MIN1PIPE: A Miniscope 1-photon-based Calcium Imaging Signal Extraction Pipeline

Jinghao Lu[1*], Chunyuan Li[3*], Jonnathan Singh-Alvarado[1], Zhe Charles Zhou[4,5], Flavio Fröhlich[4,5], Richard Mooney[1], and Fan Wang[1,2*]

[1]Department of Neurobiology, Duke University Medical Center, Duke University, Durham NC 27708, USA
[2]Department of Cell Biology, Duke University Medical Center, Duke University, Durham NC 27708, USA
[3]Department of Electrical and Computer Engineering, Duke University, Durham NC 27708, USA
[4]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill NC 27599, USA.
[5]Neurobiology Curriculum, University of North Carolina at Chapel Hill, Chapel Hill NC 27599, USA

[*]Correspondence should be addressed to: Jinghao Lu (jinghao.lu@duke.edu), Chunyuan Li (chunyuan.li@duke.edu) and Fan Wang (fan.wang@duke.edu)

SUMMARY

In vivo calcium imaging using 1-photon based miniscope and microendoscopic lens enables studies of neural activities in freely behaving animals. However, the high and fluctuating background, the inevitable movements and distortions of imaging field, and the extensive spatial overlaps of fluorescent signals emitted from imaged neurons inherent in this 1-photon imaging method present major challenges for extracting neuronal signals reliably and automatically from the raw imaging data. Here we develop a unifying algorithm called MINiscope 1-photon imaging PIPEline (MIN1PIPE) that contains several standalone modules and can handle a wide range of imaging conditions and qualities with minimal parameter tuning, and automatically and accurately isolate spatially localized neural signals. We quantitatively compare MIN1PIPE with other existing partial methods using both synthetic and real datasets obtained from different animal models, and show that MIN1PIPE has a superior performance both in terms of efficiency and precision in analyzing noisy miniscope calcium imaging data.

INTRODUCTION

In vivo calcium imaging of activities from large populations of neurons in awake and behaving animals has become one of the staple technologies in neuroscience (Cai et al., 2016; Flusberg et al., 2008; Ghosh et al., 2011). Recent advances in single-photon based miniscope technology have further enabled imaging of neural ensemble activities in freely moving animals (Cai et al., 2016; Flusberg et al., 2008; Ghosh et al., 2011), thereby allowing circuits involved in a rich repertoire of animal behaviors to be examined. For example, this technology has been successfully used in probing dynamics of neural circuits involved in innate behaviors (Betley et al., 2015; Douglass et al., 2017; Jennings et al., 2015), decision making (Pinto and Dan, 2015; Poyraz et al., 2016), motor control (Klaus et al., 2017), learning and memory (Grewe et al., 2017; Kamigaki and Dan, 2017; Kitamura et al., 2017; Roberts et al., 2017; Roy et al., 2017; Xu et al., 2016), social memory (Okuyama et al., 2016), hippocampal place coding (Ziv et al., 2013), sleep (Cox et al., 2016; Weber and Dan, 2016), bird song (Markowitz et al., 2015), and pathological processes (Berdyyeva et al., 2016).

The increasing popularity of the miniscope calcium imaging technology demands the development of a fully automatic and robust signal processing method that can reliably extract neuronal signals from the noisy single photon calcium imaging data. Ideally, the processing method 1) should be able to handle a wide range of imaging conditions (*e.g.* high fluctuating background) and results with minimal parameter tuning, and 2) should have minimal assumptions about the quality of the data, such as free of movement or distortion, or sufficiently good signal-to-noise ratio (SNR). Existing imaging processing algorithms do not meet these two criteria.

For extraction of neuronal signals, many previous methods work well in situations with high SNR and stable field of views, therefore they are best suited for processing two-photon imaging data. These algorithms include linear unsupervised basis learning methods (Mukamel et al., 2009; Reidl et al., 2007; Pachitariu et al., 2013), nonlinear unsupervised methods (Maruyama et al., 2014; Pnevmatikakis et al., 2016), and the supervised learning method (Apthorpe et al., 2016). The PCA/ICA (principal component analysis followed by independent component analysis) method (Mukamel et al., 2009) was the first attempt to automatize the signal extraction process from miniscope imaging data through manual annotations of ROIs, but this method has difficulties in delineating localized ROIs and in separating overlapping ROIs. Since single-photon based imaging collect lights from a large depth of field, overlapping neurons in different depth is not uncommon. Another method called CNMF (constraint matrix factorization framework) (Pnevmatikakis et al., 2016), which combines nonlinearity in matrix factorization with simultaneous deconvolving spike trains from calcium dynamics, returns more spatially localized maps of ROIs compared to other methods and has better performance in identifying overlapping neurons. While methods like CNMF achieve plausible results in processing two-photon imaging data, they are ill-suited for processing the single-photon based miniscope imaging because: 1) the data from miniscope imaging are dominated by noisy, uneven and fluctuating background, 2) such methods depend on sophisticated parameter tuning, especially requiring setting parameters that are unknown *a priori* in practice such as “number of neurons”. Thus, a new method that can effectively remove background yet preserve real neural signals is highly desired.

Moreover, both PCA/ICA and CNMF rely on the stable imaging field and will fail if the imaging field contains movement, but movements (including distortions) during imaging, are often inevitable. Therefore, correcting movements and distortions is another major problem needs solving before neural activity signals can be reliably extracted. Many methods were developed independently in attempt to solve this problem. For example, several approaches register frames through template matching based on the assumption that the major form of movements is translational displacement (Dubbs et al., 2016; Thévenaz et al., 1998). To eliminate such assumptions, methods with block-based displacement field estimation were developed, with image feature matching algorithms extended from Lucas-Kanade tracker (Greenberg and Kerr, 2009; Lucas and Kanade, 1981) or Hidden Markov Models (Dombeck et al., 2007; Kaifosh et al., 2013). Some of these movement correction methods have been included as a module in a larger toolbox, using such frame-wise rigid registration approaches (Kaifosh et al., 2014; Pachitariu et al., 2017). When applied to handle nonrigid movement registrations, existing methods make specific assumptions about both the form and magnitude of the potential movements that result in suboptimal performance when large deformation occurs during imaging. Furthermore, errors in the movement correction can easily propagate since these methods all register frames based on a single reference frame. Considering that extracting neural activity signals is highly dependent on removing movement artifacts, an accurate and robust movement correction module is imperative. A simple unimodal algorithm for either translation/rigid or nonrigid registration is insufficient for this purpose.

Here we develop the MINiscope 1-photon imaging signal extraction PIPEline (MIN1PIPE) that takes the very raw calcium videos as inputs, and automatically removes background

while preserving signals, corrects movements with no assumptions of the types of movements, and delivers separated neuronal ROIs as well as deconvolved calcium traces as outputs. The MIN1PIPE contains a neural enhancing module that minimizes the influence of background unevenness and fluctuations, a hierarchical movement correction module that can handle all kinds of deformation with minimal error propagation, and a seeds-cleansed neural signal extraction module that identifies the set of real ROIs and their corresponding calcium traces without setting unknown parameters *a priori* (Fig. 1). Though our MIN1PIPE is primarily developed for single-photon based miniscope imaging, individual modules can also be independently combined with other processing algorithms to improve performance in analyzing two-photon imaging.

RESULTS

Core modules in MIN1PIPE

Neural Enhancing Module: The core of MIN1PIPE relies on turning the imaging data into a stack of background free, baseline corrected frames as the first step. Due to the complex spatiotemporal properties of background dynamics, our method applies a frame-wise background estimator that is adaptive to the local properties. A natural idea can be translated from mathematical morphology and computer vision, that the foreground neural signals can be approximated by subtracting the estimated background in a denoised image. Therefore, we first remove the grainy noise inherent to the single-photon system (see example of such grainy noise in Supplementary Fig. S2) while preserving the boundary between foreground and background, and this is achieved by applying an *anisotropic diffusion* denoising operation (Perona and Malik, 1990) on the raw imaging frames. Next, we use a simple straightforward *morphological opening* operation (Serra and Vincent,

1992) as the background estimator, with the size of the structure element similar to that of the neurons in the imaging field. The opening operation removes structures smaller than the desired structure element. Subsequently, the foreground that contains all the neuronal signals with the minimal noise is computed as the difference between the denoised raw and the morphological opened frames (Fig. 1a and Supplementary Fig. S1-S2).

Movement Correction Module: After the neural signals are enhanced, we next correct for movements in the imaging videos. The problem of movement correction can essentially be broken down to image stack registration. However, without setting specific constraints on the form or magnitude of the movements, even the most efficient registration algorithms require a running time on the order of seconds to minutes per frame (Vercauteren et al., 2009). Considering that the general imaging datasets contain tens of thousands of frames, the time required for applying these sophisticated image registration methods to every frame is inconceivable. Here we develop a hierarchical video registration framework for the correction of all types of movement, without sacrificing the precision or the speed of corrections. Our framework first decomposes the imaging video into two sections: the *stable sections* whose movements can be approximated by small translational displacement, and the *non-stable sections* that contain large general deformation. This step uses the KLT tracker that estimates the displacement of potential corner-like features between two neighboring frames (Shi and Tomasi, 1994). Next, we employ three levels of different strategies to align images. At the first level, we correct the small translational displacement within each stable section using the fast Lucas-Kanade tracker, which can be performed efficiently in parallel on multiple sections (Lucas et al., 1981). We then incorporate a diffeomorphic Log-Demons image registration method which can handle large

147 deformations while preserving the local geometrical properties (Vercauteren et al., 2009).
148 At the second level, we align all the stable sections. The overall information of each section
149 is extracted to form a sectional image. The current sectional image is then aligned to a
150 reference sectional image, which is generated as a linear summation of all previously
151 registered sectional images that is the closest to the current image. The summation weights
152 are determined by least square regression between the current and previous sectional
153 images. The estimated displacement field is then applied to each frame within the current
154 section. This will be iterated until all stable sections are aligned. At the third level, we use
155 a similar process to register the individual frames within each non-stable section, which
156 can be parallelized to boost performance efficiency (Fig. 1b). This hierarchical approach
157 significantly reduces the total registration time due to the balanced assignment of different
158 methods. Importantly, the common registration error does not propagate with this approach.

