OnACID: Online Analysis of Calcium Imaging Data in Real Time

J. Friedrich*, A. Giovannucci*, M. Kaufman, A.K. Churchland, D. Chklovskii, L. Paninski, E.A. Pnevmatikakis

Abstract

Optical imaging methods using calcium indicators are critical for monitoring the activity of large neuronal populations in vivo. Imaging experiments typically generate a large amount of data that needs to be preprocessed to extract the activity of the imaged neuronal sources. While deriving algorithms for this purpose is an active area of research, most existing methods require the processing of large amounts of data at a time, rendering them vulnerable to the volume of the recorded data, and preventing real-time experimental interrogation. Here we introduce OnACID, an Online framework for the Analysis of streaming Calcium Imaging Data, including i) motion artifact correction, ii) neuronal source extraction, and iii) activity denoising and deconvolution. Our approach combines and extends previous work on online dictionary learning and calcium imaging data analysis, to deliver an automated pipeline that can discover and track the activity of hundreds of cells in real time, thereby enabling new types of closed-loop experiments. We apply our algorithm on two large scale experimental datasets, benchmark its performance on manually annotated data, and show that it outperforms a popular offline approach.

Problem formulation. Calcium imaging methods enable simultaneous measurement of the activity of thousands of neighboring neurons. To infer the neural population activity from the raw imaging data, an analysis pipeline is employed which typically involves solving the following problems: i) correcting for motion artifacts during the imaging experiment, ii) identifying/extracting the sources (neurons and axonal or dendritic processes) in the imaged field of view (FOV), and iii) denoising and deconvolving the neural activity from the dynamics of the expressed calcium indicator. The fine spatiotemporal resolution of calcium imaging comes at a data rate cost; a typical two-photon (2p) experiment generates ∼50GB of data per hour. These rates can be significantly higher for other imaging techniques, e.g. 1TB per hour for light-sheet imaging [1]. A typical experiment can quickly produce datasets larger than the available RAM. The resulting data deluge poses a significant challenge. Further, current activity extraction methods are typically applied to imaging data after the experiment is complete. However, in many cases one would prefer to run closed-loop experiments [2] - analyzing data on-the-fly to guide the next experimental steps or to control feedback - and this requires new methods for accurate real-time processing.

Approach. We present an Online, single-pass, algorithmic framework for the Analysis of Calcium Imaging Data (OnACID). Our framework is highly scalable with minimal memory requirements, as it processes the data in a streaming fashion one frame at a time, while keeping in memory a set of low dimensional sufficient statistics and a small minibatch of the last data frames. Every frame is processed in four sequential steps: i) The frame is registered against the previous denoised (and registered) frame to correct for motion artifacts. ii) The fluorescence activity of the already detected sources is tracked. iii) Newly appearing neurons and processes are detected and incorporated to the set of existing sources. iv) The fluorescence trace of each source is denoised and deconvolved to provide an estimate of the underlying spiking activity.

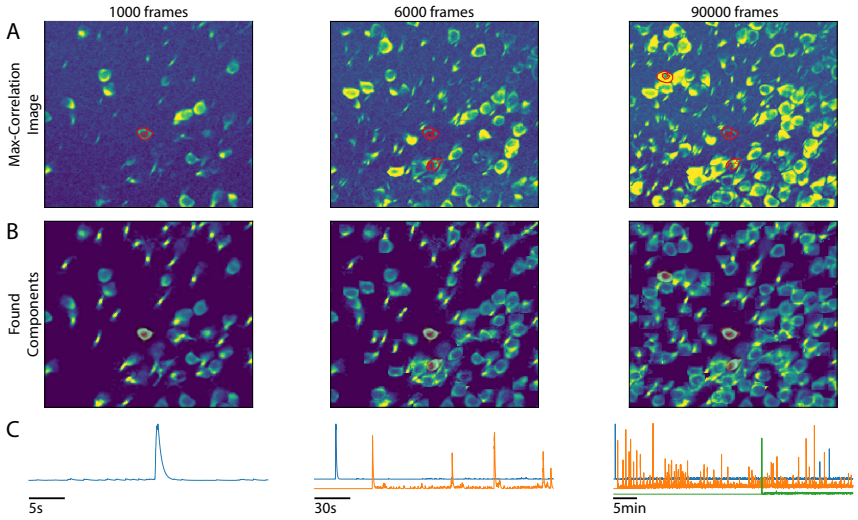

Figure 1: **Illustration of the online data analysis process.** Snapshots of the online analysis after processing 1000 frames (left), 6000 frames (middle), and 90000 frames (right). A) Max-correlation image of registered data at each snapshot point. B) Spatial footprints (shapes) of the components (neurons and processes) found by OnACID up to each point. C) Examples of neuron activity traces (marked by contours in panel A and highlighted in red in panel B). As the experiment proceeds, OnACID detects newly active neurons and tracks their activity. See also supplementary movie for an example in simulated data.

