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.

56   For extraction of neuronal signals, many previous methods work well in situations with

57   high SNR and stable field of views, therefore they are best suited for processing two-photon

58   imaging data. These algorithms include linear unsupervised basis learning methods

59   (Mukamel et al., 2009; Reidl et al., 2007; Pachitariu et al., 2013), nonlinear unsupervised

60   methods (Maruyama et al., 2014; Pnevmatikakis et al., 2016), and the supervised learning

61   method (Apthorpe et al., 2016). The PCA/ICA (principal component analysis followed by

62   independent component analysis) method (Mukamel et al., 2009) was the first attempt to

63   automatize the signal extraction process from miniscope imaging data through manual

64   annotations of ROIs, but this method has difficulties in delineating localized ROIs and in

65   separating overlapping ROIs. Since single-photon based imaging collect lights from a large

66   depth of field, overlapping neurons in different depth is not uncommon. Another method

67   called CNMF (constraint matrix factorization framework) (Pnevmatikakis et al., 2016),

68   which combines nonlinearity in matrix factorization with simultaneous deconvolving spike

69   trains from calcium dynamics, returns more spatially localized maps of ROIs compared to

70   other methods and has better performance in identifying overlapping neurons. While

71   methods like CNMF achieve plausible results in processing two-photon imaging data, they

72   are ill-suited for processing the single-photon based miniscope imaging because: 1) the

73   data from miniscope imaging are dominated by noisy, uneven and fluctuating background,

74   2) such methods depend on sophisticated parameter tuning, especially requiring setting

75   parameters that are unknown *a priori* in practice such as "number of neurons". Thus, a new

76   method that can effectively remove background yet preserve real neural signals is highly

77   desired.

4

78   Moreover, both PCA/ICA and CNMF rely on the stable imaging field and will fail if the

79   imaging field contains movement, but movements (including distortions) during imaging,

80   are often inevitable. Therefore, correcting movements and distortions is another major

81   problem needs solving before neural activity signals can be reliably extracted. Many

82   methods were developed independently in attempt to solve this problem. For example,

83   several approaches register frames through template matching based on the assumption

84   that the major form of movements is translational displacement (Dubbs et al., 2016;

85   Thévenaz et al., 1998). To eliminate such assumptions, methods with block-based

86   displacement field estimation were developed, with image feature matching algorithms

87   extended from Lucas-Kanade tracker (Greenberg and Kerr, 2009; Lucas and Kanade, 1981)

88   or Hidden Markov Models (Dombeck et al., 2007; Kaifosh et al., 2013). Some of these

89   movement correction methods have been included as a module in a larger toolbox, using

90   such frame-wise rigid registration approaches (Kaifosh et al., 2014; Pachitariu et al., 2017).

91   When applied to handle nonrigid movement registrations, existing methods make specific

92   assumptions about both the form and magnitude of the potential movements that result in

93   suboptimal performance when large deformation occurs during imaging. Furthermore,

94   errors in the movement correction can easily propagate since these methods all register

95   frames based on a single reference frame. Considering that extracting neural activity

96   signals is highly dependent on removing movement artifacts, an accurate and robust

97   movement correction module is imperative. A simple unimodal algorithm for either

98   translation/rigid or nonrigid registration is insufficient for this purpose.

99   Here we develop the MINiscope 1-photon imaging signal extraction PIPEline (MIN1PIPE)

100 that takes the very raw calcium videos as inputs, and automatically removes background

101   while preserving signals, corrects movements with no assumptions of the types of

102   movements, and delivers separated neuronal ROIs as well as deconvolved calcium traces

103   as outputs. The MIN1PIPE contains a neural enhancing module that minimizes the

104   influence of background unevenness and fluctuations, a hierarchical movement correction

105   module that can handle all kinds of deformation with minimal error propagation, and a

106   seeds-cleansed neural signal extraction module that identifies the set of real ROIs and their

107   corresponding calcium traces without setting unknown parameters *a priori* (Fig. 1).

108   Though our MIN1PIPE is primarily developed for single-photon based miniscope imaging,

109   individual modules can also be independently combined with other processing algorithms

110   to improve performance in analyzing two-photon imaging.

111   **RESULTS**

112   **Core modules in MIN1PIPE**

113   *Neural Enhancing Module*: The core of MIN1PIPE relies on turning the imaging data into

114   a stack of background free, baseline corrected frames as the first step. Due to the complex

115   spatiotemporal properties of background dynamics, our method applies a frame-wise

116   background estimator that is adaptive to the local properties. A natural idea can be

117   translated from mathematical morphology and computer vision, that the foreground neural

118   signals can be approximated by subtracting the estimated background in a denoised image.

119   Therefore, we first remove the grainy noise inherent to the single-photon system (see

120   example of such grainy noise in Supplementary Fig. S2) while preserving the boundary

121   between foreground and background, and this is achieved by applying an *anisotropic*

122   *diffusion* denoising operation (Perona and Malik, 1990) on the raw imaging frames. Next,

123   we use a simple straightforward *morphological opening* operation (Serra and Vincent,

6

124   1992) as the background estimator, with the size of the structure element similar to that of

125   the neurons in the imaging field. The opening operation removes structures smaller than

126   the desired structure element. Subsequently, the foreground that contains all the neuronal

127   signals with the minimal noise is computed as the difference between the denoised raw and

128   the morphological opened frames (Fig. 1a and Supplementary Fig. S1-S2).

129   *Movement Correction Module*: After the neural signals are enhanced, we next correct for

130   movements in the imaging videos. The problem of movement correction can essentially be

131   broken down to image stack registration. However, without setting specific constraints on

132   the form or magnitude of the movements, even the most efficient registration algorithms

133   require a running time on the order of seconds to minutes per frame (Vercauteren et al.,

134   2009). Considering that the general imaging datasets contain tens of thousands of frames,

135   the time required for applying these sophisticated image registration methods to every

136   frame is inconceivable. Here we develop a hierarchical video registration framework for

137   the correction of all types of movement, without sacrificing the precision or the speed of

138   corrections. Our framework first decomposes the imaging video into two sections: the

139   *stable sections* whose movements can be approximated by small translational displacement,

140   and the *non-stable sections* that contain large general deformation. This step uses the KLT

141   tracker that estimates the displacement of potential corner-like features between two

142   neighboring frames (Shi and Tomasi, 1994). Next, we employ three levels of different

143   strategies to align images. At the first level, we correct the small translational displacement

144   within each stable section using the fast Lucas-Kanade tracker, which can be performed

145   efficiently in parallel on multiple sections (Lucas et al., 1981). We then incorporate a

146   diffeomorphic Log-Demons image registration method which can handle large

7

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,

8

170   where the traces of real neurons usually have larger fluctuations compared to the non-

171   neuron false positive seeds. The GMM removes most background false positives without

172   losing real neurons (Supplementary Fig. S3). To further remove the remaining false

173   positives, such as the ones with abnormal background fluctuations or hemodynamics, we

174   have trained an Recurrent Neural Networks (RNNs) with LSTM module offline as the

175   classifier for calcium spikes (Supplementary Fig. S4) (LeCun et al., 2015; Hochreiter and

176   Schmidhuber, 1997). Those seeds whose traces contain RNN-identified calcium spikes,

177   regardless of their temporal locations, are classified as true positives whereas the rest are

178   deemed as false positives. After such cleansing processes, there is still a low possibility of

179   identifying multiple seeds within a single ROI. Therefore, we merge potentially redundant

180   seeds by computing the temporal similarity of seeds within their neighborhoods, and

181   preserving the ones with maximum intensity. With the cleansed set of seeds as the initial

182   position of ROIs, we next perform the iterative spatial and temporal optimizations, as

183   proposed in CNMF, to update the spatial footprints of individual ROIs, and the temporal

184   traces with deconvolved spike trains. Notably, unlike previous CNMF, where the spatial

185   footprints are sequentially updated and subtracted from the preceding residuals, we extract

186   spatial footprints from the original data that does not depend on preceding iterations.