159 ***Seeds-Cleansed Neural Signal Extraction Module***: Once movements are corrected and
160 images are aligned, the main task turns to the neural signal extraction. MIN1PIPE extracts
161 neural signals automatically in two main steps, 1) the seeds cleansing step to reliably detect
162 the set of real ROIs, and 2) an altered spatiotemporal CNMF to separate ROIs and
163 corresponding calcium traces (Pnevmatikakis et al., 2016). Previous methods contain either
164 no explicit seeds initialization step or only a coarse initialization that compromises between
165 precision and recall. In contrast, our seeds cleansing step forces the algorithm to find the
166 set of real ROIs. This is achieved by first generating an over-complete set of seeds
167 containing all potential centers of real ROIs at the cost of including false positives. This
168 over-complete set is then coarsely cleansed by applying a two-component Gaussian
169 Mixture Model (GMM) on the peak-valley difference of corresponding traces of the seeds,

where the traces of real neurons usually have larger fluctuations compared to the non-neuron false positive seeds. The GMM removes most background false positives without losing real neurons (Supplementary Fig. S3). To further remove the remaining false positives, such as the ones with abnormal background fluctuations or hemodynamics, we have trained an Recurrent Neural Networks (RNNs) with LSTM module offline as the classifier for calcium spikes (Supplementary Fig. S4) (LeCun et al., 2015; Hochreiter and Schmidhuber, 1997). Those seeds whose traces contain RNN-identified calcium spikes, regardless of their temporal locations, are classified as true positives whereas the rest are deemed as false positives. After such cleansing processes, there is still a low possibility of identifying multiple seeds within a single ROI. Therefore, we merge potentially redundant seeds by computing the temporal similarity of seeds within their neighborhoods, and preserving the ones with maximum intensity. With the cleansed set of seeds as the initial position of ROIs, we next perform the iterative spatial and temporal optimizations, as proposed in CNMF, to update the spatial footprints of individual ROIs, and the temporal traces with deconvolved spike trains. Notably, unlike previous CNMF, where the spatial footprints are sequentially updated and subtracted from the preceding residuals, we extract spatial footprints from the original data that does not depend on preceding iterations. Therefore, the information loss/duplication is reduced and the optimization procedures can be parallelized in our method.

We show example results obtained using the MIN1PIPE methodology including a raw frame, the fully processed ROIs and the example traces from ROIs (Fig. 1c).

Quantitative validation of the MIN1PIPE performance

The neural signal extraction and movement correction are two independent problems that can be tested separately. For the signal extraction, we test the performance on synthetic datasets with various signal levels, while for the movement correction, we can directly test it on real data. To measure the performance of the signal extraction, we use a scoring metric that evaluates the spatial and temporal similarity between the ground truth and the identified ROIs (see Online Methods), and calculate true positive, false positive and false negative. To measure the performance of the movement correction, we use a metric based on the average displacement of feature points between neighboring frames.

We synthesized 16 imaging videos with signal levels (SL, defined as the ratio of the amplitude between the signals and the background) ranging from 0.05 to 0.8. Each video contains 3000 frames with 100 neurons of various shapes and calcium dynamics, and background fluctuations extracted from real datasets (details in Supplementary Notes S1-S2). The properties of synthetic videos resemble those of real data, whose SL falls between the range of 0.2 and 0.8. The condition at 0.05 SL is an extreme (Supplementary Fig. S5 and Video S1-S3). We compare MIN1PIPE with PCA/ICA and CNMF. We used commercially available *Mosaic* software (Inscopix Inc.) which implements PCA/ICA method (Mukamel et al., 2009), and processed the data following the standard workflow in the software manual. In particular, we chose the number of principal components (PC) and independent components (IC) based on the suggested rate (*e.g.* 20% more ICs and 50% more PCs than the estimated number of ROIs). For CNMF, we used the default initialization strategy in (Pnevmatikakis et al., 2016).

The results of ROI detection (Fig. 2a) and the calcium traces from one example ROI obtained by different methods (Fig. 2b) at 0.2 and 0.8 SL are compared. The contours of

the identified ROIs using different methods are drawn and superimposed onto the max projection of the ground truth. For both SLs, MIN1PIPE can identify the nearly complete set of ROIs (93% and 100% for 0.2 and 0.8 SL respectively) with minimal false positives (1% and 0% respectively), whereas the other two methods detect partial subsets of ROIs (PCA/ICA: 0% at 0.2 SL and ~65% true positives at 0.8 SL; CNMF: ~32% at 0.2 SL and ~95% true positives at 0.8 SL). In addition, the extracted spatial footprints are less realistic with the PCA/ICA or previous CNMF methods. The example calcium traces indicate that MIN1PIPE has a near-optimal performance in extracting individual ROIs even at low SLs when compared to the ground truth traces (Fig. 2b). In contrast, PCA/ICA completely fails to identify ROIs when SL is low, while CNMF fails to separate neurons from overlapping ROIs. Fig. 2c summarizes the performance accuracy of these three methods at all SLs. The plots of the true positive, false positive and false negative indicate that MIN1PIPE has a significantly superior performance in all conditions, and outperforms the other two methods in the extreme conditions. Fig. 2d summarizes the spatiotemporal similarities between the extracted results and the ground truth. The clusters of point clouds confirm again that MIN1PIPE outperforms the other two methods in best resembling both the spatial and temporal properties of the ground truth signals. The results for all signal levels are shown in Supplementary Fig. S6.

To validate the performance of the movement correction in MIN1PIPE, we applied the module on video sections with large deformation movements of the imaging field. Specifically, we chose a video obtained through two-photon imaging of the ferret's posterior parietal cortex as an example due to its particularly frequent and large deformations (Supplementary Video S4 as a demo section of the full video). Fig. 2e shows

an example of 3 consecutive frames (with each frame pseudo-colored as pink, green or blue) with large nonlinear deformations between frames superimposed together. The two rigid-transform-based methods (LK and KLT Affine) by themselves fail to fully remove the large deformations, whereas MIN1PIPE (combining the Log-Demons transformation) succeeds in correcting these movements. We further quantified the extent of correction by calculating the average displacement of feature points between two neighboring frames before and after the movement correction (Fig. 2f). Before correction, the score of average displacement shows frequent burst of large deformation periods, whereas after correction, the score of displacement shows a nearly flat line. Notably, the frames with large deformations are all well aligned (Fig. 2f lower panels; Supplementary Video S4).

Application of MIN1PIPE on real miniscope imaging datasets

We next compare the performance of MIN1PIPE with PCA/ICA and CNMF methods on real datasets. We first applied the three methods to the miniscope calcium imaging data obtained using prism probe from layer 2 and 3 of the barrel cortex in freely moving mice. GCaMP6f was expressed in layer 2/3 neurons using AAV, and signals were imaged continuously over 5min when the mouse freely explored its environment. Following the general pipeline of MIN1PIPE (Online Methods), our method removed the strong uneven background structure, and automatically identified 210 putative ROI components without the need for additional manual selection (Fig. 3a, Supplementary Video S5). In comparison, PCA/ICA identified 79 ROIs whereas CNMF identified 71 ROIs. Note that with the same computer configuration and dataset (~28 GB), the CNMF ran into memory issues and could not process the full-scale video, thus we cropped a center patch as the input video to the CNMF (Fig 3b-d middle panels). The contours of the ROIs identified by the different

methods are drawn and superimposed onto the max projection of the neural enhanced data (Fig. 3b). This reveals that MIN1PIPE can identify potentially all ROIs, whereas PCA/ICA and CNMF miss a significant subset of ROIs. Meanwhile, both the PCA/ICA and CNMF have some problems in separating overlapping ROIs and/or estimating the correct shape of the ROIs, as revealed by the max projections of the extracted signals (Fig. 3c). The projection of MIN1PIPE extracted signals closely resembles those of the neural enhanced data (Fig. 3c, top). In comparison, the projection obtained using PCA/ICA has low SNR with high background signals (Fig. 3c, bottom), whereas the projection derived from CNMF shows unrealistic ROI shapes larger than the true shape of neurons, indicating that the CNMF is sensitive to the contamination of the background dynamics (Fig. 3c, middle). To further check the shape of individual ROIs, we choose to visualize one example ROI embedded in the full imaging field using different methods (Fig. 3d). MIN1PIPE delineates a well localized ROI footprint, whereas CNMF returns a less localized ROI difficult to relate to the underlying neuron. PCA/ICA, on the other hand, fails to provide a localized footprint as remnants of other ROI components can also be seen. To examine the temporal traces extracted by the different methods, we selected 10 ROI components that were identified by all three methods within the cropped image field and plotted their corresponding calcium traces (Fig 3e, individual panels marked with 1-10). Again, the spatial footprints of the 10 ROIs obtained through CNMF and PCA/ICA are less localized as described above. In terms of calcium dynamics, both MIN1PIPE and CNMF give denoised traces, whereas the traces obtained using PCA/ICA are noisy and show unrealistic negative fluctuations. Furthermore, the traces extracted using CNMF include false positive calcium events that likely reflect background noises (arrows in Fig. 3e). Supplementary

Video S6 shows the comparison of raw and processed results using three different methods of the entire imaging video.

To further illustrate the general applicability of MIN1PIPE in processing miniscope imaging data obtained over different brain areas and/or different animal models, we applied it to process calcium imaging results from Area X in zebra finch and compared the results with those obtained using the other two methods. Briefly, MIN1PIPE detected 55 ROI components, whereas CNMF detected 15 and PCA/ICA detected 35 ROIs after manual selection (Fig. 4a-b, Supplementary Video S7). Again, CNMF could only process a cropped portion of the imaging video due to the computer memory issues. All of the ROIs detected by CNMF and PCA/ICA are included in the set extracted by MIN1PIPE. In addition to false negatives, CNMF also identifies a cluster of false positive ROIs that are apparent upon visual inspections (Fig. 4b-c, e). The PCA/ICA again gives rise to nonlocalized ROI footprints that contain other potential components (Fig. 4d). It was known that Area X neurons show strong song selective activities when the bird is singing (Goldberg and Fee, 2010; Kojima and Doupe, 2007; Woolley et al., 2014; Yazaki-Sugiyama and Mooney, 2004), which can be used as partial ground truth to validate MIN1PIPE. We plot the calcium traces of the ROIs identified by MIN1PIPE, and the majority of the ROIs contain calcium events that are roughly phase-locked to the singing onsets (Fig. 4f). Notably, a subset of the neurons shows precise singing-related activities with minimal events unrelated to singing (Fig. 4g upper panel). Furthermore, we sorted neurons according to their timing of peak calcium activities during each song production event (Fig. 4g lower panel), and this analysis also reveals a subset of Area X neurons whose

306 activation patterns are closely related to the onset of the singing, consistent with previous
307 findings.

308 We also did studies (Supplementary Fig. S7-S8) showing the effectiveness of the two
309 modules (neural enhancing and seeds-cleansed signal extraction) when combined
310 separately with previous methods to improve the performance. Finally, we compare our
311 method with the CNMF-E method (Zhou et al., 2016) (Supplementary Fig. S9).

