE, Schnitzer MJ. Photon shot noise limits on optical detection of neuronal spikes and estimation of spike timing. Biophysical journal. 2013; 104(1):51–62. https://doi.org/10.1016/j.bpj.2012.07.058 PMID: 23332058

8. Vogelstein JT, Packer AM, Machado Ta, Sippy T, Babadi B, Yuste R, et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. Journal of neurophysiology. 2010; 104 (6):3691–704. https://doi.org/10.1152/jn.01073.2009 PMID: 20554834

9. Jewell S, Hocking TD, Fearnhead P, Witten D. Fast nonconvex deconvolution of calcium imaging data. arXiv preprint arXiv:1802.07380. 2018 Feb 21.

10. Yaksi E, Friedrich RW. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca 2+ imaging. Nature methods. 2006 May;3(5):377.

11. Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, et al. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. Nature Communications. 2016; 7:12190. https://doi.org/10.1038/ncomms12190 PMID: 27432255

12. Greenberg DS, Houweling AR, Kerr JN. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nature neuroscience. 2008 Jul;11(7):749.

13. Pnevmatikakis EA, Merel J, Pakman A, Paninski L. Bayesian spike inference from calcium imaging data. InSignals, Systems and Computers, 2013 Asilomar Conference on 2013 Nov 3 (pp. 349-353). IEEE.

14. Im DJ, Prakhya S, Yan J, Turaga S, Branson K. Importance Weighted Adversarial Variational Autoencoders for Spike Inference from Calcium Imaging Data. arXiv preprint arXiv:1906.03214. 2019 Jun 7.

15. Berens P, Freeman J, Deneux T, Chenkov N, McColgan T, Speiser A, Macke JH, Turaga SC, Mineault P, Rupprecht P, Gerhard S. Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. PLoS computational biology. 2018 May 21;14(5):e1006157.

16. Sebastian J, Kumar MG, Viraraghavan VS, Sur M, Murthy HA. Spike Estimation From Fluorescence Signals Using High-Resolution Property of Group Delay. IEEE Transactions on Signal Processing. 2019 Apr 2;67(11):2923-36.

17. Theis L, Berens P, Froudarakis E, Reimer J, Roma´n Roso´n M, Baden T, et al. Benchmarking Spike Rate Inference in Population Calcium Imaging. Neuron. 2016; 90(3):471–482. https://doi.org/10.1016/j.neuron.2016.04.014 PMID: 27151639

18. Theis L, Chagas AM, Arnstein D, Schwarz C, Bethge M. Beyond GLMs: a generative mixture modeling approach to neural system identification. PLoS computational biology. 2013 Nov 21;9(11):e1003356.

19. Linsley D, Linsley JW, Sharma T, Meyers N, Serre T. Learning to predict action potentials end-to-end from calcium imaging data. IEEE 52nd Annual Conference on Information Sciences and Systems (CISS), 2018, pp. 1-6.

20. Nallapati R, Zhou B, Gulcehre C, Xiang B. Abstractive text Summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023. 2016 Feb 19.

21. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Advances in neural information processing systems 2014 (pp. 3104-3112).

22. Chan W, Jaitly N, Le Q, Vinyals O. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on 2016 Mar 20 (pp. 4960-4964). IEEE.

23. Chiu CC, Sainath TN, Wu Y, Prabhavalkar R, Nguyen P, Chen Z, Kannan A, Weiss RJ, Rao K, Gonina K, Jaitly N. State-of-the-art speech recognition with sequence-to-sequence models. arXiv preprint arXiv:1712.01769. 2017 Dec 5.

24. Venkataramani S, Casebeer J, Smaragdis P. Adaptive front-ends for end-to-end source separation. InProc. NIPS 2017.

25. Luo Y, Mesgarani N. TasNet: time-domain audio separation network for real-time, single-channel speech separation. arXiv preprint arXiv:1711.00541. 2017 Nov 1.

26. Fu SW, Tsao Y, Lu X, Kawai H. Raw waveform-based speech enhancement by fully convolutional networks. arXiv preprint arXiv:1703.02205. 2017 Mar 7.

27. Rethage D, Pons J, Serra X. A Wavenet for speech denoising. arXiv preprint arXiv:1706.07162. 2017 Jun 22.

28. Friedrich J, Zhou P, Paninski L. Fast online deconvolution of calcium imaging data. PLoS Comput Biol. 2017; 13(3):e1005423. https://doi.org/10.1371/journal.pcbi.1005423 PMID: 28291787

29. Svoboda K, Project G. Simultaneous imaging and loose-seal cell-attached electrical recordings from neurons expressing a variety of genetically encoded calcium indicators; 2015. Available from: http://dx.doi.org/10.6080/K02R3PMN.

30. Cotton RJ, Froudarakis E, Storer P, Saggau P, Tolias AS. Three-dimensional mapping of microcircuit correlation structure. Frontiers in neural circuits. 2013; 7:151. https://doi.org/10.3389/fncir.2013.00151 PMID: 24133414

31. Friedrich J, Paninski L. Fast active set methods for online spike inference from calcium imaging. Advances In Neural Information Processing Systems; 2016. p. 1984–1992.

32. Speiser A, Yan J, Archer E, Buesing L, Turaga SC, Macke JH. Fast amortized inference of neural activity from calcium imaging data with variational autoencoders. Advances in Neural Information Processing Systems. vol. 30; 2017

33. Kümmerer M, Wallis TS, Bethge M. Saliency Benchmarking: Separating Models, Maps and Metrics. arXiv preprint arXiv:1704.08615. 2017 Apr 27.

34. Kuemmerer M, Wallis T, Bethge M. Information-theoretic model comparison unifies saliency metrics. Proceedings of the National Academy of Science. 2015; 112(52):16054–16059. https://doi.org/10.1073/pnas.1510393112

35. Reynolds S, Schultz SR, Dragotti PL. CosMIC: A Consistent Metric for Spike Inference from Calcium Imaging. bioRxiv. 2017; 238592. https://doi.org/10.1101/238592

36. Freeman J. Open source tools for large-scale neuroscience. Current Opinion in Neurobiology. 2015;32:156–163. https://doi.org/10.1016/j.conb.2015.04.002 PMID: 25982977

37. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron. 2016; 89(2):285–299. https://doi.org/10.1016/j.neuron.2015.11.037 PMID: 26774160

38. Pachitariu M, Stringer C, Schro¨der S, Dipoppa M, Rossi LF, Carandini M, et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv. 2016; https://doi.org/10.1101/061507

39. Sasaki T, Takahashi N, Matsuki N, Ikegaya Y. Fast and accurate detection of action potentials from somatic calcium fluctuations. J Neurophysiol. 2008; 100(3):1668–1676. https://doi.org/10.1152/jn.00084.2008 PMID: 18596182

40. GENIE project, Janelia Research Campus, HHMI; Karel Svoboda (contact). Simultaneous imaging and loose-seal cell-attached electrical recordings from neurons expressing a variety of genetically encoded calcium indicators. CRCNS.org; 2015. Available from: http://dx.doi.org/10.6080/K02R3PMN.

41. Quan T, Lv X, Liu X, Zeng S. Reconstruction of burst activity from calcium imaging of neuronal population via Lq minimization and interval screening. Biomed Opt Express. 2016; 7(6):2103–2117. https://doi.org/10.1364/BOE.7.002103 PMID: 27375930

42. Chollet, François and others. Keras (2015). `https://keras.io`

43. Girija, SS. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org, 2015.

44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V Vanderplas J. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12(Oct):2825-30.

45. Oliphant TE. Python for scientific computing. Computing in Science & Engineering. 2007 May;9(3).