

# Learning to predict action potentials end-to-end from calcium imaging data

Drew Linsley  
CLPS department  
Brown University  
Providence, RI, USA  
drew\_linsley@brown.edu

Jeremy W. Linsley  
J David Gladstone Institutes  
University of California San Francisco  
San Francisco, California  
jeremy.linsley@gladstone.ucsf.edu

Tarun Sharma Nathan Meyers Thomas Serre  
CLPS department  
Brown University  
Providence, RI, USA  
thomas\_serre@brown.edu

**Abstract**—Advances in two-photon calcium ( $\text{Ca}^{2+}$ ) imaging technologies are rapidly increasing the fidelity and scale of in-vivo neural recordings from behaving animals. These methodologies hold great promises for uncovering the neural computations that support behavior. One remaining challenge is the need to estimate the timing of action potentials (APs) from measured cellular  $\text{Ca}^{2+}$  traces revealed by genetically encoded calcium indicators. Current approaches rely on manually pre-processing  $\text{Ca}^{2+}$  imaging data, with routines for cell segmentation and signal normalization preceding deconvolution. Here, we show that by leveraging recent advances in recurrent neural networks it is possible to learn to detect APs end-to-end directly from raw  $\text{Ca}^{2+}$  recordings. Our novel model achieves state-of-the-art performance on this task. We further demonstrate that the relative success of the approach is based in part on the system’s ability to learn spatio-temporal features that are typically associated with biophysical events known to contribute to the AP process.

**Index Terms**—calcium imaging, spike prediction, convolutional networks, recurrent networks

## I. INTRODUCTION

Calcium ( $\text{Ca}^{2+}$ ) imaging techniques hold the promise to transform neuroscience by allowing large-scale recording of neural activity deep within the nervous system. Intracellular  $\text{Ca}^{2+}$  concentration is highly dynamic and the rapid changes in membrane potentials that lead to action potentials (APs) are typically accompanied with brief and transient changes in cytosolic  $\text{Ca}^{2+}$ . Thus,  $\text{Ca}^{2+}$  concentrations constitute a prime signal for tracking neuronal activity.

The development of fluorescent dyes and buffers that fluoresce when bound to  $\text{Ca}^{2+}$  [17] has led to the creation of a diverse array of  $\text{Ca}^{2+}$  indicators. These indicators vary in their capacity to accurately signal  $\text{Ca}^{2+}$  levels within the cell [2]. In contrast to dyes, which must be loaded into the cell body and fade over time, genetically encoded  $\text{Ca}^{2+}$  indicators (GECIs) allow the genetic targeting of neurons of interest, they are less invasive, and can provide stable signal throughout the life of the neuron, allowing for precise monitoring and dissection of neuronal circuits [12], [14]. The compatibility of GECIs with two-photon microscopy has also enabled  $\text{Ca}^{2+}$  imaging deep within brain tissues at subcellular resolution, even in awake behaving animals [3]. However, there is currently no gold standard method for inferring or deconvolving neural activity (in the form of APs) from  $\text{Ca}^{2+}$  signals, even for popular GECIs such as the GCaMP variants [16].

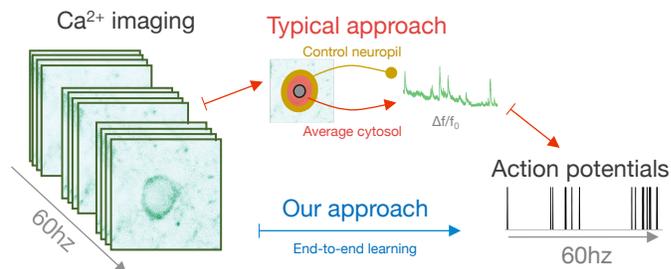


Fig. 1. We describe a learning approach for inferring APs from  $\text{Ca}^{2+}$  imaging data. Typical approaches begin with a series of heuristic preprocessing steps which include 1) segmenting the neuron by hand or with imprecise algorithms and taking the average intensity across its cytosol (red ring; middle panel), 2) measuring surrounding neuropil activity and correcting for it in the neuron’s activity trace (yellow ring; middle panel), and 3) normalizing the signal across time to  $\Delta f/f_0$ . The resulting processed GCaMP6 trace is used for AP deconvolution. In contrast, our approach removes the entire preprocessing stage and operates directly on raw unprocessed GCaMP6 images.

A variety of biological nuisances and technical difficulties make AP detection in GCaMP signal difficult: (i)  $\text{Ca}^{2+}$  kinetics and indicator responses are slower than APs [3]; (ii) Subthreshold  $\text{Ca}^{2+}$  events are not correlated with APs, yet are picked by indicators [7]; (iii) GCaMP is not spatially or temporally standardized (i.e. expression varies across the cell body and the basal intensity varies across time), impeding any comparison between intra- and inter-cellular signal changes and how they relate to APs [13].