312 **DISCUSSION**

313 The key advances that set MIN1PIPE apart from the previous imaging processing methods
314 are the following. First, MIN1PIPE solves the full range of problems for signal extraction
315 in single-photon miniscope imaging with one pipeline. Specifically, we have developed
316 innovative and robust modules to solve different problems including: the neural enhancing
317 module that uses a novel algorithm to remove the ultra-high and fluctuating background
318 characteristic of single photon imaging, the novel hierarchical movement correction
319 module that is capable of efficiently registering any types of deformations of the imaging
320 field, and the seeds-cleansed neural signal extraction module that utilizes GMM and
321 pretrained RNN to enable automatic identification and extraction of ROIs and calcium
322 traces. Second, MIN1PIPE eliminates the need for heuristically setting many parameters
323 that are not only unknown *a priori*, but also influence the performance of the downstream
324 processing steps, such as setting the number of neurons, a central parameter required by all
325 previous neural signal extraction methods. The importance of this should not be neglected,
326 because pre-setting unknown parameters can become problematic in practice. For example,
327 overestimating the number of neurons may result in false positives in identified ROIs
328 before the calcium trace extraction step, and also result in the unnecessary consumption of

329 computing time. These false positives can only be removed with laborious manual selection
330 without a robust seeds-cleansing step. On the other hand, underestimating the number of
331 neurons will likely lead to false negatives that can never be identified by the downstream
332 steps. Therefore, tweaking this parameter is inevitable in practice using previous methods.
333 While we do not claim that MIN1PIPE completely eliminates the need for manually
334 pruning identified ROIs, our method does only involve minimal manual interference. Third,
335 MIN1PIPE contains a minimal set of parameters that are easy to interpret and error-tolerant
336 (Online Methods). These include a set of simplified and fixed parameters applicable to
337 various conditions of the popular miniscope platforms (*e.g.* Inscopix nVista, UCLA open-
338 source miniscope) in our brick algorithm. In addition, all modules use definitive criteria
339 independent of various imaging datasets, which ensures the robustness that the previous
340 combination of setting the number of neurons and the serial initialization procedure could
341 not provide.

342 In summary, MIN1PIPE provides a generally applicable high-performance toolbox with
343 the modular framework to handle and process miniscope imaging data. Interesting future
344 works may integrate more advanced methods to further improve the precision, such as
345 stricter choices of the kernel of anisotropic diffusion (Chen et al., 2011; Tsiotsios and
346 Petrou, 2013), considering spectral invariants to handle very large deformations during
347 non-rigid registration (Lombaert et al., 2014), and using more biologically valid calcium
348 dynamics deconvolution methods (Speiser et al., 2017). Additionally, a more robust RNN
349 classifier for seeds cleansing can be trained with more available datasets.

ACKNOWLEDGEMENT

We thank the suggestions and comments from members of the Wang Lab and Mooney Lab at Duke University. The study is supported by NIH grants DP1-MH103908, R21NS101441, NS077986 to F.W.

AUTHOR CONTRIBUTIONS

J.L. and F.W. conceived and designed the project; J.L. designed and implemented all modules of MIN1PIPE; C.L. designed the RNN classifier; J.L. generated the synthetic data; J.L., F.W., J.S.-A., R.M., Z.Z. and F.F. designed and/or conducted in vivo imaging experiments; J.L. and C.L. analyzed the data; J.L., C.L., F.W. wrote the manuscript.

359 **REFERENCES**

360 Apthorpe, N., Riordan, A., Aguilar, R., Homann, J., Gu, Y., Tank, D. and Seung, H.S.
361 (2016). Automatic neuron detection in calcium imaging data using convolutional networks.
362 In *Proc. Advances in Neural Information Processing Systems. 29*, 3270-3278.

363 Berdyyeva, T.K., Frady, E.P., Nassi, J.J., Aluisio, L., Cherkas, Y., Otte, S., Wyatt, R.M.,
364 Dugovic, C., Ghosh, K.K., Schnitzer, M.J., et al. (2016). Direct Imaging of Hippocampal
365 Epileptiform Calcium Motifs Following Kainic Acid Administration in Freely Behaving
366 Mice. *Front. Neurosci. 10*, 53.

367 Betley, J.N., Xu, S., Cao, Z.F.H., Gong, R., Magnus, C.J., Yu, Y. and Sternson, S.M. (2015).
368 Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature 521*,
369 180-185.

370 Cai, D.J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M.,
371 La-Vu, M., Lou, J., et al. (2016). A shared neural ensemble links distinct contextual
372 memories encoded close in time. *Nature 534*, 115-118.

373 Chen, D., MacLachlan, S. and Kilmer, M. (2011). Iterative parameter-choice and multigrid
374 methods for anisotropic diffusion denoising. *SIAM J. Sci. Comput. 33*, 2972-2994.

375 Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and
376 Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for
377 statistical machine translation. In *Proc. Conference on Empirical Methods in Natural*
378 *Language Processing*, 1724-1734.

379 Cox, J., Pinto, L. and Dan, Y. (2016). Calcium imaging of sleep–wake related neuronal
380 activity in the dorsal pons. *Nat. Commun. 7*, 10763.

381 Ding, C.H., Li, T. and Jordan, M.I. (2010). Convex and semi-nonnegative matrix
382 factorizations. *IEEE Trans. Pattern Analysis and Machine Intelligence 32*, 45-55.

383 Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L. and Tank, D.W. (2007).
384 Imaging large-scale neural activity with cellular resolution in awake, mobile
385 mice. *Neuron 56*, 43-57.

386 Douglass, A.M., Kucukdereli, H., Ponserre, M., Markovic, M., Gründemann, J., Strobel,
387 C., Morales, P.L.A., Conzelmann, K.K., Lüthi, A. and Klein, R. (2017). Central amygdala
388 circuits modulate food consumption through a positive valence mechanism. *bioRxiv*,
389 145375.

390 Dubbs, A., Guevara, J. and Yuste, R. (2016). moco: Fast motion correction for calcium
391 imaging. *Front. Neuroinform. 10*, 6.

392 Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann.*
393 *Stat. 32*, 407-499.

394 Flusberg, B.A., Nimmerjahn, A., Cocker, E.D., Mukamel, E.A., Barretto, R.P., Ko, T.H.,
395 Burns, L.D., Jung, J.C. and Schnitzer, M.J. (2008). High-speed, miniaturized fluorescence
396 microscopy in freely moving mice. *Nat. Methods 5*, 935.

397 Ghosh, K.K., Burns, L.D., Cocker, E.D., Nimmerjahn, A., Ziv, Y., El Gamal, A. and
398 Schnitzer, M.J. (2011). Miniaturized integration of a fluorescence microscope. *Nat.*
399 *Methods 8*, 871-878.

Goldberg, J. H., and Fee, M. S. (2010). Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia. *J. Neurophysiol. 103*, 2002-2014.

Greenberg, D.S. and Kerr, J.N. (2009). Automated correction of fast motion artifacts for two-photon imaging of awake animals. *J. Neurosci. Methods 176*, 1-15.

Grewe, B.F., Gründemann, J., Kitch, L.J., Lecoq, J.A., Parker, J.G., Marshall, J.D., Larkin, M.C., Jercog, P.E., Grenier, F., Li, J.Z., et al. (2017). Neural ensemble dynamics underlying a long-term associative memory. *Nature 543*, 670-675.

Hahn, T.T., Sakmann, B. and Mehta, M.R. (2006). Phase-locking of hippocampal interneurons' membrane potential to neocortical up-down states. *Nat. Neurosci 9*, 1359-1361.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput. 9*, 1735-1780.

Jennings, J.H., Ung, R.L., Resendez, S.L., Stamatakis, A.M., Taylor, J.G., Huang, J., Veleta, K., Kantak, P.A., Aita, M., Shilling-Scrivo, K., et al. (2015). Visualizing hypothalamic network dynamics for appetitive and consummatory behaviors. *Cell 160*, 516-527.

Kaifosh, P., Lovett-Barron, M., Turi, G.F., Reardon, T.R. and Losonczy, A. (2013). Septo-hippocampal GABAergic signaling across multiple modalities in awake mice. *Nat. Neurosci. 16*, 1182-1184.

Kaifosh, P., Zaremba, J.D., Danielson, N.B. and Losonczy, A. (2014). SIMA: Python software for analysis of dynamic fluorescence imaging data. *Front. Neuroinform. 8*, 80.

Kamigaki, T. and Dan, Y. (2017). Delay activity of specific prefrontal interneuron subtypes modulates memory-guided behavior. *Nat. Neurosci.* 20, 854-863.

Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M., Redondo, R.L. and Tonegawa, S. (2017). Engrams and circuits crucial for systems consolidation of a memory. *Science 356*, 73-78.

Klaus, A., Martins, G.J., Paixao, V.B., Zhou, P., Paninski, L. and Costa, R.M. (2017). The spatiotemporal organization of the striatum encodes action space. *Neuron 95*, 1171-1180.

Kojima, S. and Doupe, A.J. (2007). Song selectivity in the pallial-basal ganglia song circuit of zebra finches raised without tutor song exposure. *J. Neurophysiol*. 98, 2099-2109.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature 521*, 436-444.

Lombaert, H., Grady, L., Pennec, X., Ayache, N. and Cheriet, F. (2014). Spectral log-demons: diffeomorphic image registration with very large deformations. *Int. J. Comput. Vis. 107*, 254-271.

Lucas, B.D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc. 7th International Joint Conference on Artificial Intelligence*, 674-679.

Markowitz, J.E., Liberti III, W.A., Guitchounts, G., Velho, T., Lois, C. and Gardner, T.J. (2015). Mesoscopic patterns of neural activity support songbird cortical sequences. *PLoS Biol. 13*, e1002158.

441 Maruyama, R., Maeda, K., Moroda, H., Kato, I., Inoue, M., Miyakawa, H. and Aonishi, T.
442 (2014). Detecting cells using non-negative matrix factorization on calcium imaging
443 data. *Neural Networks 55*, 11-19.

444 Mukamel, E.A., Nimmerjahn, A. and Schnitzer, M.J. (2009). Automated analysis of
445 cellular signals from large-scale calcium imaging data. *Neuron 63*, 747-760.

446 Okuyama, T., Kitamura, T., Roy, D.S., Itohara, S. and Tonegawa, S. (2016). Ventral CA1
447 neurons store social memory. *Science 353*, 1536-1541.

448 Pachitariu, M., Packer, A.M., Pettit, N., Dalgleish, H., Hausser, M. and Sahani, M. (2013).
449 Extracting regions of interest from biological images with convolutional sparse block
450 coding. In *Proc. Advances in Neural Information Processing Systems 26*, 1745-1753.

451 Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L.F., Carandini, M. and
452 Harris, K.D. (2016). Suite2p: beyond 10,000 neurons with standard two-photon
453 microscopy. *bioRxiv*, 061507.

454 Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic
455 diffusion. In *IEEE Trans. Pattern Analysis and Machine Intelligence 12*, 629-639.