187   Therefore, the information loss/duplication is reduced and the optimization procedures can

188   be parallelized in our method.

189   We show example results obtained using the MIN1PIPE methodology including a raw

190   frame, the fully processed ROIs and the example traces from ROIs (Fig. 1c).

191   **Quantitative validation of the MIN1PIPE performance**

9

192   The neural signal extraction and movement correction are two independent problems that

193   can be tested separately. For the signal extraction, we test the performance on synthetic

194   datasets with various signal levels, while for the movement correction, we can directly test

195   it on real data. To measure the performance of the signal extraction, we use a scoring metric

196   that evaluates the spatial and temporal similarity between the ground truth and the

197   identified ROIs (see Online Methods), and calculate true positive, false positive and false

198   negative. To measure the performance of the movement correction, we use a metric based

199   on the average displacement of feature points between neighboring frames.

200   We synthesized 16 imaging videos with signal levels (SL, defined as the ratio of the

201   amplitude between the signals and the background) ranging from 0.05 to 0.8. Each video

202   contains 3000 frames with 100 neurons of various shapes and calcium dynamics, and

203   background fluctuations extracted from real datasets (details in Supplementary Notes S1-

204   S2). The properties of synthetic videos resemble those of real data, whose SL falls between

205   the range of 0.2 and 0.8. The condition at 0.05 SL is an extreme (Supplementary Fig. S5

206   and Video S1-S3). We compare MIN1PIPE with PCA/ICA and CNMF. We used

207   commercially available *Mosaic* software (Inscopix Inc.) which implements PCA/ICA

208   method (Mukamel et al., 2009), and processed the data following the standard workflow

209   in the software manual. In particular, we chose the number of principal components (PC)

210   and independent components (IC) based on the suggested rate (*e.g.* 20% more ICs and 50%

211   more PCs than the estimated number of ROIs). For CNMF, we used the default

212   initialization strategy in (Pnevmatikakis et al., 2016).

213   The results of ROI detection (Fig. 2a) and the calcium traces from one example ROI

214   obtained by different methods (Fig. 2b) at 0.2 and 0.8 SL are compared. The contours of

10

215   the identified ROIs using different methods are drawn and superimposed onto the max

216   projection of the ground truth. For both SLs, MIN1PIPE can identify the nearly complete

217   set of ROIs (93% and 100% for 0.2 and 0.8 SL respectively) with minimal false positives

218   (1% and 0% respectively), whereas the other two methods detect partial subsets of ROIs

219   (PCA/ICA: 0% at 0.2 SL and ~65% true positives at 0.8 SL; CNMF: ~32% at 0.2 SL and

220   ~95% true positives at 0.8 SL). In addition, the extracted spatial footprints are less realistic

221   with the PCA/ICA or previous CNMF methods. The example calcium traces indicate that

222   MIN1PIPE has a near-optimal performance in extracting individual ROIs even at low SLs

223   when compared to the ground truth traces (Fig. 2b). In contrast, PCA/ICA completely fails

224   to identify ROIs when SL is low, while CNMF fails to separate neurons from overlapping

225   ROIs. Fig. 2c summarizes the performance accuracy of these three methods at all SLs. The

226   plots of the true positive, false positive and false negative indicate that MIN1PIPE has a

227   significantly superior performance in all conditions, and outperforms the other two

228   methods in the extreme conditions. Fig. 2d summarizes the spatiotemporal similarities

229   between the extracted results and the ground truth. The clusters of point clouds confirm

230   again that MIN1PIPE outperforms the other two methods in best resembling both the

231   spatial and temporal properties of the ground truth signals. The results for all signal levels

232   are shown in Supplementary Fig. S6.

233   To validate the performance of the movement correction in MIN1PIPE, we applied the

234   module on video sections with large deformation movements of the imaging field.

235   Specifically, we chose a video obtained through two-photon imaging of the ferret's

236   posterior parietal cortex as an example due to its particularly frequent and large

237   deformations (Supplementary Video S4 as a demo section of the full video). Fig. 2e shows

238   an example of 3 consecutive frames (with each frame pseudo-colored as pink, green or

239   blue) with large nonlinear deformations between frames superimposed together. The two

240   rigid-transform-based methods (LK and KLT Affine) by themselves fail to fully remove

241   the large deformations, whereas MIN1PIPE (combining the Log-Demons transformation)

242   succeeds in correcting these movements. We further quantified the extent of correction by

243   calculating the average displacement of feature points between two neighboring frames

244   before and after the movement correction (Fig. 2f). Before correction, the score of average

245   displacement shows frequent burst of large deformation periods, whereas after correction,

246   the score of displacement shows a nearly flat line. Notably, the frames with large

247   deformations are all well aligned (Fig. 2f lower panels; Supplementary Video S4).

248   **Application of MIN1PIPE on real miniscope imaging datasets**

249   We next compare the performance of MIN1PIPE with PCA/ICA and CNMF methods on

250   real datasets. We first applied the three methods to the miniscope calcium imaging data

251   obtained using prism probe from layer 2 and 3 of the barrel cortex in freely moving mice.

252   GCaMP6f was expressed in layer 2/3 neurons using AAV, and signals were imaged

253   continuously over 5min when the mouse freely explored its environment. Following the

254   general pipeline of MIN1PIPE (Online Methods), our method removed the strong uneven

255   background structure, and automatically identified 210 putative ROI components without

256   the need for additional manual selection (Fig. 3a, Supplementary Video S5). In comparison,

257   PCA/ICA identified 79 ROIs whereas CNMF identified 71 ROIs. Note that with the same

258   computer configuration and dataset (~28 GB), the CNMF ran into memory issues and could

259   not process the full-scale video, thus we cropped a center patch as the input video to the

260   CNMF (Fig 3b-d middle panels). The contours of the ROIs identified by the different

12

261 methods are drawn and superimposed onto the max projection of the neural enhanced data

262 (Fig. 3b). This reveals that MIN1PIPE can identify potentially all ROIs, whereas PCA/ICA

263 and CNMF miss a significant subset of ROIs. Meanwhile, both the PCA/ICA and CNMF

264 have some problems in separating overlapping ROIs and/or estimating the correct shape of

265 the ROIs, as revealed by the max projections of the extracted signals (Fig. 3c). The

266 projection of MIN1PIPE extracted signals closely resembles those of the neural enhanced

267 data (Fig. 3c, top). In comparison, the projection obtained using PCA/ICA has low SNR

268 with high background signals (Fig. 3c, bottom), whereas the projection derived from

269 CNMF shows unrealistic ROI shapes larger than the true shape of neurons, indicating that

270 the CNMF is sensitive to the contamination of the background dynamics (Fig. 3c, middle).

271 To further check the shape of individual ROIs, we choose to visualize one example ROI

272 embedded in the full imaging field using different methods (Fig. 3d). MIN1PIPE delineates

273 a well localized ROI footprint, whereas CNMF returns a less localized ROI difficult to

274 relate to the underlying neuron. PCA/ICA, on the other hand, fails to provide a localized

275 footprint as remnants of other ROI components can also be seen. To examine the temporal

276 traces extracted by the different methods, we selected 10 ROI components that were

277 identified by all three methods within the cropped image field and plotted their

278 corresponding calcium traces (Fig 3e, individual panels marked with 1-10). Again, the

279 spatial footprints of the 10 ROIs obtained through CNMF and PCA/ICA are less localized

280 as described above. In terms of calcium dynamics, both MIN1PIPE and CNMF give

281 denoised traces, whereas the traces obtained using PCA/ICA are noisy and show unrealistic

282 negative fluctuations. Furthermore, the traces extracted using CNMF include false positive

283 calcium events that likely reflect background noises (arrows in Fig. 3e). Supplementary

284 Video S6 shows the comparison of raw and processed results using three different methods

285 of the entire imaging video.