The typical approach to AP deconvolution tries to compensate for GCaMP imaging deficiencies and the indirect relationship between  $\text{Ca}^{2+}$  and APs with a series of preprocessing steps (Fig. 1). Broadly, these steps convert raw GCaMP images into a spatially and temporally normalized activity “trace” from which APs can be more readily inferred (“Typical approach” in Fig. 1). Preprocessing steps include segmenting the cell body from the background, averaging the recorded intracellular  $\text{Ca}^{2+}$  signal, correcting for contaminating signals from surrounding neuropil, and normalizing the trace to basal fluorescence across time. To overcome these nuisances, hand-tuning each preprocessing step is critical for existing deconvolution algorithms to succeed. Current methods also suffer because averaging GCaMP intensities across the cell yields a very noisy signal for predicting APs (due to cellular mechanisms unrelated to APs which are also associated with

the release of  $\text{Ca}^{2+}$  and hence distort the average [11].)

Here, we describe a novel approach which leverages recent end-to-end learning methods and alleviates many of the aforementioned limitations. We take advantage of recent developments in recurrent neural network architectures to make the following contributions: **(1)** We describe novel model architectures based on gated recurrent units (GRUs) [4] that deconvolve APs from raw or minimally processed  $\text{Ca}^{2+}$  imaging data. **(2)** We demonstrate that GRUs trained directly on raw imaging data outperform the state-of-the-art [16]. **(3)** By visualizing spatio-temporal features used by the GRUs to deconvolve APs from  $\text{Ca}^{2+}$  imaging data, we confirm that the neural networks successfully learn a meaningful association between APs and changes in cytosolic  $\text{Ca}^{2+}$  signals.

## II. DATASET

Experiments were performed on an open source 2-photon  $\text{Ca}^{2+}$  imaging dataset [3], [6]. The dataset contains simultaneous imaging of GCaMP6s indicator combined with intracellular electrophysiology, which serves as a ground truth for learning the relationship between GCaMP signals and APs. The dataset contains 9 recording sessions total derived from 4 neurons in layer 2/3 of the primary visual cortex in anesthetized mice.  $\text{Ca}^{2+}$  recordings were sampled at 60Hz in a  $30 \times 30 \mu\text{m}^2$  ( $256 \times 256$  pixel) field of view. Electrophysiological recordings were made at 60kHz, thresholded for APs, and then resampled at 60Hz.

We manually inspected the GCaMP6 images and found that all images from one of the neurons as well as those from a late recording session in a second neuron were visually different compared to the others. Abnormalities in these images included an anomalous cell morphology and an atypical distribution of GCaMP across the cell consistent with cell death. For these reasons, we excluded these sessions from our analyses. In total, we kept 7/9 available recordings in the dataset, which spanned 3 different neurons.

This dataset was split into disjoint video segments used for training and testing. The first 80% of the video recording was used for training and final 20% was used for testing. For the training set, 166-ms sequences of images (10 consecutive frames) were sampled with no overlap. We first converted the AP counts across each sequence to a binary indicator (“1” if an AP occurred during a sequence, “0” otherwise). APs occurred in approximately 5% of the sequences, so we oversampled these until they were close in proportion to those with No AP [1]. This procedure expanded the size of the training set from 8,064 to 15,386 sequences. For the test set, we increased the total number of sequences by sampling every 33ms (causing significant overlap between test sequences), which yielded 10,076 test sequences. This provided a more difficult and robust test of AP deconvolution.

We created three different preparations of the training and test splits of data, each of which was designed for a specific model described in detail below and referred to as dataset A, B, and C. (A) Raw GCaMP image sequences, which were normalized by subtracting the scalar mean activity across