456 Pinto, L. and Dan, Y. (2015). Cell-type-specific activity in prefrontal cortex during goal-
457 directed behavior. *Neuron 87*, 437-450.

458 Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon,
459 T., Mu, Y., Lacefield, C., Yang, W., et al. (2016). Simultaneous denoising, deconvolution,
460 and demixing of calcium imaging data. *Neuron 89*, 285-299.

461 Poyraz, F.C., Holzner, E., Bailey, M.R., Meszaros, J., Kenney, L., Kheirbek, M.A., Balsam,
462 P.D. and Kellendonk, C. (2016). Decreasing striatopallidal pathway function enhances
463 motivation by energizing the initiation of goal-directed action. *J. Neurosci. 36*, 5988-6001.

464 Reidl, J., Starke, J., Omer, D.B., Grinvald, A. and Spors, H. (2007). Independent
465 component analysis of high-resolution imaging data identifies distinct functional
466 domains. *Neuroimage 34*, 94-108.

467 Roberts, T.F., Hisey, E., Tanaka, M., Kearney, M.G., Chattree, G., Yang, C.F., Shah, N.M.
468 and Mooney, R. (2017). Identification of a motor-to-auditory pathway important for vocal
469 learning. *Nat. Neurosci.* 20, 978-986.

470 Roy, D.S., Kitamura, T., Okuyama, T., Ogawa, S.K., Sun, C., Obata, Y., Yoshiki, A. and
471 Tonegawa, S. (2017). Distinct neural circuits for the formation and retrieval of episodic
472 memories. *Cell 170*, 1000-1012.

473 Serra, J. and Vincent, L. (1992). An overview of morphological filtering. *Circuits, Systems,*
474 *and Signal Processing 11*, 47-108.

475 Shi, J. and Tomasi, C. (1994). Good features to track. *IEEE Conference on Computer*
476 *Vision and Pattern Recognition*, 593-600.

477 Speiser, A., Yan, J., Archer, E.W., Buesing, L., Turaga, S.C. and Macke, J.H. (2017).
478 Fast amortized inference of neural activity from calcium imaging data with variational
479 autoencoders. In *Proc. Advances in Neural Information Processing Systems 30*, 4027-
480 4037.

481 Thévenaz, P., Ruttimann, U.E. and Unser, M. (1998). A pyramid approach to subpixel
482 registration based on intensity. *IEEE Trans. Image Processing 7*, 27-41.

483 Tsiotsios, C. and Petrou, M. (2013). On the choice of the parameters for anisotropic
484 diffusion in image processing. *Pattern Recognition 46*, 1369-1381.

485 Van Den Boomgaard, R. and Van Balen, R. (1992). Methods for fast morphological image
486 transforms using bitmapped binary images. *CVGIP: Graphical Models and Image*
487 *Processing 54*, 252-258.

488 Vercauteren, T., Pennec, X., Perchant, A. and Ayache, N. (2009). Diffeomorphic demons:
489 Efficient non-parametric image registration. *Neuroimage 45*, 61-72.

490 Weber, F. and Dan, Y. (2016). Circuit-based interrogation of sleep control. *Nature 538*,
491 51-59.

492 Woolley, S. C., Rajan, R., Joshua, M., and Doupe, A. J. (2014). Emergence of context-
493 dependent variability across a basal ganglia network. *Neuron 82*, 208-223.

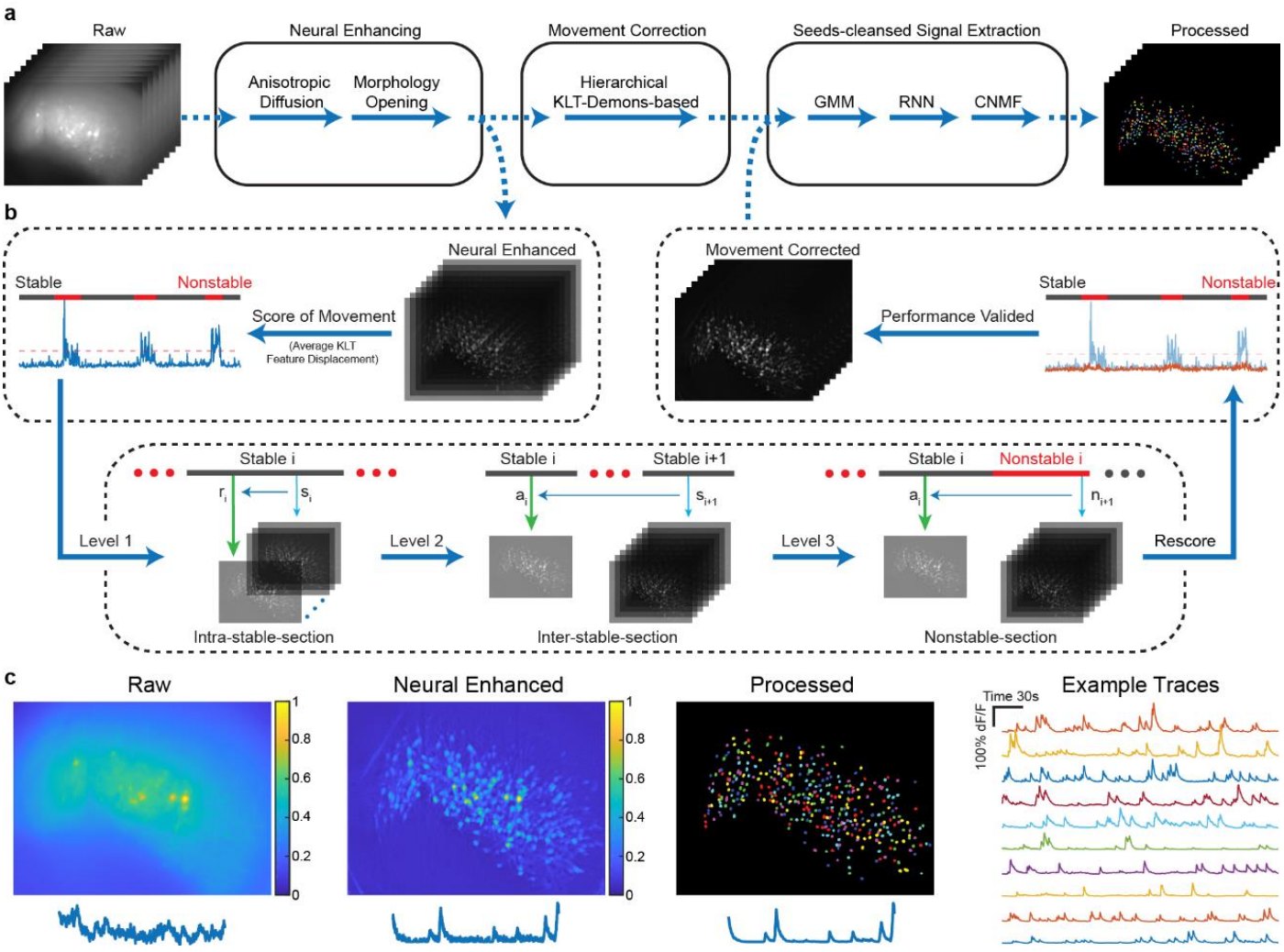
494 Xu, C., Krabbe, S., Gründemann, J., Botta, P., Fadok, J.P., Osakada, F., Saur, D., Grewe,
495 B.F., Schnitzer, M.J., Callaway, E.M., et al. (2016). Distinct hippocampal pathways
496 mediate dissociable roles of context in memory retrieval. *Cell 167*, 961-972.

497 Yazaki-Sugiyama, Y. and Mooney, R. (2004). Sequential learning from multiple tutors and
498 serial retuning of auditory neurons in a brain area important to birdsong learning. *J.*
499 *Neurophysiol, 92*, 2771-2788.

Zhou, P., Resendez, S.L., Stuber, G.D., Kass, R.E. and Paninski, L. (2016). Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *arXiv*, 1605.07266.

Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A. and Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci. 16*, 264-266.