286 To further illustrate the general applicability of MIN1PIPE in processing miniscope

287 imaging data obtained over different brain areas and/or different animal models, we applied

288 it to process calcium imaging results from Area X in zebra finch and compared the results

289 with those obtained using the other two methods. Briefly, MIN1PIPE detected 55 ROI

290 components, whereas CNMF detected 15 and PCA/ICA detected 35 ROIs after manual

291 selection (Fig. 4a-b, Supplementary Video S7). Again, CNMF could only process a

292 cropped portion of the imaging video due to the computer memory issues. All of the ROIs

293 detected by CNMF and PCA/ICA are included in the set extracted by MIN1PIPE. In

294 addition to false negatives, CNMF also identifies a cluster of false positive ROIs that are

295 apparent upon visual inspections (Fig. 4b-c, e). The PCA/ICA again gives rise to

296 nonlocalized ROI footprints that contain other potential components (Fig. 4d). It was

297 known that Area X neurons show strong song selective activities when the bird is singing

298 (Goldberg and Fee, 2010; Kojima and Doupe, 2007; Woolley et al., 2014; Yazaki-

299 Sugiyama and Mooney, 2004), which can be used as partial ground truth to validate

300 MIN1PIPE. We plot the calcium traces of the ROIs identified by MIN1PIPE, and the

301 majority of the ROIs contain calcium events that are roughly phase-locked to the singing

302 onsets (Fig. 4f). Notably, a subset of the neurons shows precise singing-related activities

303 with minimal events unrelated to singing (Fig. 4g upper panel). Furthermore, we sorted

304 neurons according to their timing of peak calcium activities during each song production

305 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

15

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

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.

400   Goldberg, J. H., and Fee, M. S. (2010). Singing-related neural activity distinguishes four

401   classes of putative striatal neurons in the songbird basal ganglia. *J. Neurophysiol. 103*,

402   2002-2014.

403   Greenberg, D.S. and Kerr, J.N. (2009). Automated correction of fast motion artifacts for

404   two-photon imaging of awake animals. *J. Neurosci. Methods 176*, 1-15.

405   Grewe, B.F., Gründemann, J., Kitch, L.J., Lecoq, J.A., Parker, J.G., Marshall, J.D., Larkin,

406   M.C., Jercog, P.E., Grenier, F., Li, J.Z., et al. (2017). Neural ensemble dynamics

407   underlying a long-term associative memory. *Nature 543*, 670-675.

408   Hahn, T.T., Sakmann, B. and Mehta, M.R. (2006). Phase-locking of hippocampal

409   interneurons' membrane potential to neocortical up-down states. *Nat. Neurosci 9*, 1359-

410   1361.

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

412   1735-1780.

413   Jennings, J.H., Ung, R.L., Resendez, S.L., Stamatakis, A.M., Taylor, J.G., Huang, J.,

414   Veleta, K., Kantak, P.A., Aita, M., Shilling-Scrivo, K., et al. (2015). Visualizing

415   hypothalamic network dynamics for appetitive and consummatory behaviors. *Cell 160*,

416   516-527.

417   Kaifosh, P., Lovett-Barron, M., Turi, G.F., Reardon, T.R. and Losonczy, A. (2013). Septo-

418   hippocampal GABAergic signaling across multiple modalities in awake mice. *Nat.*

419   *Neurosci. 16*, 1182-1184.

420   Kaifosh, P., Zaremba, J.D., Danielson, N.B. and Losonczy, A. (2014). SIMA: Python

421   software for analysis of dynamic fluorescence imaging data. *Front. Neuroinform. 8*, 80.

422   Kamigaki, T. and Dan, Y. (2017). Delay activity of specific prefrontal interneuron subtypes

423   modulates memory-guided behavior. *Nat. Neurosci.* 20, 854-863.

424   Kitamura, T., Ogawa, S.K., Roy, D.S., Okuyama, T., Morrissey, M.D., Smith, L.M.,

425   Redondo, R.L. and Tonegawa, S. (2017). Engrams and circuits crucial for systems

426   consolidation of a memory. *Science 356*, 73-78.

427   Klaus, A., Martins, G.J., Paixao, V.B., Zhou, P., Paninski, L. and Costa, R.M. (2017). The

428   spatiotemporal organization of the striatum encodes action space. *Neuron 95*, 1171-1180.

429   Kojima, S. and Doupe, A.J. (2007). Song selectivity in the pallial-basal ganglia song circuit

430   of zebra finches raised without tutor song exposure. *J. Neurophysiol.* 98, 2099-2109.

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

432   Lombaert, H., Grady, L., Pennec, X., Ayache, N. and Cheriet, F. (2014). Spectral log-

433   demons: diffeomorphic image registration with very large deformations. *Int. J. Comput.*

434   *Vis. 107*, 254-271.

435   Lucas, B.D. and Kanade, T. (1981). An iterative image registration technique with an

436   application to stereo vision. In *Proc. 7th International Joint Conference on Artificial*

437   *Intelligence*, 674-679.

438   Markowitz, J.E., Liberti III, W.A., Guitchounts, G., Velho, T., Lois, C. and Gardner, T.J.

439   (2015). Mesoscopic patterns of neural activity support songbird cortical sequences. *PLoS*

440   *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.

500   Zhou, P., Resendez, S.L., Stuber, G.D., Kass, R.E. and Paninski, L. (2016). Efficient and

501   accurate extraction of in vivo calcium signals from microendoscopic video data. *arXiv,*

502   1605.07266.

503   Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A.

504   and Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat.*

505   *Neurosci. 16*, 264-266.