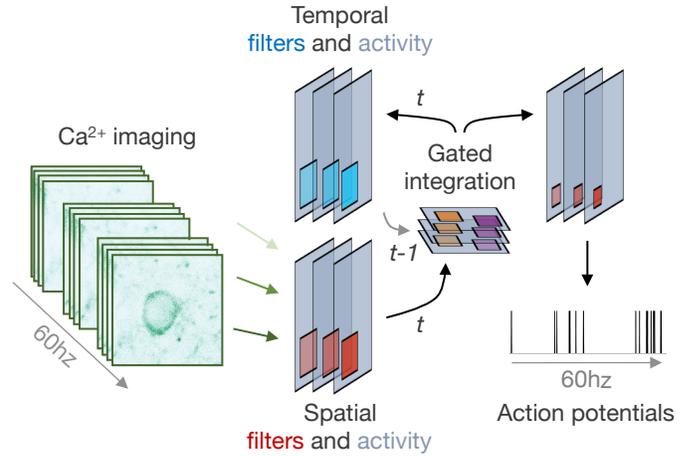


Fig. 2. We developed a recurrent convolutional neural network (2d GRU) for end-to-end learning of AP deconvolution from  $\text{Ca}^{2+}$  imaging data. The model consists of three layers of computations to map GCaMP image sequences to APs. First, a convolutional gated recurrent unit (GRU) with feedforward filters extracts visual features from images in the sequence. Temporal filters and gates accumulate GCaMP signals across images that are diagnostic for AP deconvolution. The final “temporal activity” captures information present throughout the image sequence, and is passed through another convolutional filtering layer that extracts more complex visual features (far right). Finally, a fully connected readout layer learns to predict whether an AP occurred during the image sequence. Colored squares denote convolutional filters, and translucent gray squares denote feature maps. Red corresponds to spatial convolution, blue to time convolutions, orange and purple to recurrent gate convolutions. The readout layer is not shown.

all training sequences; (B) neuropil-corrected traces of  $\text{Ca}^{2+}$  activity taken from the segmented cell body provided by [3] and (C) an additional version of (B) with activity normalized to basal fluorescence across time. This temporal normalization was implemented as standard  $\Delta f/f_0$  given by:

$$\Delta f/f_0 = (f_t - f_0)/f_0 \quad (1)$$

where  $f_t$  denotes the preprocessed activity value at time  $t$ , and  $f_0$  denotes the mean activity across the preceding 1 second of recordings (60 frames). In words,  $\Delta f/f_0$  normalizes activity at time  $t$  with the running mean  $f_0$ .

## III. PROPOSED MODELS

It is typically assumed that successful AP deconvolution from  $\text{Ca}^{2+}$  imaging data requires (i) knowledge of the relationship between neuronal  $\text{Ca}^{2+}$  dynamics and APs as well as (ii) an explicit model of the measurement noise that plagues  $\text{Ca}^{2+}$  imaging data. This challenge is usually met with several manual preprocessing steps before deconvolving APs from imaging data. We hypothesized that recent developments in recurrent neural networks could yield an alternative end-to-end learning solution – bypassing the need for any manual intervention.

### A. GRU Model

We introduce the 2d gated recurrent unit (2d GRU), an architecture that pairs spatial convolutional feature extraction with a recurrent temporal memory with the goal of learning complex and long-range spatio-temporal dependencies. The

proposed 2d GRU comprises a 3-layer hierarchy (Fig. 2): 1) a convolutional GRU, 2) a convolutional layer, and 3) a fully connected readout layer. The model receives a sequence of GCaMP6 images  $\mathbf{x}_s(t) = (x_{t-t_0}, \dots, x_t)$  as input, where  $t$  represents the endpoint of the sequence. This sequence, taken from dataset A, is mapped onto an indicator variable  $y_s$  which encodes whether it coincided with an AP. All models were trained and tested on 166-ms long sequences (10 consecutive frames, so  $t_0 = 9$ ).

The model’s convolutional GRU layer modifies a typical gated recurrent unit [4] with convolutional operations. Its dynamics are governed by the following equations:

$$r_t = \sigma(W_r * x_t + U_r * h_{t-1} + b_r) \quad (2)$$

$$z_t = \sigma(W_z * x_t + U_z * h_{t-1} + b_z) \quad (3)$$

$$\tilde{h}_t = \zeta(U_{\tilde{h}} * (r_t \odot h_{t-1}) + W_{\tilde{h}} * x_t + b_x) \quad (4)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (5)$$

Let  $r_t$ ,  $z_t$ ,  $\tilde{h}_t$ ,  $h_t$  and  $x_t$  denote a reset gate activity, an update gate activity, a temporal “hidden state” activity, a candidate temporal hidden state activity, and an input GCaMP video frame at time  $t$ , respectively (this will be a single frame from an  $\mathbf{x}_s(t)$  sequence.) These activities are derived from convolutions between an input frame  $x_t$  with the kernels  $W_*$  and  $h_{t-1}$  or the hidden state activity  $h_t$  with the kernels  $U_*$ . Kernels  $W_x$  and  $U_h$  are of size  $11 \times 11 \times 1 \times 96$  as in height  $\times$  width  $\times$  number of input channels  $\times$  number of filters used to compute the spatial and temporal recurrent hidden state activity. Lastly,  $b_r$ ,  $b_z$ , and  $b_x$  are bias terms. This architecture is depicted in Fig. 2, with  $W_*$  signifying spatial filters,  $U_*$  temporal filters, and  $h_*$  temporal activities. Element-wise multiplication is signified by  $\odot$ .