506 Figure 1

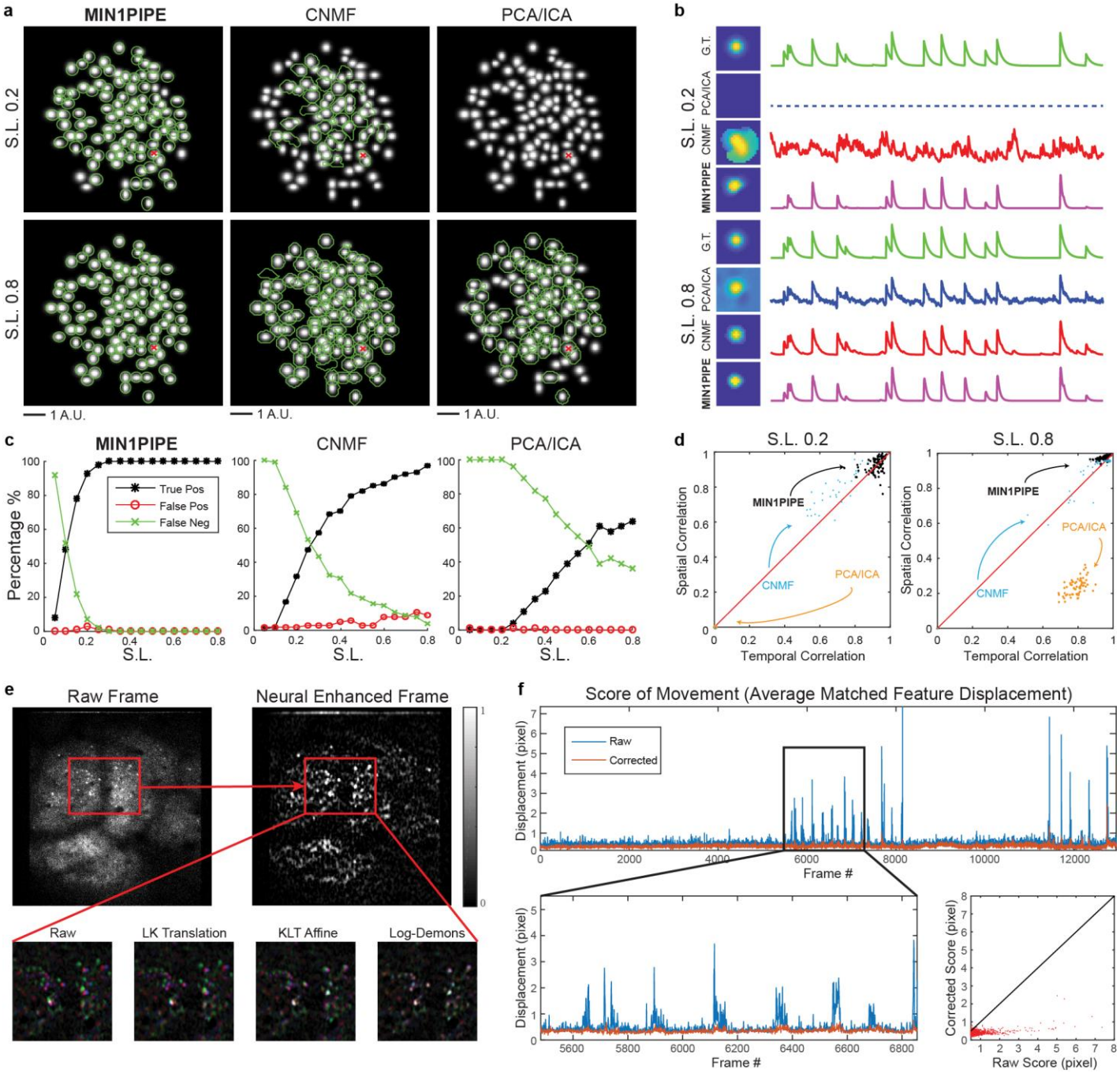


507

508 **Fig. 1.** The general pipeline and demonstrations of MIN1PIPE. **a.** The overall structure of
509 MIN1PIPE. MIN1PIPE takes the very raw miniscope imaging data freshly collected from
510 the imaging system as inputs, and returns fully processed ROI components with spatial
511 footprints and temporal calcium traces as outputs. The data are processed in series by neural
512 enhancing, hierarchical movement correction, and seeds-cleansed signal extraction
513 modules. Each module is composed of specific brick functions. **b.** A zoom-in of the
514 hierarchical movement correction module. The module is KLT-Log Demons based. It first
515 scores all the neural enhanced frames and divides them into stable and nonstable sections,

then register frames at three different levels. The movement corrected frames are then fed into the seeds-cleansed neural signal extraction module. **c.** The max projecting demonstrations of MIN1PIPE. The raw imaging video contains large, dominant background fluctuation, the neural enhanced video contains mainly neural signals, and the fully processed video contains only extracted and denoised neural signals as independent ROI components. Example traces are randomly selected from the processed video. See also Fig. S1-S4.

523 Figure 2

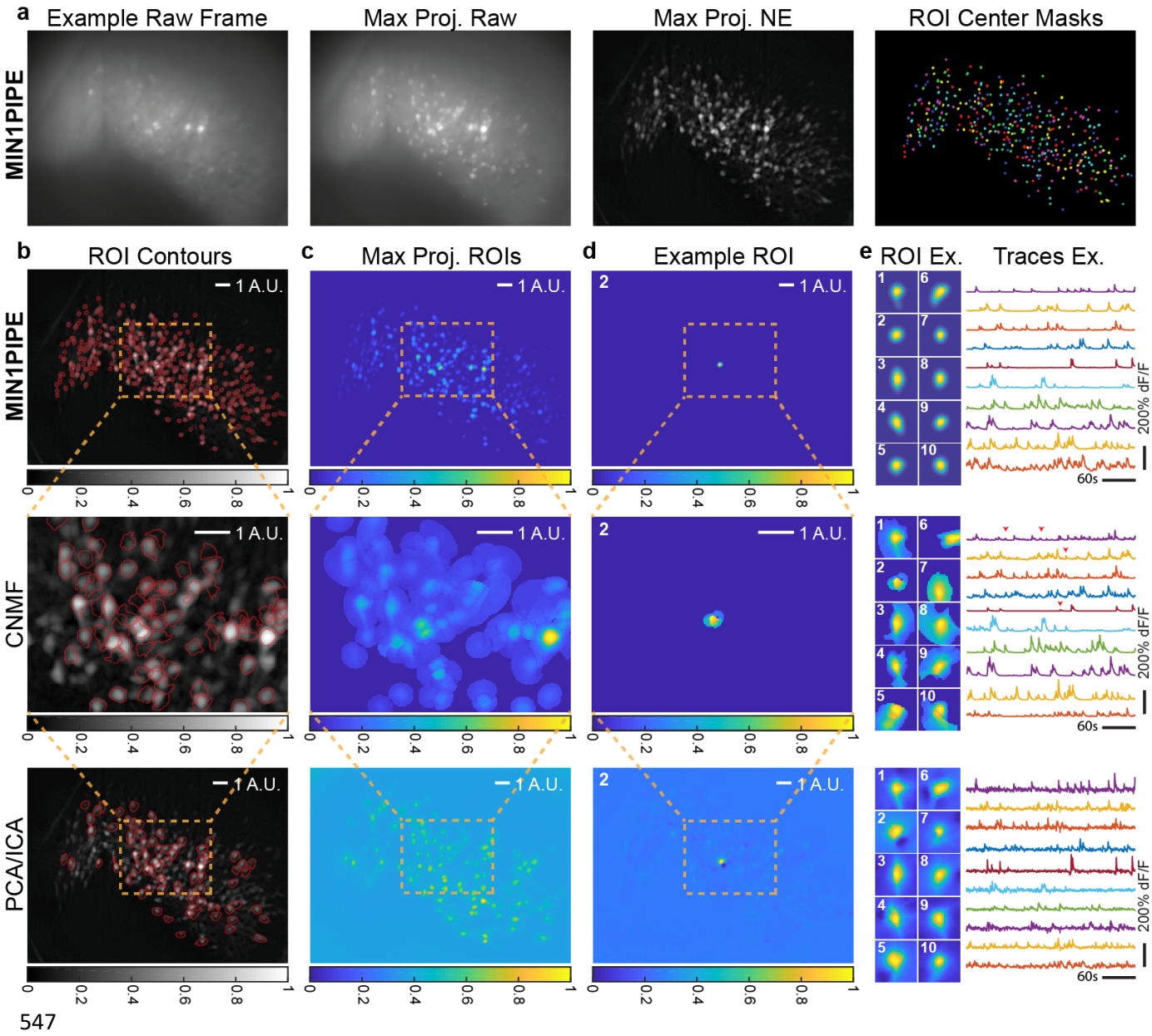


524

525 **Fig. 2.** Quantification of MIN1PIPE. **a-d.** The quantitative performance comparison
526 between MIN1PIPE and the other methods on simulated datasets. **a.** The identified ROI
527 contours of the three methods at two representative signal levels. S.L.: signal level (the

ratio of the amplitude between the signals and the background). A.U.: arbitrary unit. **b.** The trace of an example ROI component identified by the three methods and the ground truth. G.T.: ground truth. **c.** The identification precision and accuracy of the three methods. **d.** The spatiotemporal identification accuracy with the three methods at two representative signal levels. The similarity of the spatial footprints and temporal traces between the results of the three methods and the ground truth of each ROI is plotted as a dot in the figure. These quantifications confirm that MIN1PIPE not only outperforms the other two methods in all aspects, but also verify that MIN1PIPE can provide satisfactory neural signal extraction results of miniscope imaging data. **e-f.** The quantification of the movement correction module. **e.** The demonstration of various image registration methods in a consecutive three frames from two-photon imaging data. The deformation registration algorithm we use in the module (Log-Demons) successfully corrects all the nonrigid deformations within the frames that cannot be corrected by the methods that assume translation or rigid transformation. The black and white color indicate overlapping of the same structure between the frames, whereas other colors indicate nonaligned structures. **f.** The score of movement before and after the correction. In general, the hierarchical correction steps register large deformations while preserves the stable frames. See also Fig. S5-S9.

546 Figure 3

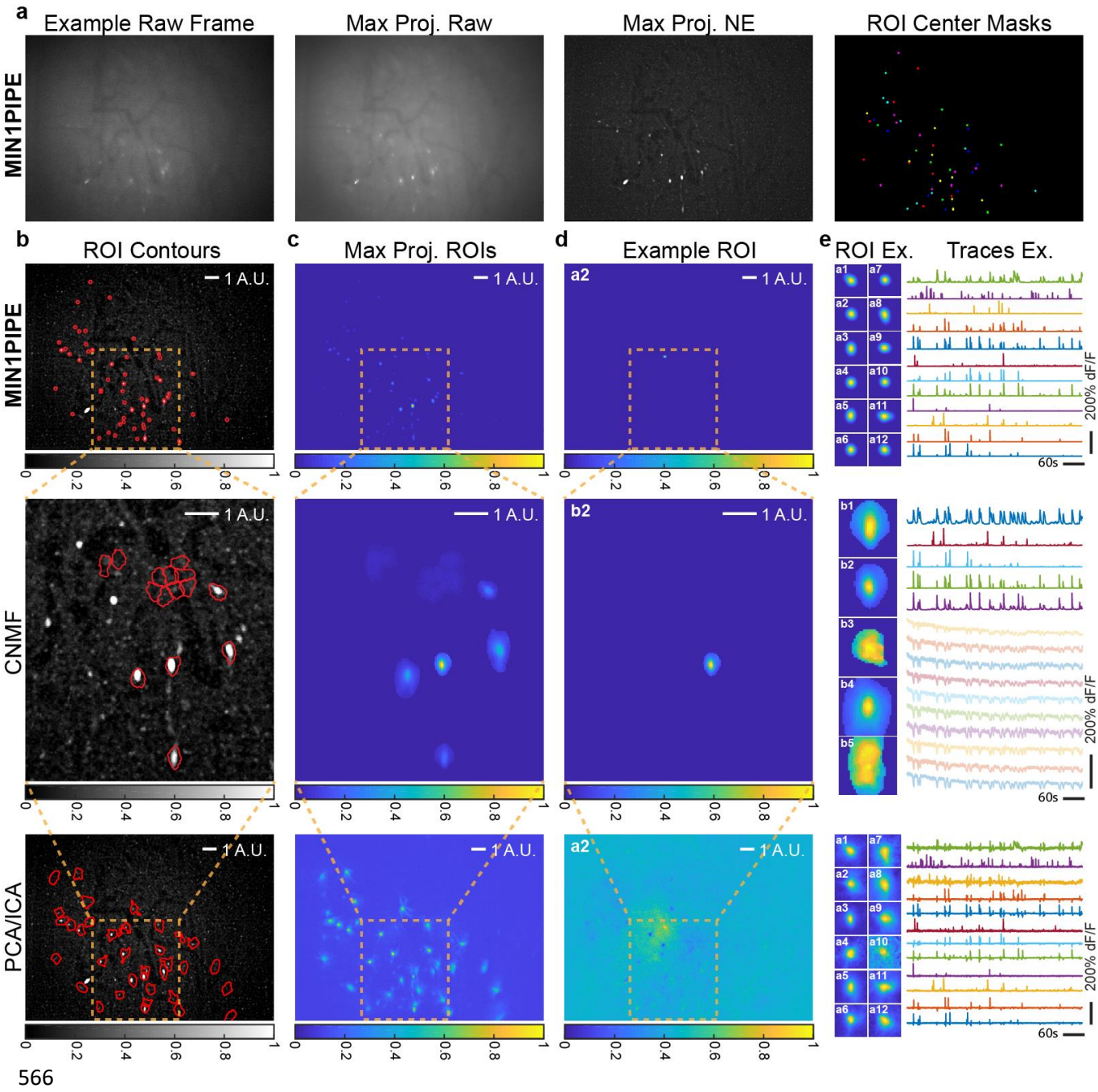


547

548 **Fig. 3.** Comparing different methods using miniscope imaging data from the mouse barrel
549 cortex. **a.** Demo of MIN1PIPE in processing imaging data from the barrel cortex. We show
550 an example raw frame, the max projection of raw video, the max projection of the neural
551 enhanced video, and the colored ROI masks. **b.** The identified ROI contours superimposed
552 on the max projection of neural enhanced video. MIN1PIPE returns a set of ROIs with