506   Figure 1

507
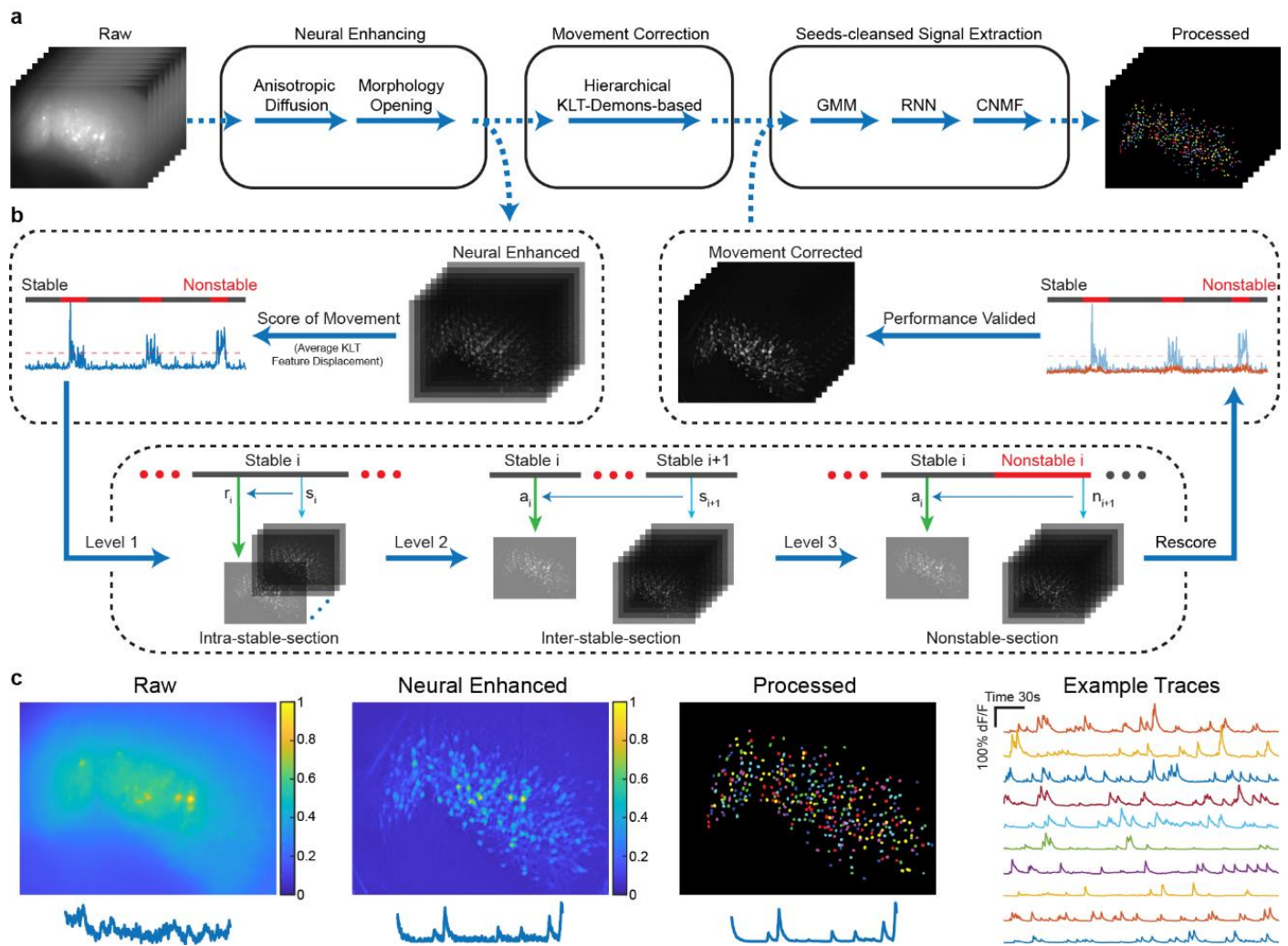
Fig. 1. The general pipeline and demonstrations of MIN1PIPE. **a.** The overall structure of

MIN1PIPE. MIN1PIPE takes the very raw miniscope imaging data freshly collected from

the imaging system as inputs, and returns fully processed ROI components with spatial

footprints and temporal calcium traces as outputs. The data are processed in series by neural

enhancing, hierarchical movement correction, and seeds-cleansed signal extraction

modules. Each module is composed of specific brick functions. **b.** A zoom-in of the

hierarchical movement correction module. The module is KLT-Log Demons based. It first

scores all the neural enhanced frames and divides them into stable and nonstable sections,

516   then register frames at three different levels. The movement corrected frames are then fed

517   into the seeds-cleansed neural signal extraction module. **c.** The max projecting

518   demonstrations of MIN1PIPE. The raw imaging video contains large, dominant

519   background fluctuation, the neural enhanced video contains mainly neural signals, and the

520   fully processed video contains only extracted and denoised neural signals as independent

521   ROI components. Example traces are randomly selected from the processed video. See also

522   Fig. S1-S4.

523   Figure 2
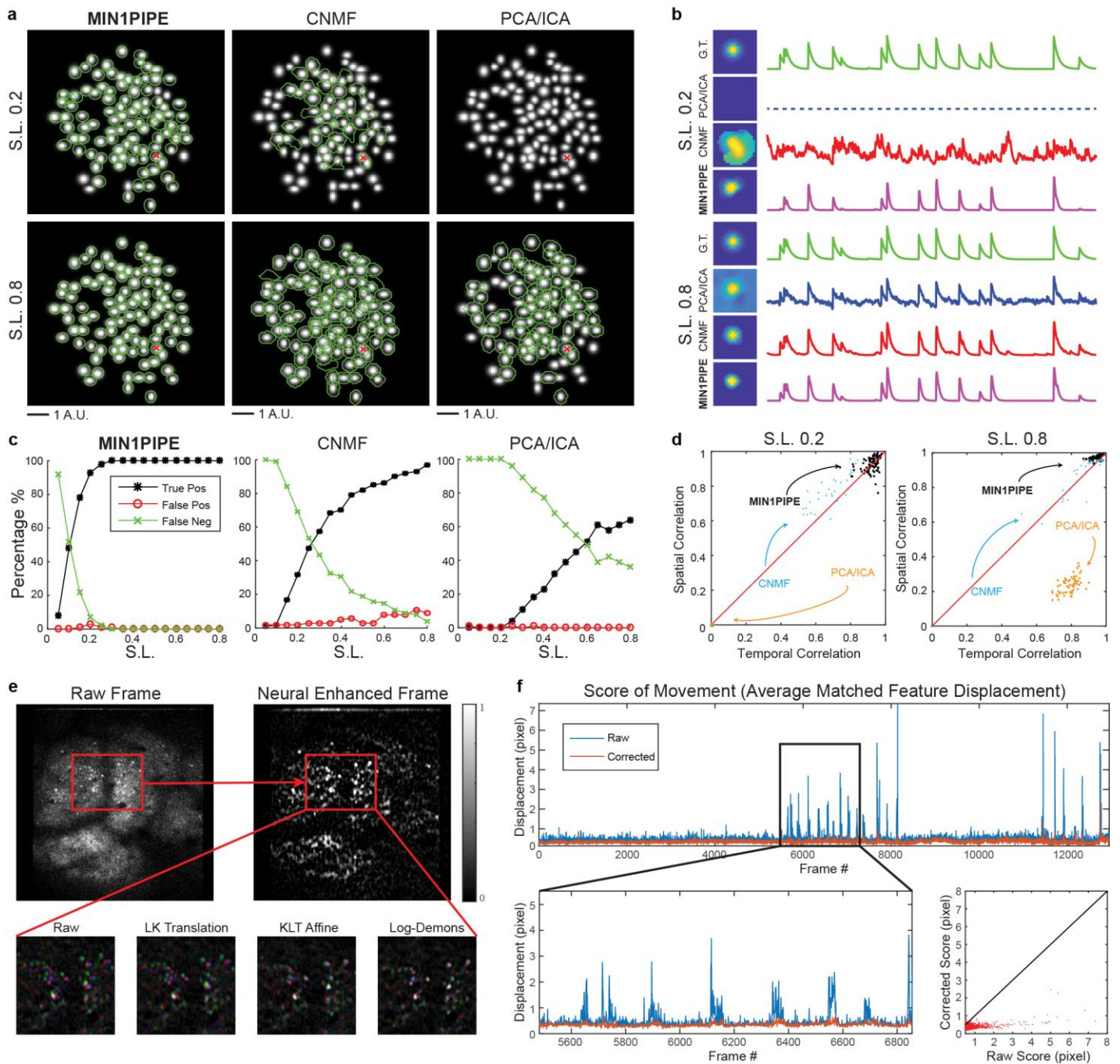
524

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

528   ratio of the amplitude between the signals and the background). A.U.: arbitrary unit. **b.** The

529   trace of an example ROI component identified by the three methods and the ground truth.

530   G.T.: ground truth. **c.** The identification precision and accuracy of the three methods. **d.**

531   The spatiotemporal identification accuracy with the three methods at two representative

532   signal levels. The similarity of the spatial footprints and temporal traces between the results

533   of the three methods and the ground truth of each ROI is plotted as a dot in the figure.

534   These quantifications confirm that MIN1PIPE not only outperforms the other two methods

535   in all aspects, but also verify that MIN1PIPE can provide satisfactory neural signal

536   extraction results of miniscope imaging data. **e-f.** The quantification of the movement

537   correction module. **e.** The demonstration of various image registration methods in a

538   consecutive three frames from two-photon imaging data. The deformation registration

539   algorithm we use in the module (Log-Demons) successfully corrects all the nonrigid

540   deformations within the frames that cannot be corrected by the methods that assume

541   translation or rigid transformation. The black and white color indicate overlapping of the

542   same structure between the frames, whereas other colors indicate nonaligned structures. **f.**

543   The score of movement before and after the correction. In general, the hierarchical

544   correction steps register large deformations while preserves the stable frames. See also Fig.