To reduce the number of parameters in the model, we used separable convolutions to compute all gate activities (approximating a full-convolution kernel with separate spatial and channel filters.) This approach significantly cuts down on parameters, which we found critical for successful training.  $W_i^d$  and  $U_j^d$  are spatial filters and  $W_i^p$  and  $U_j^p$  are channel filters. Spatial filters are of size  $3 \times 3 \times 1 \times 1$  and channel filters are of size  $1 \times 1 \times 1 \times f$ . To facilitate training, we also initialized  $W_x$  with weights from the first layer of the popular AlexNet [9].

Nonlinearities in the convolutional layers are incorporated to help the GRU learn complex spatial and temporal regularities. The nonlinear function  $\sigma(\cdot)$  is an element-wise sigmoid function used to calculate  $r_t$  and  $z_t$ . This effectively transforms unit activities into gates which zero out irrelevant information from the hidden state activity at the previous time step (the reset gate,  $r_t$ ), and help gate elements of the hidden states which should be preserved across time (the update gate,  $z_t$ ). The nonlinear function  $\zeta(\cdot)$  rectifies the updated hidden state with a scaled exponential linear unit [8] (SeLU), which drives activities towards zero mean and unit variance and which we found to improve model training. The final hidden state activity of the convolutional GRU is fed to an additional

convolutional layer with kernel size  $3 \times 3 \times 96 \times 128$  (far right of Fig. 2). A SeLU is also applied to these activities.

The 2d GRU is trained by adding one or multiple fully connected layers to its final convolutional layers to yield an output binary classification label (AP vs. No AP) for a sequence of GCaMP6 images  $\mathbf{x}_s(t)$ . We experiment with two different loss functions for training the model end-to-end. The first version of the model was trained for binary classification (AP vs. No AP). A single fully connected layer was used in this model, and we minimized the binary cross-entropy between its output prediction  $\tilde{y}_s$  and the ground truth  $y_s$ . We refer to this as the “single loss” version.

For the second version of the model, we tested whether a latent representation of the  $\text{Ca}^{2+}$  kinetics specific to each neuron in our dataset could improve AP deconvolution performance. This was implemented by training a 2d GRU with two distinct fully connected layers: one for detecting APs across  $\mathbf{x}_s(t)$  as in the single loss model, and another for identifying the neuron pictured in  $\mathbf{x}_s(t)$ . Because there were three neurons in this dataset, this second loss was the categorical cross-entropy between the predicted neuron identity  $\tilde{y}_{sid}$  and the actual identity  $y_{sid}$ . This second version of the model (“dual loss”) was trained by minimizing the combination of these two losses. The assumption behind this loss is that forcing the network to recognize the different cells may force it to learn implicit associations between cell morphology and AP detection.

We compared the 2d GRU to an architecture that was trained on 1d activity traces from dataset B. This model, which we refer to as the 1d GRU, uses the same sequence of operations as the 2d GRU but modified for 1d data. It consists of a GRU followed by a 1d convolution kernel of width 3 and finally one or multiple fully connected layers. Separate versions of the 1d GRU were trained with single and dual losses.

Both the 1d and 2d GRU models were trained with TensorFlow and the Nadam optimizer [5] (which augments the popular Adam optimizer with Nesterov momentum). Model training continued until the Nadam optimizer converged, which took approximately 100 epochs. The 2d GRU was trained on images that were resized with area interpolation to  $192 \times 192$ . Images were then augmented with random crops to  $164 \times 164$  and random left/right and up/down flips. Image augmentation is critical for training, and we anticipate that more aggressive routines could improve performance beyond what we report.

Since the 2d GRU had significantly more model parameters than the 1d GRU, we attempted to increase the fairness of the comparison between the two models by optimizing over 1d GRU models. This procedure, which was repeated for both single and dual loss versions of the 1d GRU, involved selecting the top-performing 1d GRU across 20 different configurations and 4 learning rates. This procedure was not repeated for the 2d GRU since it took significantly longer to train. Although our model optimization procedure is favorable to the 1d GRU, we still find a clear advantage for the 2d GRU over it in AP deconvolution (see Results).