Our online approach allows us to employ a very simple yet effective motion correction scheme: each *denoised* dataframe can be used to register the next incoming noisy dataframe. A standard approach for source extraction is to model the fluorescence as product of a matrix A of spatial footprints and a matrix C of temporal activities. In the CNMF framework of [3] the matrices are further constraint by the spatial locality of each neuron and the calcium dynamics. CNMF operates by alternating optimization of A and C. This framework can be adapted to a data streaming setup using the online NMF algorithm of [4]. To further denoise and deconvolve the neural activity from the dynamics of the indicator we use the OASIS algorithm [5]. To account for a variable number of sources in an online NMF setting, we keep a buffer that contains the last instances of the residual signal and search for the point in space that explains the maximum variance. New candidate components $\mathbf{a}_{\text{new}}$, and $\mathbf{c}_{\text{new}}$ are estimated by performing a local rank-1 NMF of the residual matrix.

Results. We verified first on simulated data that the algorithm successfully detected and tracked all active sources, plus one false positive. To enable interesting closed loop experiments we need to extract spikes with a short time-lag. Lags of 2-5 yielded already similar results as the solution with unrestricted lag. By balancing the computational load over frames, the frame processing rate remained higher than 30Hz for each individual frame.

Next we considered a larger scale (116K frames, 512×512 pixels) real 2p calcium imaging dataset taken at 30Hz (64min long). We collected manual annotations from two independent labelers. The labelers found respectively 928 and 875 ROIs. OnACID found 752 sources (734 after filtering for size). We also compared with the CNMF algorithm applied to the whole dataset. To quantify performance we used a precision/recall framework. The online algorithm not only matches but outperforms the offline approach of CNMF, reaching high performance values ($F_1 = 0.76$ and 0.79 against the two manual annotations, as opposed to 0.65 and 0.66 for CNMF). The two annotations matched closely with each other ($F_1 = 0.89$), indicating high reliability. The matches and mismatches between OnACID and Labeler 1 on a 200×200 pixel part of the FOV are shown in Fig. 2A. Further, OnACID produced traces that are highly correlated with their offline counterparts, suggesting that the online algorithm can detect new neurons once they become active and then reliably track their activity. In addition to being more accurate, OnACID is also considerably faster as it required ∼48 minutes, i.e., $\sim 2\times$ faster than real time on average, to analyze the full dataset as opposed to ∼1.7 hours for the offline approach and ∼10 hours for each of the annotators (only spatial component selection). For this dataset, rigid motion correction was also performed according to the simple method of aligning each frame to the denoised (and registered) background from the previous frame. Fig. 2B shows that this approach produced strikingly similar results to an offline template based, rigid motion correction method [6].

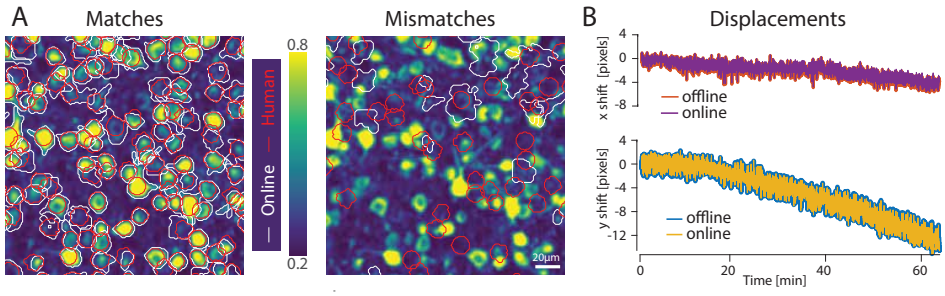

Figure 2: **Application to an *in vivo* 64min long parietal cortex dataset.** A-left) Matched inferred locations between the online algorithm (white) and the manual annotation of Labeler 1 (red). A-right) False positive (white) and false negative (red) mismatches between the online algorithm and a manual annotation. B) Displacement vectors estimated by OnACID during motion registration compared to a template based algorithm. OnACID estimates the same motion vectors at a sub-pixel resolution.

Conclusion. OnACID processes large calcium imaging datasets efficiently while requiring merely a minimal memory footprint. Significantly, the online fashion and speed of the algorithm enable real time identification and processing of hundreds of neurons in a typical 2p experiment, affording novel designs of closed-loop experiments.

[1] Ahrens et al. *Nat Methods*, 10(5):413–420, 2013.
[2] Grosenick et al. *Neuron*, 86(1):106–139, 2015.
[3] Pnevmatikakis et al. *Neuron*, 89(2):285–299, 2016.
[4] Mairal et al. *J Mach Learn Res*, 11:19–60, 2010.
[5] Friedrich et al. *PLoS Comput Biol*, 13(3):e1005423, 2017.
[6] Pnevmatikakis and Giovannucci. *J Neurosci Methods*, 291:83–94, 2017.