545   S5-S9.

546   Figure 3

547
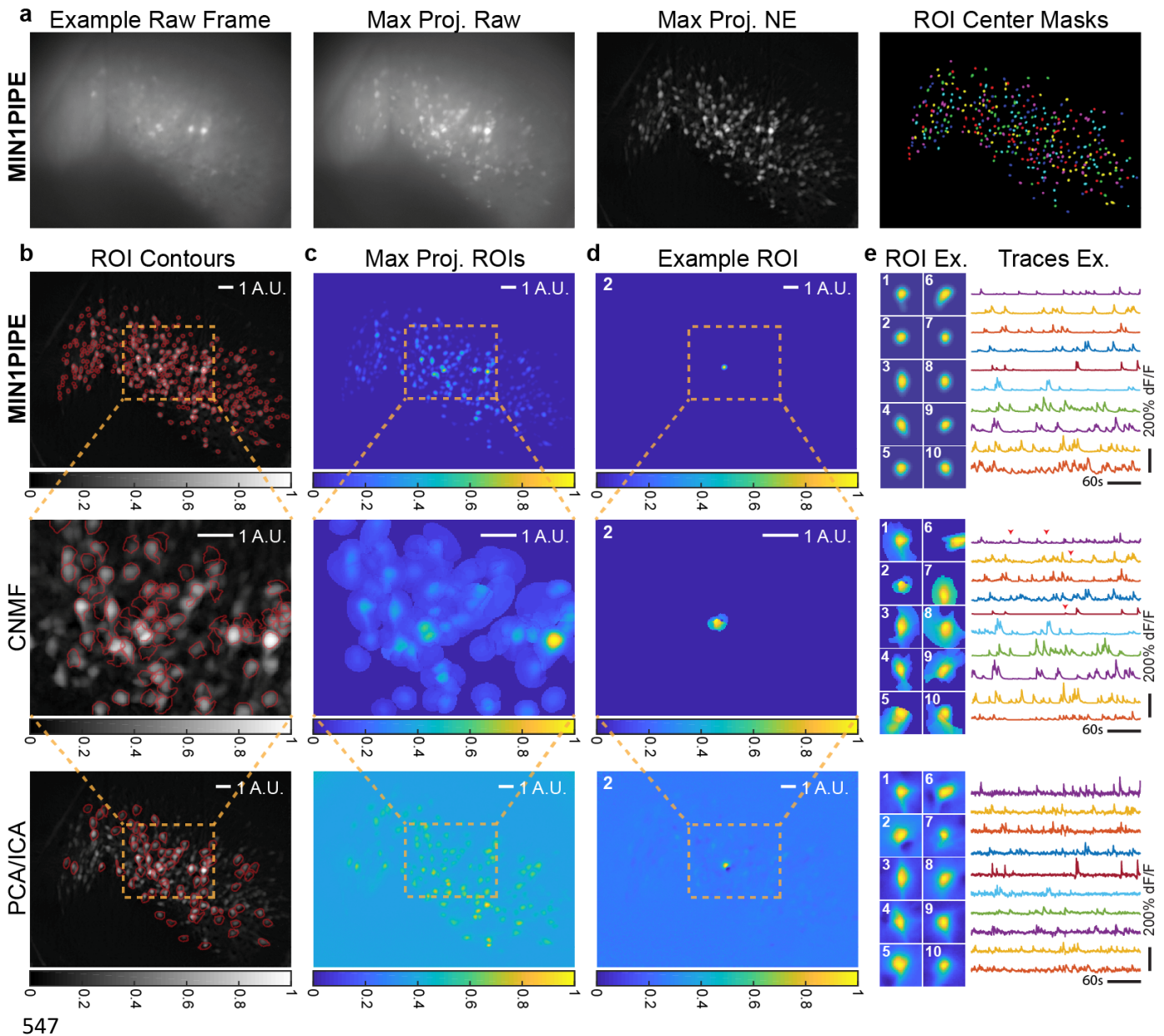
Fig. 3. Comparing different methods using miniscope imaging data from the mouse barrel

cortex. **a.** Demo of MIN1PIPE in processing imaging data from the barrel cortex. We show

an example raw frame, the max projection of raw video, the max projection of the neural

enhanced video, and the colored ROI masks. **b.** The identified ROI contours superimposed

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
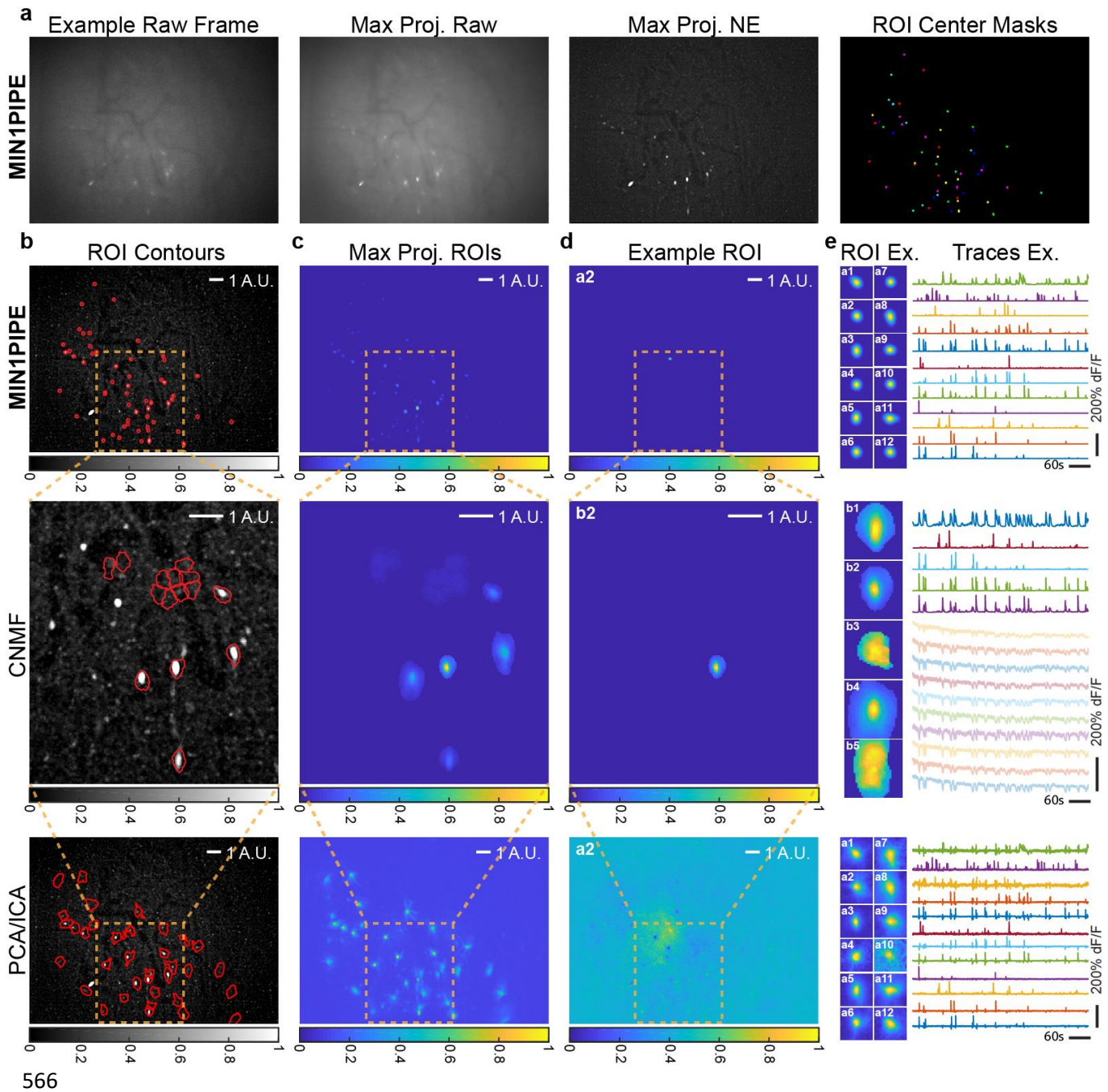
564   events indicated by red arrows).