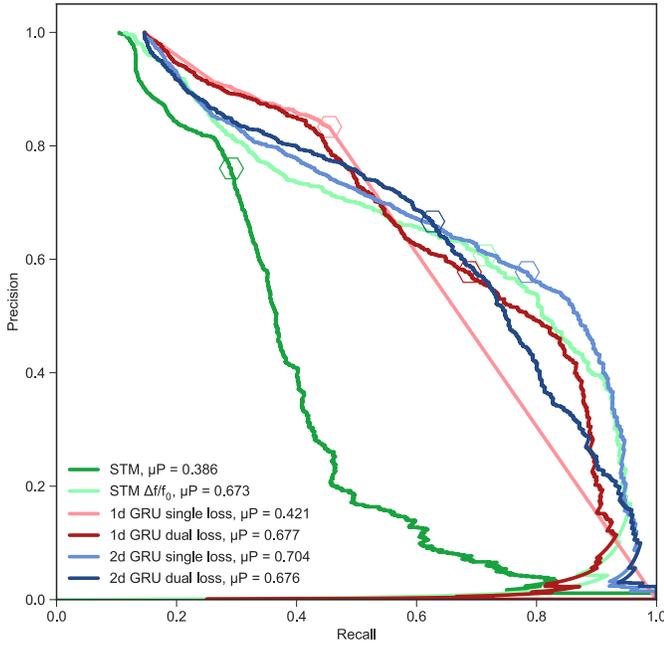


Fig. 3. End-to-end deconvolution on unprocessed GCaMP6 images outperforms typical approaches. Our novel 2d GRU achieves better average precision ( $\mu P$ ) than both a benchmark Spike Triggered Mixture (STM) model for deconvolution [16] and a similar GRU architecture trained on activity traces instead of images (“1d GRU”). The 2d models were trained on raw  $Ca^{2+}$  imaging data. Of the remaining models, most were trained on neuropil-corrected activity traces, with the exception of the STM  $\Delta f/f_0$ , which was trained on temporally normalized versions of the corrected traces. Precision is the ratio of true positives to all positive decisions. Recall is the ratio of true positives detections to total positive APs.  $\mu P$  is a weighted mean of precisions achieved across the depicted thresholds. Hexagons mark the point on each curve closest to the top-right hand corner, which is the optimal trade-off between precision and recall.

### B. Benchmark model

We used the spike-triggered mixture (STM) model for spike inference [16] as our baseline. The approach was shown to outperform other representative approaches, which included simple deconvolution as well as generative and discriminative approaches. Since our task was to detect the occurrence of an AP given a series of GCaMP6 images, and not to estimate the firing rate per se, we used a logistic non-linearity instead of an exponential one as done in the original study. The model assumes Bernoulli distributed APs given by the following equation:

$$\lambda(\bar{\mathbf{x}}_s) = \sum_{k=1}^K \text{logistic} \left( \sum_{m=1}^M \beta_{km} (\mathbf{u}_m^T \bar{\mathbf{x}}_s)^2 + \mathbf{w}_k^T \bar{\mathbf{x}}_s + \mathbf{b}_k \right), \quad (6)$$

where  $\bar{\mathbf{x}}_s(t)$  is a vector of preprocessed fluorescence values obtained by averaging over each of the frames in the image sequence  $\mathbf{x}_s(t)$ .  $\mathbf{w}_k$  and  $\mathbf{u}_m$  are a set of  $K$  linear and  $M$  quadratic features learned with supervision from training data. The quadratic feature term ensures that the model is expressive enough to learn complex non-linear relationships between calcium traces and spikes.

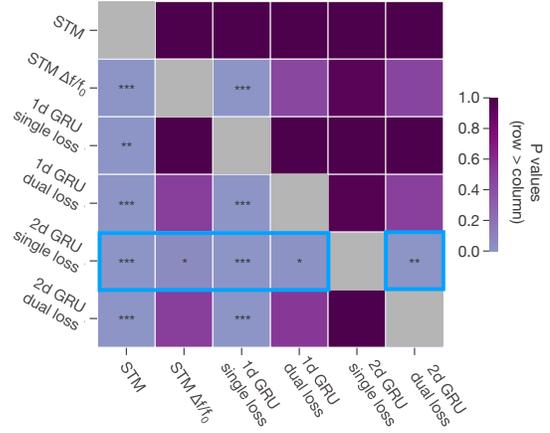


Fig. 4. Statistical comparison of model  $\mu P$  scores for AP deconvolution from  $Ca^{2+}$  activity. Cells are color coded to denote  $p$ -values derived from randomization tests that measured the likelihood that the performance of a model on a given row is superior to that of a model on a given column. The 2d GRU single loss model, outlined in blue, recorded a significantly higher  $\mu P$  score than any other model. \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ .

We used the same data preprocessing and hyper parameters as [16]. For increased robustness against potential local minima, we trained 5 different STM models and used the mode of their predictions for evaluation. Separate STMs were trained on datasets B and C. We validated this modified STM by training it on the same datasets as [16], for which we found comparable AP deconvolution performance ( $\rho = 0.49$ ) to what was reported in the paper [16] ( $\rho = 0.45$  to  $\rho = 0.55$ ).

## IV. EVALUATION

### A. Prediction accuracy

Fig. 3 shows precision-recall curves obtained for the different models for AP deconvolution. These curves show the trade-off between correctly detecting APs vs. detecting all APs in the dataset as the decision threshold is varied. An optimally performing model will have a curve approaching the top-right hand corner of the axes. The point on each curve closest to the top-right hand corner is marked with a hexagon. The markers show that the 2d GRU single loss model (light blue) is closest to this benchmark. These curves also demonstrate that the 2d GRUs outperform all other models at nearly all precision and recall thresholds except when recall is high, where the 1d GRUs outperform all other models.

We also summarized performance of the models as average precision (commonly known as AP, but referred to as  $\mu P$  here to avoid confusion with the abbreviation used for action potentials), which is a popular metric to evaluate computer vision systems (see e.g. [10]). We compared model performance in  $\mu P$  by entering these scores into randomization tests, which measured the probability of the difference in  $\mu P$  between any two models occurring due to chance. Briefly, this procedure involved first measuring the true difference in  $\mu P$  between any two models. Next, for each pair of predictions made by the two models, we randomly swapped the identities of their groups (having been produced by the STM vs. the 2d GRU) and