553 well-shaped, localized spatial footprints. CNMF and PCA/ICA identify only a small subset
554 of true positives (71 and 79 respectively), and either the footprints are not localized, or the
555 run into memory issues. The projection map is shown in grayscale colormap. **c.** The max
556 projections of all identified ROI footprints. This further demonstrates the general properties
557 of the extracted neural components. **d.** The spatial footprint of an example ROI. MIN1PIPE
558 extracted the most localized component that is close to the real shape shown in the neural
559 enhanced projection, whereas CNMF detected the noise-sensitive result and PCA/ICA
560 detected the unrealistic component as a mix of several other components. **e.** The temporal
561 traces of ten examples randomly selected from the ROIs that are detected by all three
562 methods. A similar conclusion can be drawn in spatial footprints, whereas MIN1PIPE
563 shows the best performance in temporal trace extraction (some potential false positive
564 events indicated by red arrows).

565 Figure 4



566

567 **Fig. 4.** Comparing different methods using miniscope imaging data from the Area X in

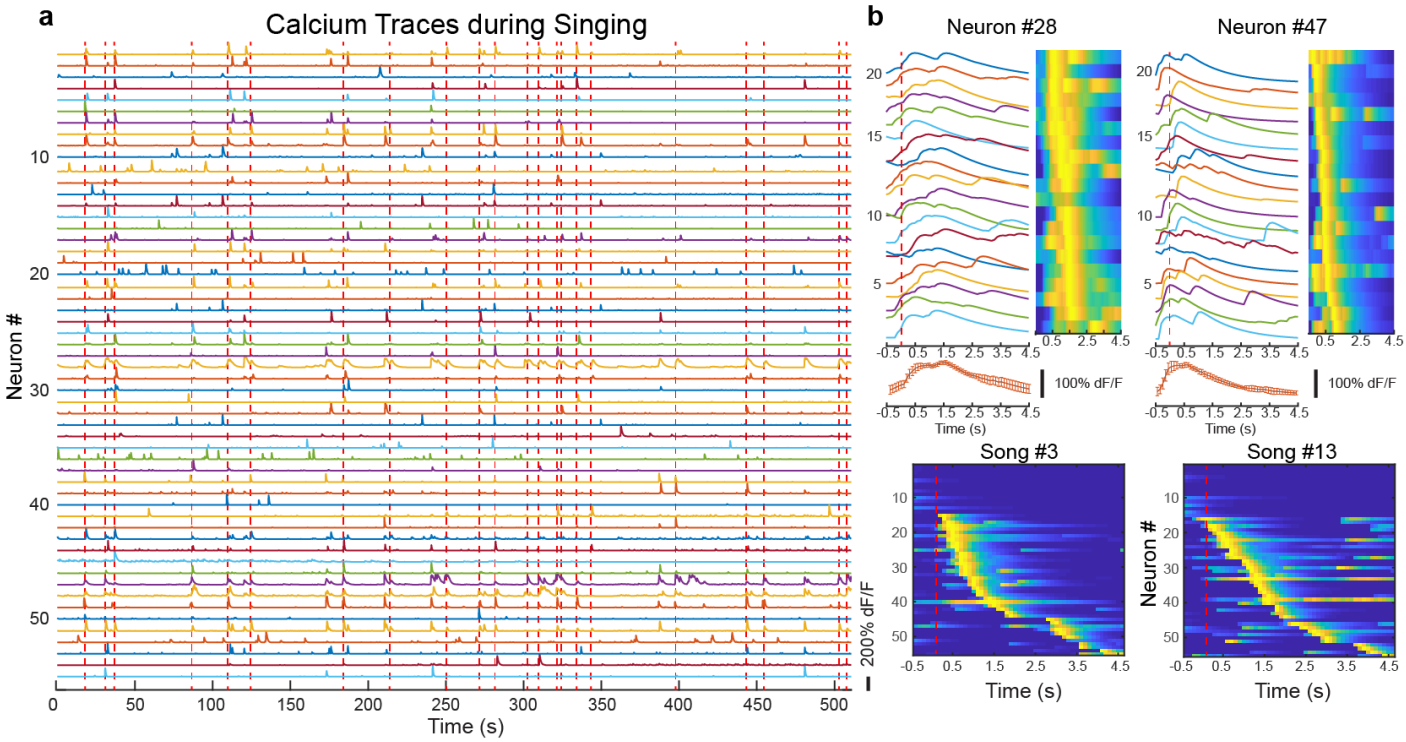
568 zebra finch. **a.** Demo of MIN1PIPE in processing data from Area X in the zebra finch brain.

569 We show an example of the raw frame, the max projection of raw video, the max projection

570 of the neural enhanced video, and the colored ROI masks. **b.** The identified ROI contours

571 superimposed on the max projection of neural enhanced video. MIN1PIPE returns a set of
572 ROIs with well-shaped, localized spatial footprints. CNMF and PCA/ICA identify only a
573 small subset of true positives (15 and 35 respectively), and either the footprints are not
574 localized, or the processes run into memory issues. The projection map is shown in
575 grayscale colormap. **c.** The max projections of all the identified ROI footprints. **d.** The
576 spatial footprint of an example ROI. MIN1PIPE extracted the most localized component
577 that is close to the real shape shown in the neural enhanced projection, whereas PCA/ICA
578 extracted the unrealistic component as a mix of several other components. CNMF on the
579 other hand, did not detect this ROI. **e.** The temporal traces of twelve examples selected
580 from the ROIs detected by both MIN1PIPE and PCA/ICA. A similar conclusion can be
581 drawn in spatial footprints, whereas MIN1PIPE shows the best performance of temporal
582 trace extraction. Faded traces in CNMF indicate the traces of the false positives.

583 Figure 5



584

585 **Fig. 5.** Analysis of neural correlation in Area X of zebra finch to singing behavior. **a.** The
586 traces of all automatically detected ROIs with MIN1PIPE superimposed with complete
587 singing onsets in red dashed lines. Note that incomplete song events are not labeled and
588 taken into analysis. **b.** Upper panel: two example neurons with precise song selectivity.
589 The traces represent the neuronal activity of the first 20 complete song singing events with
590 a window of 0.5 second before and 4.5 seconds after the song onset. The heatmap is another
591 way of visualizing the analysis. The error bar graphs at the bottom show the trial average
592 traces of the two neurons. Lower panel: population activity pattern of two example song
593 singing events. The neurons are sorted to the latency of the peak. Red dashed lines indicate
594 the onset of songs.

ONLINE METHODS

NEURAL ENHANCING The neural enhancing module contains two steps of framewise operations. We denote $\mathbf{X} \in \mathbb{R}^{P\times T}$ (P is number of pixels per frame, and T is number of frames) as the raw video after converting each two dimensional frame into a vector, and $\mathbf{I} \in \mathbb{R}^{M\times J}$ (M and J are the height and width of a frame respectively) as a frame before vectorization.

Denoising To first remove the spatial noise embedded in each frame introduced by the photoelectric process of the sensor, we apply denoising operation on each frame $\mathbf{I}$ using anisotropic diffusion [36].For a given diffusion time τ, the evolution follows the equation $\frac{\partial \mathbf{I}}{\partial \tau} = \mathrm{div}(\mathbf{C}\nabla\mathbf{I}) \triangleq \nabla\mathbf{C}\cdot\nabla\mathbf{I} + \mathbf{C}\Delta\mathbf{I}$ where $\mathbf{C}$ is the diffusive coefficient matrix depending on pixels and τ. By choosing concrete form of $\mathbf{C}$ and τ, we can control the smoothing level along or perpendicular to the boundaries between neurons and the background. Due to the simple structures in the cleaned imaging field, we choose the classical Perona-Malik filter $\mathbf{C} = \exp\left(-\frac{\|\nabla\mathbf{I}\|^2}{\kappa^2}\right)$, where κ controls the threshold of high-contrast. The diffusivity is selected to preferentially smooth high-contrast regions, and κ and τ are chosen to allow high tolerance. We use $\kappa = 0.5$ and $\tau = 0.5$ with a step size $\delta t = 0.05$ as the default parameters in anisotropic diffusion. The output image $\mathbf{I^s}$ contains reduced spatial noise while preserves the boundary information of ROIs.

Background Removal Based on the observation that neuronal ROIs are small in size compared to background structures, the gray-scale morphological opening operator with a binary structuring element matrix $\mathbf{\Phi}$ can well estimate the background $\mathbf{B}$ from the frame $\mathbf{I}$ adaptively. The opening operator is the combination of erosion $\ominus$ and dilation $\oplus$: $\mathbf{B} = (\mathbf{I}\ominus\mathbf{\Phi})\oplus\mathbf{\Phi}$ (Van Den Boomgaard and Van Balen, 1992), where the morphological erosion returns minimum value $(\mathbf{I}\ominus\mathbf{\Phi})(x,y) = \min_{(s,t)\in\mathbf{\Phi}}\{\mathbf{I}(x+s,y+t)\}$ and dilation returns

maximum value $(\mathbf{I} \oplus \mathbf{\Phi})(x, y) = \max_{(s,t)\in\mathbf{\Phi}}\{\mathbf{I}(x+s, y+t)\}$ within the same structuring window at each point. In practice, the choice of the structuring element should be comparable to the overall size of the neurons in the imaging field. We choose the structuring element to be of disk shape with a default size of 9 (pixels). The foreground is then computed as: $\mathbf{I^f} = \mathbf{I} - \mathbf{B}$. After the dynamic background removal, $\mathbf{I^f}$ contains only neuronal information with minimal background noise corruption.

MOVEMENT CORRECTION The movement correction module performs a hierarchical registration framework over the frames of the imaging video. The neural enhanced video is first scored with a measuring metric on the relative displacement between two neighboring frames. The video is then segmented into stable or nonstable sections based on the score, followed by three levels of registration: the intra-stable-section, inter-stable-section and nonstable-section registration.