565   Figure 4

566

Fig. 4. Comparing different methods using miniscope imaging data from the Area X in zebra finch. **a.** Demo of MIN1PIPE in processing data from Area X in the zebra finch brain. We show an example of the raw frame, the max projection of raw video, the max projection of the neural enhanced video, and the colored ROI masks. **b.** The identified ROI contours

32

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
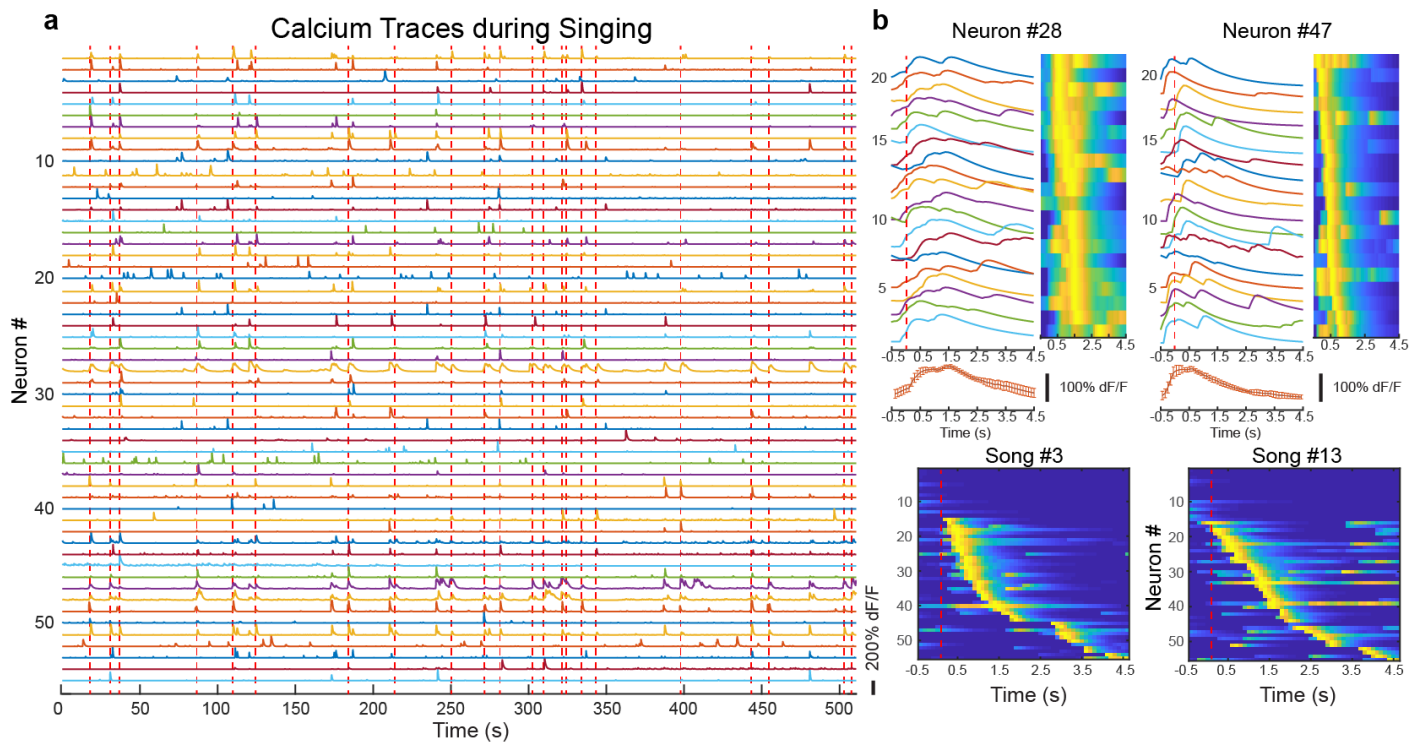
Fig. 5. Analysis of neural correlation in Area X of zebra finch to singing behavior. **a.** The

traces of all automatically detected ROIs with MIN1PIPE superimposed with complete

singing onsets in red dashed lines. Note that incomplete song events are not labeled and

taken into analysis. **b.** Upper panel: two example neurons with precise song selectivity.

The traces represent the neuronal activity of the first 20 complete song singing events with

a window of 0.5 second before and 4.5 seconds after the song onset. The heatmap is another

way of visualizing the analysis. The error bar graphs at the bottom show the trial average

traces of the two neurons. Lower panel: population activity pattern of two example song

singing events. The neurons are sorted to the latency of the peak. Red dashed lines indicate

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 $\mathbf{I}^{\mathrm{s}}$ 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 $\boldsymbol{\Phi}$ can well estimate the background $\mathbf{B}$ from the frame $\mathbf{I}$ adaptively. The opening operator is the combination of erosion $\ominus$ and dilation $\oplus$: $\mathbf{B} = (\mathbf{I} \ominus \boldsymbol{\Phi}) \oplus \boldsymbol{\Phi}$ (Van Den Boomgaard and Van Balen, 1992), where the morphological erosion returns minimum value $(\mathbf{I} \ominus \boldsymbol{\Phi})(x, y) = \min_{(s,t)\in\boldsymbol{\Phi}}\{\mathbf{I}(x + s, y + t)\}$ and dilation returns

1

619   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

620   window at each point. In practice, the choice of the structuring element should be comparable

621   to the overall size of the neurons in the imaging field. We choose the structuring element

622   to be of disk shape with a default size of 9 (pixels). The foreground is then computed as:

623   $\mathbf{I}^{\mathbf{f}} = \mathbf{I} - \mathbf{B}$. After the dynamic background removal, $\mathbf{I}^{\mathbf{f}}$ contains only neuronal information

624   with minimal background noise corruption.

625   **MOVEMENT CORRECTION**   The movement correction module performs a hierarchi-

626   cal registration framework over the frames of the imaging video. The neural enhanced video

627   is first scored with a measuring metric on the relative displacement between two neighboring

628   frames. The video is then segmented into stable or nonstable sections based on the score,

629   followed by three levels of registration: the intra-stable-section, inter-stable-section and

630   nonstable-section registration.

631   **Scoring Metric**   We track features containing corner information of every two neighboring

632   frames using the KLT tracker, and then calculate the average displacement of these features.

633   The KLT tracker first selects good features to track, where the feature points contains enough

634   information on both directions within the neighborhood p. The problem can be formulated

635   as $\nabla p = 0$ over the neighborhood. A good feature point is selected if the smaller eigenvalue

636   $\sigma_{min} \geq \sigma_0$ and the *condition number* $\kappa = \frac{\sigma_{max}}{\sigma_{min}} \leq \kappa_{max}$. Then the algorithm tracks the

637   feature points based on Newton-Raphson minimization on the normal equation:

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

638   where

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

2

639   $J_t^{\mathbf{f}}$ and $I_t^{\mathbf{f}}$ are the moving and fixed frame of the tth registration pair after neural enhancing

640   respectively (in this case, $J_t^{\mathbf{f}} = I_{t+1}^{\mathbf{f}}$), and $\omega(\boldsymbol{x})$ is the window of the neighborhood. For the

641   scoring, we choose 31 pixels as the default window size of the neighborhood. To segment

642   the frames into stable/nonstable sections, we compute the minimum value between the 3

643   times of median absolute deviation (MAD) above median and upper limit threshold (0.5

644   pixel as default) to be the threshold of segmentation. Any frame whose score is above

645   the threshold is considered as one belonging to some stable section. The frames that are

646   connected together are labeled as stable sections while the remaining connected components

647   are nonstable ones.