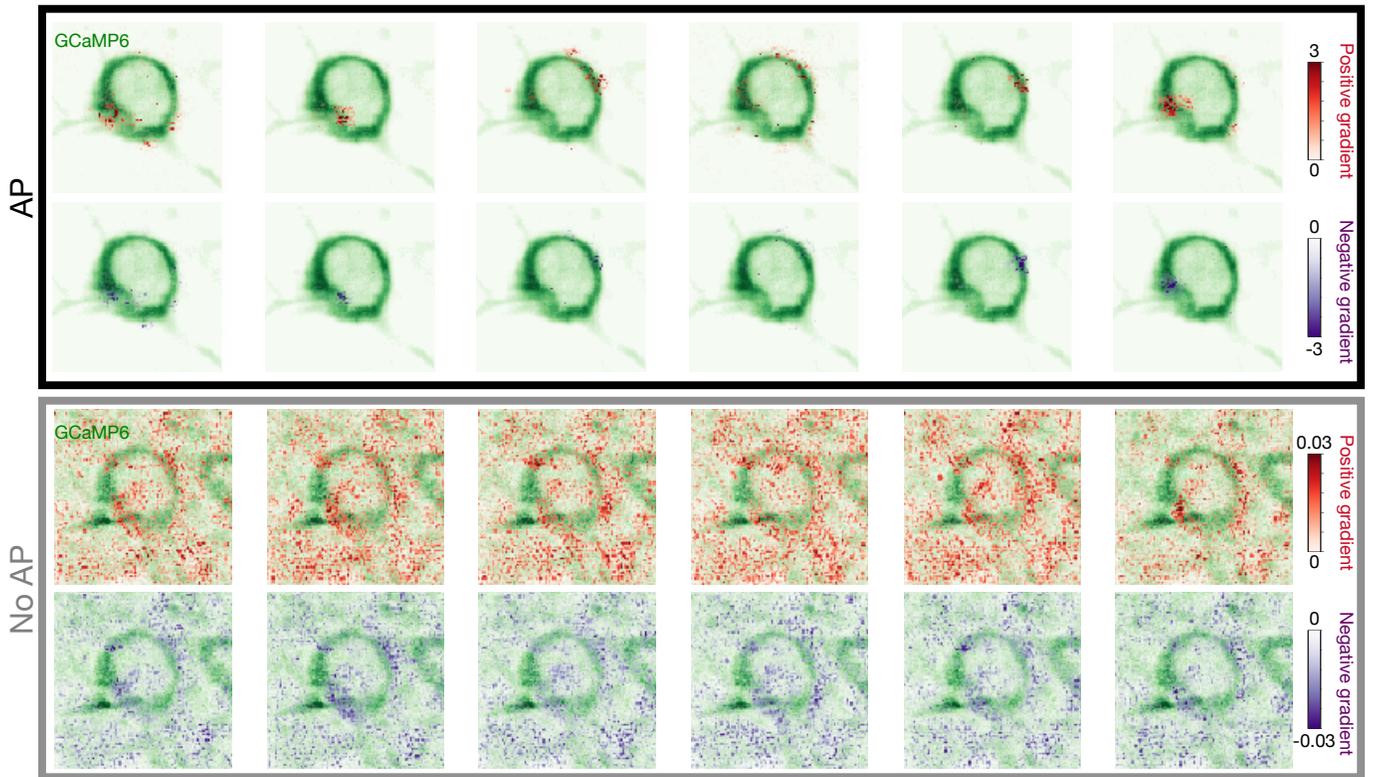


Fig. 5. The 2d GRU learns to associate calcium transients in the cytosol with APs. Calculating the gradient of activity in the 2d GRU decision layer for AP vs. No AP decisions with respect to GCaMP6 images yields maps that explain model decisions. Separately visualizing positive gradients from the negative gradients identifies pixels that positively (red overlay) and negatively (purple overlay) contribute to the model’s decisions for AP vs. No AP. Each panel depicts the mean GCaMP6 activity over different 166-ms sequences. The model identifies APs based on calcium transients in the cytosol over time, and in the process ignores background activity. When the model predicts No AP, it coincides with roughly uniform dispersion of gradient across the image with a notable absence of relevant signals in the cytosol. Colorbars on the right depict the range of gradient image values. Note that gradient images for sequences where an AP was detected are orders of magnitude larger than those where an AP was not detected, suggesting that the model made its decisions based on the amount of accumulated visual evidence for APs over time.

recorded the new difference in  $\mu P$  scores between them. This process was repeated 1,000 times to build a null distribution of  $\mu P$  differences between the models, and a  $p$  value was calculated as the proportion of randomized  $\mu P$  differences that exceeded the true difference.

This analysis revealed that the 2d GRU was significantly better at AP detection than any other model tested, including the baseline STM (4). Note that the STM only performed well when trained on pre-processed  $\Delta f/f_0$  signals as opposed to raw traces as done with the 1d GRU. Indeed, the 1d GRU performs similarly on raw traces to how the STM model performed on pre-processed  $\Delta f/f_0$  traces. This suggests that recurrent neural network models are capable of learning an appropriate normalization across time for AP deconvolution. This is one major benefit of the end-to-end learning approach as it circumvents the need to specify a temporal window by hand for computing  $\Delta f/f_0$ .

### B. Diagnostic features

What are the visual features present in  $Ca^{2+}$  image sequences that the 2d GRU model uses to predict APs? One approach that tries to identify the visual features that drive neural network decisions is based on the computation of

gradient images [15] which highlight individual pixel locations that influence model decisions. For our purpose, the procedure involved calculating the gradient of model decisions with respect to sequences of GCaMP6 images. We split sequences of gradient images into separate positive and negative gradient value volumes, which identified pixels based on their positive or negative contributions to the model’s decisions. Averaging across the multiple frames in each volume yielded heatmaps of positive and negative spatiotemporal feature importance.

Gradient images derived from the 2d GRU show that the model learned to focus on biologically meaningful visual features for AP deconvolution (Fig. IV-A): AP decisions seem to be associated with localized changes in cytosolic  $Ca^{2+}$ , while No-AP decisions are associated with gradients distributed nearly uniformly across the entire frame. In other words, significant cytosolic events trigger the detection of AP events while the absence of cytosolic events trigger No-AP event predictions. This demonstrates a key benefit of an end-to-end learning approach: the model learns how to focus on calcium transients to make its decision without manual preprocessing.

## V. DISCUSSION

We showed that end-to-end learning of AP deconvolution from  $\text{Ca}^{2+}$  imaging data is possible. This raises the exciting possibility that the set of manual heuristics developed for pre-processing  $\text{Ca}^{2+}$  imaging data may be learned entirely from data – bypassing the need for human intervention in the AP deconvolution pipeline. Time and time again, end-to-end learning with neural networks has proven capable of removing humans from the loop in various computer vision tasks.

We found that our novel 2d GRU trained to predict APs learns to base its decisions on meaningful regions in neurons, focusing on cytosolic  $\text{Ca}^{2+}$  transients. Our novel 2d GRU also exemplifies a toolbox of computations needed for this task: convolutional filters for extracting relevant spatial features and a memory (hidden state) for accumulating relevant signals possibly distributed across the cell body across time. More work is needed to optimize this model and answer many of the questions we could not answer because of a lack of data. For instance, training the model on longer image sequences could improve its deconvolution by exposing it to temporal dependencies in neuronal function that are more difficult to learn over the 166-ms image sequences used here.