Scoring Metric We track features containing corner information of every two neighboring frames using the KLT tracker, and then calculate the average displacement of these features. The KLT tracker first selects good features to track, where the feature points contains enough information on both directions within the neighborhood p. The problem can be formulated as $\nabla p = 0$ over the neighborhood. A good feature point is selected if the smaller eigenvalue $\sigma_{min} \geq \sigma_0$ and the *condition number* $\kappa = \frac{\sigma_{max}}{\sigma_{min}} \leq \kappa_{max}$. Then the algorithm tracks the feature points based on Newton-Raphson minimization on the normal equation:

$$\mathbf{A}\boldsymbol{s} = \boldsymbol{b}$$

where

$$\mathbf{A} = \Sigma_{\boldsymbol{x}} \nabla \mathbf{J}_t^{\mathbf{f}}(\boldsymbol{x}) \nabla \mathbf{J}_t^{\mathbf{f}}(\boldsymbol{x})^\top \omega(\boldsymbol{x} - \boldsymbol{x}_{\mathbf{I}_t^{\mathbf{f}}}) \quad \textit{and} \quad \boldsymbol{b} = \Sigma_{\boldsymbol{x}} \nabla \mathbf{J}_t^{\mathbf{f}}(\boldsymbol{x}) [\mathbf{I}_t^{\mathbf{f}}(\boldsymbol{x}) - \mathbf{J}_t^{\mathbf{f}}(\boldsymbol{x})]^\top \omega(\boldsymbol{x} - \boldsymbol{x}_{\mathbf{I}_t^{\mathbf{f}}})$$

$J_t^{\mathtt{f}}$ and $I_t^{\mathtt{f}}$ are the moving and fixed frame of the tth registration pair after neural enhancing respectively (in this case, $J_t^{\mathtt{f}} = I_{t+1}^{\mathtt{f}}$), and $\omega(\boldsymbol{x})$ is the window of the neighborhood. For the scoring, we choose 31 pixels as the default window size of the neighborhood. To segment the frames into stable/nonstable sections, we compute the minimum value between the 3 times of median absolute deviation (MAD) above median and upper limit threshold (0.5 pixel as default) to be the threshold of segmentation. Any frame whose score is above the threshold is considered as one belonging to some stable section. The frames that are connected together are labeled as stable sections while the remaining connected components are nonstable ones.

Intra-stable-section Registration Displacements with small magnitude or consistent directions can be approximated as translational displacement. Therefore, after the segmentation, movements within the stable sections can be considered translational. Within each stable section, an efficient displacement matching method with subpixel resolution is used to register the neighboring frames. Here we use a similar Lucas-Kanade tracker to track the center of the moving frames with a sufficiently large window (80% of the frame size) between two neighboring frames. However, when frame number within a section grows, the registration error may cumulate for later frames. Therefore, in practice we update the fixed frame as reference every second and all moving frames within that period are aligned to the fixed frame. This is acceptable due to the slow calcium dynamics so that the imaging field remains essentially similar within a second.

Inter-stable-section Registration The imaging field is stable within each stable section after the intra-section registration. However, the imaging fields of different stable sections are not necessarily similar. Therefore, we need to align different stable sections without specific assumptions of the movement type. We first extract the overall information of each stable section by averaging the frames with significant neuronal activation, we then sort

these sectional images $\mathbf{I}_i^{\mathtt{sec}}$ according to their similarity. To register the ith sectional image based on the sorted similarity, a reference image is generated by calculating the weights $\hat{x}$ using the least square regression between the registered frames and the current sectional image:

$$\mathbf{A}^{\mathtt{ref}} x = \boldsymbol{b}^{\mathtt{cur}}$$

where

$$\mathbf{A}^{\mathtt{ref}} = (\mathsf{vec}(\mathbf{I}_1^{\mathtt{sec}}), \cdots, \mathsf{vec}(\mathbf{I}_{i-1}^{\mathtt{sec}})) \quad and \quad \boldsymbol{b}^{\mathtt{cur}} = \mathsf{vec}(\mathbf{I}_i^{\mathtt{sec}})$$

The reference image is then reshaped into two dimensional matrix $\mathbf{R}_i^{\mathtt{sec}}$ from $\boldsymbol{b}^{\mathtt{ref}} = \mathbf{A}^{\mathtt{ref}}\hat{x}$. Once we have the moving and fixed sectional images, we apply rigid preregistration to the moving image using similar KLT tracker. To finally correct for the nonrigid deformation, we apply the diffeomorphic Log-Demons to the preregistered image $\hat{\mathbf{I}}_i^{\mathtt{sec}}$ based on the reference $\mathbf{R}_i^{\mathtt{sec}}$. In general, the diffeomorphic Log-Demons is a non-parametric registration method, which aims at finding the displacement of all pixels by minimizing the global energy:

$$E(c, s) = \frac{1}{\sigma_i^2} Sim(\mathbf{F}, \mathbf{M} \circ c) + \frac{1}{\sigma_x^2} dist(s, c)^2 + \frac{1}{\sigma_T^2} Reg(s)$$

It consists of the similarity, correspondence and regularization items, where $\mathbf{F}$ and $\mathbf{M}$ are fixed and moving image respectively, and σ_i, σ_x and σ_T account for similarity, spatial uncertainty on the correspondence and regularization level. In specific, Log-Demons represents everything in log domain. It spells out two regularizing terms, fluid and diffusion, and uses a Lie group structure to impose diffeomorphism. In our implementation, we choose $\sigma_i = \sigma_x = 1$, and $\sigma_{fluid} = \sigma_{diffusion} = 3$. Because all the frames within the same stable section share the same imaging field, the deformation field of the current stable section is then applied to warp all the frames. After registering all the stable sections iteratively, all the stable frames are efficiently aligned.

Nonstable-section Registration At the third level, only nonstable sections remain to be aligned. We use a similar approach as in the inter-section registration, where we first preregister the current frame to the reference image, and then apply Log-Demons to finely align the frame. Similarly, the reference image is computed based on some of the registered frames. For the ith nonstable section, this set of frames includes the last frame of the ith stable section, the first frame of the $i+1$th stable section, and the frames that have already been registered in the ith nonstable section.

SEEDS-CLEANSED SIGNAL EXTRACTION We generate and cleanse the potential seeds of ROIs in this section, and create the initial spatial regions and the time series of ROIs for later refinement of the spatiotemporal signals.

Over-complete Seeds Initialization To generate the initial set containing centers of all potential ROIs, we construct a *randomized max pooling* process to create an over-complete set of seeds. This process first randomly selects a portion of frames $\mathbf{I}^{\mathbf{s}}_{\boldsymbol{\alpha}} = \{\mathbf{I}^{\mathbf{s}}_t\}_{t\in\boldsymbol{\alpha}}, \boldsymbol{\alpha} \subseteq \{1,\ldots,T\}$, where $\boldsymbol{\alpha}$ is a randomized subset of frame number. We then compute the max-projection map across selected frames $\mathbf{I}^{\max} = \max(\mathbf{I}^{\mathbf{s}}_{\boldsymbol{\alpha}})$, and further detect all the local max points on this map as $\mathcal{S}$. We repeat the above procedures multiple times, and collect the union of each map $\mathcal{S}$ as the final over-complete set of neural seeds $\mathcal{S}_{\text{seeds}}$. Our randomized max pooling can improve the true positives compared with max-pooling over the entire video, because a real seed can be buried in the uneven florescence of an ROI over a long period but can most likely be discovered in a small temporal vicinity. To exclude false positive seeds from the set, we propose the following two-stage algorithm for seeds refinement.

Seeds Refinement with GMM The temporal properties of ROIs and non-ROIs can be very different. In one aspect, the ROIs often have prominent peak-valley difference d, while non-ROIs tend to be less spiky. We assume $\{d_s | s \in \mathcal{S}_{\text{seeds}}\}$ are generated from a mixture

of two Gaussians $d_s \sim \sum_{i=1}^{2} \omega_i \mathcal{N}(d|\mu_i, \sigma_i^2)$, where $\mu_i, \sigma_i^2, \omega_i$ indicate the mean, variance, mixture proportion of the ith Gaussian component. Therefore, we can cluster the seeds based on their probabilities of belonging to each component, and only consider those having higher probabilities to the Gaussian with a larger mean as positive seeds.

Seeds Refinement with LSTM To select true ROI seeds from the over-complete set, we need to classify the seeds based on their patterns of neuronal calcium dynamics. We propose to employ Recurrent Neural Networks (RNNs) (LeCun et al., 2015) for calcium signal sequence classification. We offline trained the RNNs with a separate training dataset, composed of both positive and negative sequence chunks of length $T_0 = 100$, obtained with 10Hz frame rate. The training data for the RNN module were selected from a separate real dataset. The labels were manually selected by experienced neuroscientists with a conservative standard: the positive trials were the ones with most obvious calcium dynamics that were aligned to the peak of the spike, whereas the negative trials were randomly selected from the rest of the data. The training dataset contained 1000 positive and 1000 negative labels, and both the validation and the test set contained 600 balanced trials. Specifically, consider training data $\mathcal{D} = \{\mathbf{D}_1, \cdots, \mathbf{D}_N\}$, where $\mathbf{D}_n \triangleq (\boldsymbol{y}_n, l_n)$, with input sequence $\boldsymbol{y}_n$ and output label $l_n \in \{0, 1\}$. Our goal is to learn model parameters $\boldsymbol{\theta}$ to best characterize the mapping from $\boldsymbol{y}_n$ to l_n with likelihood $p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{D}_n|\boldsymbol{\theta})$. In our setting for sequence classification, the input is a sequence, $\boldsymbol{y} = \{y_1, \ldots, y_{T_0}\}$, where y_t is the input data at time t. There is a corresponding hidden state vector $\boldsymbol{h}_t \in \mathbb{R}^K$ at each time t, obtained by recursively applying the *transition function* $\boldsymbol{h}_t = g(\boldsymbol{h}_{t-1}, \boldsymbol{y}_t; \mathbf{W}, \mathbf{U})$. $\mathbf{W}$ is *encoding weights*, and $\mathbf{U}$ is *recurrent weights*. The ouput c for our classification is defined as the corresponding *decoding function* $p(c|\boldsymbol{h}_{T_0}; \mathbf{V}) = \sigma(\mathbf{V}\boldsymbol{h}_{T_0})$, where $\sigma(\cdot)$ denotes the *logistic sigmoid function*, and $\mathbf{V}$ is *decoding weights*.