648   **Intra-stable-section Registration**   Displacements with small magnitude or consistent

649   directions can be approximated as translational displacement. Therefore, after the segmen-

650   tation, movements within the stable sections can be considered translational. Within each

651   stable section, an efficient displacement matching method with subpixel resolution is used

652   to register the neighboring frames. Here we use a similar Lucas-Kanade tracker to track

653   the center of the moving frames with a sufficiently large window (80% of the frame size)

654   between two neighboring frames. However, when frame number within a section grows, the

655   registration error may cumulate for later frames. Therefore, in practice we update the fixed

656   frame as reference every second and all moving frames within that period are aligned to the

657   fixed frame. This is acceptable due to the slow calcium dynamics so that the imaging field

658   remains essentially similar within a second.

659   **Inter-stable-section Registration**   The imaging field is stable within each stable section

660   after the intra-section registration. However, the imaging fields of different stable sections

661   are not necessarily similar. Therefore, we need to align different stable sections without

662   specific assumptions of the movement type. We first extract the overall information of each

663   stable section by averaging the frames with significant neuronal activation, we then sort

3

664 these sectional images $\mathbf{I}_i^{\mathrm{sec}}$ according to their similarity. To register the ith sectional image

665 based on the sorted similarity, a reference image is generated by calculating the weights $\hat{x}$

666 using the least square regression between the registered frames and the current sectional

667 image:

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

668 where

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

669 The reference image is then reshaped into two dimensional matrix $\mathbf{R}_i^{\mathrm{sec}}$ from $\boldsymbol{b}^{\mathrm{ref}} = \mathbf{A}^{\mathrm{ref}}\hat{x}$.

670 Once we have the moving and fixed sectional images, we apply rigid preregistration to the

671 moving image using similar KLT tracker. To finally correct for the nonrigid deformation, we

672 apply the diffeomorphic Log-Demons to the preregistered image $\hat{\mathbf{I}}_i^{\mathrm{sec}}$ based on the reference

673 $\mathbf{R}_i^{\mathrm{sec}}$. In general, the diffeomorphic Log-Demons is a non-parametric registration method,

674 which aims at finding the displacement of all pixels by minimizing the global energy:

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

675 It consists of the similarity, correspondence and regularization items, where $\mathbf{F}$ and $\mathbf{M}$ are

676 fixed and moving image respectively, and σ_i, σ_x and σ_T account for similarity, spatial

677 uncertainty on the correspondence and regularization level. In specific, Log-Demons

678 represents everything in log domain. It spells out two regularizing terms, fluid and diffusion,

679 and uses a Lie group structure to impose diffeomorphism. In our implementation, we choose

680 $\sigma_i = \sigma_x = 1$, and $\sigma_{fluid} = \sigma_{diffusion} = 3$. Because all the frames within the same stable

681 section share the same imaging field, the deformation field of the current stable section is

682 then applied to warp all the frames. After registering all the stable sections iteratively, all

683 the stable frames are efficiently aligned.

4

684 **Nonstable-section Registration**   At the third level, only nonstable sections remain to be

685 aligned. We use a similar approach as in the inter-section registration, where we first

686 preregister the current frame to the reference image, and then apply Log-Demons to finely

687 align the frame. Similarly, the reference image is computed based on some of the registered

688 frames. For the ith nonstable section, this set of frames includes the last frame of the ith

689 stable section, the first frame of the $i + 1$th stable section, and the frames that have already

690 been registered in the ith nonstable section.

691 **SEEDS-CLEANSED SIGNAL EXTRACTION**   We generate and cleanse the potential

692 seeds of ROIs in this section, and create the initial spatial regions and the time series of

693 ROIs for later refinement of the spatiotemporal signals.

694 **Over-complete Seeds Initialization**   To generate the initial set containing centers of all

695 potential ROIs, we construct a *randomized max pooling* process to create an over-complete

696 set of seeds. This process first randomly selects a portion of frames $\mathbf{I}^{\mathsf{s}}_{\boldsymbol{\alpha}} = \{\mathbf{I}^{\mathsf{s}}_t\}_{t \in \boldsymbol{\alpha}}, \boldsymbol{\alpha} \subseteq$

697 $\{1, \ldots, T\}$, where $\boldsymbol{\alpha}$ is a randomized subset of frame number. We then compute the max-

698 projection map across selected frames $\mathbf{I}^{\max} = \max(\mathbf{I}^{\mathsf{s}}_{\boldsymbol{\alpha}})$, and further detect all the local max

699 points on this map as $\mathcal{S}$. We repeat the above procedures multiple times, and collect the

700 union of each map $\mathcal{S}$ as the final over-complete set of neural seeds $\mathcal{S}_{\text{seeds}}$. Our randomized

701 max pooling can improve the true positives compared with max-pooling over the entire video,

702 because a real seed can be buried in the uneven florescence of an ROI over a long period but

703 can most likely be discovered in a small temporal vicinity. To exclude false positive seeds

704 from the set, we propose the following two-stage algorithm for seeds refinement.

705 **Seeds Refinement with GMM**   The temporal properties of ROIs and non-ROIs can be

706 very different. In one aspect, the ROIs often have prominent peak-valley difference d, while

707 non-ROIs tend to be less spiky. We assume $\{d_s | s \in \mathcal{S}_{\text{seeds}}\}$ are generated from a mixture

5

708   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,

709   mixture proportion of the ith Gaussian component. Therefore, we can cluster the seeds

710   based on their probabilities of belonging to each component, and only consider those having

711   higher probabilities to the Gaussian with a larger mean as positive seeds.

712   **Seeds Refinement with LSTM**   To select true ROI seeds from the over-complete set,

713   we need to classify the seeds based on their patterns of neuronal calcium dynamics. We

714   propose to employ Recurrent Neural Networks (RNNs) (LeCun et al., 2015) for calcium

715   signal sequence classification. We offline trained the RNNs with a separate training dataset,

716   composed of both positive and negative sequence chunks of length $T_0 = 100$, obtained

717   with 10Hz frame rate. The training data for the RNN module were selected from a separate

718   real dataset. The labels were manually selected by experienced neuroscientists with a

719   conservative standard: the positive trials were the ones with most obvious calcium dynamics

720   that were aligned to the peak of the spike, whereas the negative trials were randomly selected

721   from the rest of the data. The training dataset contained 1000 positive and 1000 negative

722   labels, and both the validation and the test set contained 600 balanced trials. Specifically,

723   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 $\boldsymbol{y}_n$

724   and output label $l_n \in \{0, 1\}$. Our goal is to learn model parameters $\boldsymbol{\theta}$ to best characterize

725   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

726   sequence classification, the input is a sequence, $\boldsymbol{y} = \{y_1, \ldots, y_{T_0}\}$, where y_t is the input

727   data at time t. There is a corresponding hidden state vector $\boldsymbol{h}_t \in \mathbb{R}^K$ at each time t, obtained

728   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*

729   *weights*, and $\mathbf{U}$ is *recurrent weights*. The ouput c for our classification is defined as the

730   corresponding *decoding function* $p(c|\boldsymbol{h}_{T_0}; \mathbf{V}) = \sigma(\mathbf{V}\boldsymbol{h}_{T_0})$, where $\sigma(\cdot)$ denotes the *logistic*

731   *sigmoid function*, and $\mathbf{V}$ is *decoding weights*.