Most importantly, the dataset used in the present study pales in size and variety compared to standard datasets used in modern computer vision and natural language processing benchmarks. While the described 2d GRU outperforms the state-of-the-art, it is clear that the model is underfitting. This is likely due to both the small number of unique image sequences for training, and the small proportion of these sequences in which an AP occurred. The latter limitation motivated us to train these models on AP detection across a brief image sequence rather than the more data intensive and functionally informative task of predicting the timing of APs. We also suspect that exposing the model to more neurons than the three tested here could facilitate our dual loss training routine, which gives the model an implicit signal for  $\text{Ca}^{2+}$  kinetics that are specific to different types of neurons. Given how computer vision has benefited from an emphasis on increasing dataset size, we expect that larger and more varied  $\text{Ca}^{2+}$  imaging datasets will significantly improve end-to-end learning for AP deconvolution and provide greater flexibility in task specification, especially for the neural networks we employ here (versus other methods like our STM baseline).

Rapid advancements in the design of better DCN architectures and training routines have bolstered the development of computer vision systems that can learn directly from images to solve visual tasks. We demonstrate that this approach can be readily applied to the unsolved task of deconvolving APs from  $\text{Ca}^{2+}$  imaging, outperforming the standard approach and positioning end-to-end learning as a potential way forward.

## ACKNOWLEDGMENT

This work was supported by NSF early career award [grant number IIS-1252951] and DARPA young faculty award [grant number YFA N66001-14-1-4037]. DL was supported in part by the Brown Institute for Brain Sciences and the Initiative for Computation in Brain and Mind.

## REFERENCES

- [1] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.*, 6(1):20–29, June 2004.
- [2] M. D. Bootman, K. Rietdorf, T. Collins, S. Walker, and M. Sanderson.  $\text{Ca}^{2+}$ -sensitive fluorescent dyes and intracellular  $\text{Ca}^{2+}$  imaging. *Cold Spring Harb. Protoc.*, 2013(2):83–99, Feb. 2013.
- [3] T.-W. Chen, T. J. Wardill, Y. Sun, S. R. Pulver, S. L. Renninger, A. Baohan, E. R. Schreier, R. A. Kerr, M. B. Orger, V. Jayaraman, L. L. Looger, K. Svoboda, and D. S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, July 2013.
- [4] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN Encoder-Decoder for statistical machine translation. June 2014.
- [5] T. Dozat. Incorporating nesterov momentum into adam. Feb. 2016.
- [6] GENIE project, Janelia Farm Campus, HHMI; Karel Svoboda (contact). Simultaneous imaging and loose-seal cell-attached electrical recordings from neurons expressing a variety of genetically encoded calcium indicators. 2015.
- [7] C. Grienberger and A. Konnerth. Imaging calcium in neurons. *Neuron*, 73(5):862–885, Mar. 2012.
- [8] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-Normalizing neural networks. June 2017.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [10] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Lawrence Zitnick, and P. Dollár. Microsoft COCO: Common objects in context. May 2014.
- [11] J. W. Linsley, I.-U. Hsu, L. Groom, V. Yarotsky, M. Lavorato, E. J. Horstick, D. Linsley, W. Wang, C. Franzini-Armstrong, R. T. Dirksen, et al. Congenital myopathy results from misregulation of a muscle  $\text{Ca}^{2+}$  channel by mutant *stac3*. *Proceedings of the National Academy of Sciences*, 114(2):E228–E236, 2017.
- [12] A. Miyawaki, O. Griesbeck, R. Heim, and R. Y. Tsien. Dynamic and quantitative  $\text{Ca}^{2+}$  measurements using improved cameleons. *Proc. Natl. Acad. Sci. U. S. A.*, 96(5):2135–2140, Mar. 1999.
- [13] N. O’Connor and R. B. Silver. Ratio imaging: practical considerations for measuring intracellular  $\text{Ca}^{2+}$  and  $\text{pH}$  in living cells. *Methods Cell Biol.*, 114:387–406, 2013.
- [14] O. Sadakane, Y. Masamizu, A. Watakabe, S.-I. Terada, M. Ohtsuka, M. Takaji, H. Mizukami, K. Ozawa, H. Kawasaki, M. Matsuzaki, and T. Yamamori. Long-Term Two-Photon calcium imaging of neuronal populations with subcellular resolution in adult non-human primates. *Cell Rep.*, 13(9):1989–1999, Dec. 2015.
- [15] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *ICLR Workshop*, 2014.
- [16] L. Theis, P. Berens, E. Froudarakis, J. Reimer, M. Román Rosón, T. Baden, T. Euler, A. S. Tolias, and M. Bethge. Benchmarking spike rate inference in population calcium imaging. *Neuron*, 90(3):471–482, May 2016.
- [17] R. Y. Tsien. New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures. *Biochemistry*, 19(11):2396–2404, May 1980.