The transition function $g(\cdot)$ can be implemented with a *gated* activation function, such as LSTM (Hochreiter and Schmidhuber, 1997) or a Gated Recurrent Unit (GRU) (Cho et al.,

2014). Both LSTM and GRU have been proposed to address the issue of learning long-term sequential dependencies. Each LSTM unit has a cell containing a state $\mathtt{c}_t$ at time t. This cell can be viewed as a memory unit. Reading or writing the memory unit is controlled through sigmoid gates: input gate $\mathtt{i}_t$, forget gate $\mathtt{f}_t$, and output gate $\mathtt{o}_t$. The hidden units $\boldsymbol{h}_t$ are updated as follows:

$$
\begin{aligned}
\mathtt{i}_t &= \sigma(\mathbf{W}_i \boldsymbol{y}_t + \mathbf{U}_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i)\,, & \tilde{\mathtt{c}}_t &= \tanh(\mathbf{W}_c \boldsymbol{y}_t + \mathbf{U}_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c)\,, \\
\mathtt{f}_t &= \sigma(\mathbf{W}_f \boldsymbol{y}_t + \mathbf{U}_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f)\,, & \mathtt{c}_t &= \mathtt{f}_t \odot \mathtt{c}_{t-1} + \mathtt{i}_t \odot \tilde{\mathtt{c}}_t\,, \\
\mathtt{o}_t &= \sigma(\mathbf{W}_o \boldsymbol{y}_t + \mathbf{U}_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o)\,, & \boldsymbol{h}_t &= \mathtt{o}_t \odot \tanh(\mathtt{c}_t)\,,
\end{aligned}
\tag{1}
$$

where $\odot$ represents the element-wise matrix multiplication operator. Note that the training of RNNs is completed off-line, only the efficient testing stage is performed for seeds refinement.

In the testing stage, given an input $\tilde{\boldsymbol{y}}$ (with missing label $\tilde{l}$), the estimate for the output is $p(\tilde{l}|\tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = \arg\max \log p(\mathcal{D}|\boldsymbol{\theta})$. In our practical application, the testing sequence $\bar{\boldsymbol{y}} \in \mathbb{R}^T$ is often of length $T > T_0$. We first convert it into a bag of subsequences $\{\tilde{\boldsymbol{y}}_i\}_{i=1}^{T-T_0+1}$, with a sliding window of width T_0 and moving step size 1. The well trained LSTMs are then used to label the subsequences. We consider $\bar{\boldsymbol{y}}$ as positive if at least one subsequence in its bag is classified as "calcium spike" $(l = 1)$, otherwise negative. Intuitively, this means that we only care about the patterns of neural spikes, regardless of their temporal positions.

Seeds Merging Once the set of seeds is cleansed, there is still a low possibility of identifying multiple seeds within a single ROI. Therefore, we merge all these redundant seeds by computing the temporal similarity of seeds within their neighborhood, and preserving the one with maximum intensity. Specifically, we compute the similarity based on phase-locking information (Hahn et al., 2006). For a sequence, we extract the instantaneous phase dynamics using Hilbert transform, and only consider the subsequences containing prominent peaks, because all the pixels within the same ROI are highly correlated only during the calcium spiking periods, but not necessarily during baseline period. After seeds merging,

we obtain K seeds, as the number of ROIs in our MIN1PIPE.

Spatial and Temporal Initialization The time series of the kth seed is used as initial guess of temporal signal $\hat{\boldsymbol{s}}_k$. The spatial map of the corresponding ROI, $\hat{\boldsymbol{r}}_k$, is estimated by pooling neighbor pixels with temporal similarity above a threshold. Because of our seeds generating and cleansing approach, the spatiotemporal initialization in our method can be readily parallelized. Then we fine-tune the corresponding spatial and temporal guess, by applying semi-NMF:

$$\min \|\boldsymbol{z}_k - \hat{\boldsymbol{r}}_k \hat{\boldsymbol{s}}_k^\top\|$$

where $\boldsymbol{z}_k \subset \mathbf{I}^{\mathtt{s}}$ is a region containing the kth ROI. The spatial map and the temporal signal are updated respectively:

$$\boldsymbol{r}_k \leftarrow \boldsymbol{z}_k \boldsymbol{s}_k (\boldsymbol{s}_k^\top \boldsymbol{s}_k)^{-1}, \qquad \boldsymbol{s}_k \leftarrow \boldsymbol{s}_k \sqrt{\frac{(\boldsymbol{r}_k \boldsymbol{z}_k)^+ + \boldsymbol{s}_k (\boldsymbol{r}_k^\top \boldsymbol{r}_k)^-}{(\boldsymbol{r}_k \boldsymbol{z}_k)^- + \boldsymbol{s}_k (\boldsymbol{r}_k^\top \boldsymbol{r}_k)^+}}$$

where $\mathbf{M}^+ \triangleq \frac{|\mathbf{M}|+\mathbf{M}}{2}$ and $\mathbf{M}^- \triangleq \frac{|\mathbf{M}|-\mathbf{M}}{2}$ for any matrix $\mathbf{M}$ (Ding et al., 2010). In total, we now have K ROIs $\mathbf{S}^0 = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_K]$, and temporal signals $\mathbf{R}^0 = [\boldsymbol{r}_1, \cdots, \boldsymbol{r}_K]$. Similarly, the background parameters $\boldsymbol{b}^0$, $\boldsymbol{f}^0$ are also estimated by the same semi-NMF procedure, using $(\mathbf{I}^{\mathtt{s}} - \sum_{k=1}^K \boldsymbol{r}_k \boldsymbol{s}_k^\top)$. $\mathbf{S}^0, \mathbf{R}^0, \boldsymbol{b}^0, \boldsymbol{f}^0$ are then used as initializations of the spatiotemporal signal refinement.

Spatiotemporal Signal Refinement We perform the iterative spatial and temporal optimizations to update the spatial footprints of individual ROIs, and the temporal traces with deconvolved spike trains, as proposed in CNMF. Specifically, for the neural enhanced video $\mathbf{X}^{\mathtt{f}}$, the method decomposes $\mathbf{X}^{\mathtt{f}}$ into a spatial dictionary $\mathbf{R}^{\mathtt{f}} \in \mathbb{R}^{P \times K}$ representing individual ROIs, and corresponding temporal dynamics matrix $\mathbf{S}^{\mathtt{f}} \in \mathbb{R}^{T \times K}$, in addition to

the background $\mathbf{B}^{\mathtt{f}}$ and the background dynamics $\mathbf{E}$:

$$\mathbf{X}^{\mathtt{f}} = \mathbf{R}^{\mathtt{f}}{\mathbf{S}^{\mathtt{f}}}^{\top} + \mathbf{B}^{\mathtt{f}} + \mathbf{E}$$

Similarly, with the rank-1 assumption on $\mathbf{B}^{\mathtt{f}}$, CNMF decomposes the background $\mathbf{B}^{\mathtt{f}} = \boldsymbol{b}^{\mathtt{f}}{\boldsymbol{f}^{\mathtt{f}}}^{\top}$, where $\boldsymbol{b}^{\mathtt{f}} \in \mathbb{R}^{P}$ and $\boldsymbol{f}^{\mathtt{f}} \in \mathbb{R}^{T}$. Meanwhile, $\mathbf{S}^{\mathtt{f}}$ is also correlated with underlying action potential events: $\mathbf{A}^{\mathtt{f}} = {\mathbf{S}^{\mathtt{f}}}^{\top}\mathbf{G}^{\mathtt{f}}$, where $\mathbf{A}^{\mathtt{f}} \in \mathbb{R}^{K\times T}$, and $\mathbf{G}^{\mathtt{f}} \in \mathbb{R}^{T\times T}$ is the coefficient matrix of the low-order autoregression. The variables $\mathbf{R}^{\mathtt{f}}, \mathbf{S}^{\mathtt{f}}, \boldsymbol{b}^{\mathtt{f}}, \boldsymbol{f}^{\mathtt{f}}$ are therefore estimated via iteratively alternating between the following two steps [26].

Estimating Spatial Variables Given the estimates of temporal variables ${\mathbf{S}^{\mathtt{f}}}^{(\ell-1)}$ and ${\boldsymbol{f}^{\mathtt{f}}}^{(\ell-1)}$ from the last iteration, the spatial parameters can be updated by solving the problem:

$$\min_{\mathbf{R}^{\mathtt{f}},\boldsymbol{b}^{\mathtt{f}}} \|\mathbf{R}^{\mathtt{f}}\|_1,\ s.t.\ \mathbf{R}^{\mathtt{f}}, \boldsymbol{b}^{\mathtt{f}} \geq 0$$
$$\|\mathbf{X}^{\mathtt{f}}(i,:) - \mathbf{R}^{\mathtt{f}}(i,:){\mathbf{S}^{\mathtt{f}}}^{(\ell-1)\top} - \boldsymbol{b}^{\mathtt{f}}(i){\boldsymbol{f}^{\mathtt{f}}}^{(\ell-1)\top}\| \leq \sigma_i\sqrt{T}$$

where $\mathbf{X}^{\mathtt{f}}(i,:)$ is the ith row of the matrix $\mathbf{X}^{\mathtt{f}}$, $\boldsymbol{b}^{\mathtt{f}}(i,:)$ is the ith element of the vector. $\epsilon_i\sqrt{T}$ is the empirically selected noise residual constraint of the corresponding pixel. This is essentially a basis pursuit denoising problem, and it is solved by the least angular regression (Efron et al., 2004) in implementation.

Estimating Temporal Variables Given the estimates of spatial variables $\mathbf{R}^{\mathtt{f}}$, $\boldsymbol{b}^{\mathtt{f}}$ and temporal parameters ${\mathbf{S}^{\mathtt{f}}}^{(\ell-1)}$, ${\boldsymbol{f}^{\mathtt{f}}}^{(\ell-1)}$ from the last iteration, the temporal parameters can be updated by solving the problem:

$$\min_{\mathbf{S}^{\mathtt{f}},\boldsymbol{f}^{\mathtt{f}}} \sum_{k=1}^{K} \mathbf{1}^{\top}\mathbf{G}^{\mathtt{f}}\mathbf{S}^{\mathtt{f}},\ s.t.\ \mathbf{G}^{\mathtt{f}}\boldsymbol{s}_k^{\mathtt{f}} \geq 0, k = 1,\cdots,K$$
$$\|\mathbf{X}^{\mathtt{f}}(i,:) - \mathbf{R}^{\mathtt{f}}(i,:){\mathbf{S}^{\mathtt{f}}}^{(\ell-1)} - \boldsymbol{b}^{\mathtt{f}}(i){\boldsymbol{f}^{\mathtt{f}}}^{(\ell-1)\top}\| \leq \sigma_i\sqrt{T}$$

Unlike in CNMF, where the spatial footprints are serially updated and subtracted from the preceding residuals, we extract spatial footprints from the original data that does not depend on preceding iterations. Therefore, the information loss/duplication is reduced and the optimization procedures can be parallelized in our method.

Data availability The data that support the findings of this study are available from the corresponding authors upon request.

Code availability The codes will be freely available upon publication, and the beta version is available for reviewers upon request.