732   The transition function $g(\cdot)$ can be implemented with a *gated* activation function, such as

733   LSTM (Hochreiter and Schmidhuber, 1997) or a Gated Recurrent Unit (GRU) (Cho et al.,

734   2014). Both LSTM and GRU have been proposed to address the issue of learning long-term

735   sequential dependencies. Each LSTM unit has a cell containing a state $\mathtt{c}_t$ at time t. This cell

736   can be viewed as a memory unit. Reading or writing the memory unit is controlled through

737   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

738   updated as follows:

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

739   where $\odot$ represents the element-wise matrix multiplication operator. Note that the training of

740   RNNs is completed off-line, only the efficient testing stage is performed for seeds refinement.

741   In the testing stage, given an input $\tilde{\boldsymbol{y}}$ (with missing label $\tilde{l}$), the estimate for the output is

742    $p(\tilde{l}|\tilde{\boldsymbol{y}}, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}} = \arg\max \log p(\mathcal{D}|\boldsymbol{\theta})$. In our practical application, the testing sequence

743    $\bar{\boldsymbol{y}} \in \mathbb{R}^T$ is often of length $T > T_0$. We first convert it into a bag of subsequences $\{\tilde{\boldsymbol{y}}_i\}_{i=1}^{T-T_0+1}$,

744   with a sliding window of width T_0 and moving step size 1. The well trained LSTMs are then

745   used to label the subsequences. We consider $\bar{\boldsymbol{y}}$ as positive if at least one subsequence in its

746   bag is classified as "calcium spike" ($l = 1$), otherwise negative. Intuitively, this means that

747   we only care about the patterns of neural spikes, regardless of their temporal positions.

748   **Seeds Merging**   Once the set of seeds is cleansed, there is still a low possibility of identi-

749   fying multiple seeds within a single ROI. Therefore, we merge all these redundant seeds

750   by computing the temporal similarity of seeds within their neighborhood, and preserving

751   the one with maximum intensity. Specifically, we compute the similarity based on phase-

752   locking information (Hahn et al., 2006). For a sequence, we extract the instantaneous phase

753   dynamics using Hilbert transform, and only consider the subsequences containing prominent

754   peaks, because all the pixels within the same ROI are highly correlated only during the

755   calcium spiking periods, but not necessarily during baseline period. After seeds merging,

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

757 **Spatial and Temporal Initialization**   The time series of the kth seed is used as initial

758 guess of temporal signal $\hat{s}_k$. The spatial map of the corresponding ROI, $\hat{r}_k$, is estimated by

759 pooling neighbor pixels with temporal similarity above a threshold. Because of our seeds

760 generating and cleansing approach, the spatiotemporal initialization in our method can be

761 readily parallelized. Then we fine-tune the corresponding spatial and temporal guess, by

762 applying semi-NMF:

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

763 where $\boldsymbol{z}_k \subset \mathbf{I}^\mathsf{s}$ is a region containing the kth ROI. The spatial map and the temporal signal

764 are updated respectively:

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

765 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

766 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,

767 the background parameters $\boldsymbol{b}^0, \boldsymbol{f}^0$ are also estimated by the same semi-NMF procedure,

768 using $(\mathbf{I}^\mathsf{s} - \sum_{k=1}^K \boldsymbol{r}_k \boldsymbol{s}_k^\top)$. $\mathbf{S}^0, \mathbf{R}^0, \boldsymbol{b}^0, \boldsymbol{f}^0$ are then used as initializations of the spatiotemporal

769 signal refinement.

770 **Spatiotemporal Signal Refinement**   We perform the iterative spatial and temporal op-

771 timizations to update the spatial footprints of individual ROIs, and the temporal traces

772 with deconvolved spike trains, as proposed in CNMF. Specifically, for the neural enhanced

773 video $\mathbf{X}^\mathsf{f}$, the method decomposes $\mathbf{X}^\mathsf{f}$ into a spatial dictionary $\mathbf{R}^\mathsf{f} \in \mathbb{R}^{P \times K}$ representing

774 individual ROIs, and corresponding temporal dynamics matrix $\mathbf{S}^\mathsf{f} \in \mathbb{R}^{T \times K}$, in addition to

8

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

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

776   Similarly, with the rank-1 assumption on $\mathbf{B}^{\mathrm{f}}$, CNMF decomposes the background $\mathbf{B}^{\mathrm{f}} =$

777    $\boldsymbol{b}^{\mathrm{f}}\boldsymbol{f}^{\mathrm{f}^{\top}}$, where $\boldsymbol{b}^{\mathrm{f}} \in \mathbb{R}^{P}$ and $\boldsymbol{f}^{\mathrm{f}} \in \mathbb{R}^{T}$. Meanwhile, $\mathbf{S}^{\mathrm{f}}$ is also correlated with underlying

778   action potential events: $\mathbf{A}^{\mathrm{f}} = \mathbf{S}^{\mathrm{f}^{\top}}\mathbf{G}^{\mathrm{f}}$, where $\mathbf{A}^{\mathrm{f}} \in \mathbb{R}^{K \times T}$, and $\mathbf{G}^{\mathrm{f}} \in \mathbb{R}^{T \times T}$ is the coefficient

779   matrix of the low-order autoregression. The variables $\mathbf{R}^{\mathrm{f}}, \mathbf{S}^{\mathrm{f}}, \boldsymbol{b}^{\mathrm{f}}, \boldsymbol{f}^{\mathrm{f}}$ are therefore estimated

780   via iteratively alternating between the following two steps [26].

781   **Estimating Spatial Variables**   Given the estimates of temporal variables $\mathbf{S}^{\mathrm{f}^{(\ell-1)}}$ and

782    $\boldsymbol{f}^{\mathrm{f}^{(\ell-1)}}$ from the last iteration, the spatial parameters can be updated by solving the problem:

$$\min_{\mathbf{R}^{\mathrm{f}}, \boldsymbol{b}^{\mathrm{f}}} \|\mathbf{R}^{\mathrm{f}}\|_{1}, \; s.t. \; \mathbf{R}^{\mathrm{f}}, \boldsymbol{b}^{\mathrm{f}} \geq 0$$

$$\|\mathbf{X}^{\mathrm{f}}(i,:) - \mathbf{R}^{\mathrm{f}}(i,:)\mathbf{S}^{\mathrm{f}^{(\ell-1)\top}} - \boldsymbol{b}^{\mathrm{f}}(i)\boldsymbol{f}^{\mathrm{f}^{(\ell-1)\top}}\| \leq \sigma_{i}\sqrt{T}$$

783   where $\mathbf{X}^{\mathrm{f}}(i,:)$ is the ith row of the matrix $\mathbf{X}^{\mathrm{f}}$, $\boldsymbol{b}^{\mathrm{f}}(i,:)$ is the ith element of the vector. $\epsilon_{i}\sqrt{T}$

784   is the empirically selected noise residual constraint of the corresponding pixel. This is

785   essentially a basis pursuit denoising problem, and it is solved by the least angular regression

786   (Efron et al., 2004) in implementation.

787   **Estimating Temporal Variables**   Given the estimates of spatial variables $\mathbf{R}^{\mathrm{f}}, \boldsymbol{b}^{\mathrm{f}}$ and

788   temporal parameters $\mathbf{S}^{\mathrm{f}^{(\ell-1)}}, \boldsymbol{f}^{\mathrm{f}^{(\ell-1)}}$ from the last iteration, the temporal parameters can be

789   updated by solving the problem:

$$\min_{\mathbf{S}^{\mathrm{f}}, \boldsymbol{f}^{\mathrm{f}}} \sum_{k=1}^{K} \mathbf{1}^{\top}\mathbf{G}^{\mathrm{f}}\mathbf{S}^{\mathrm{f}}, \; s.t. \; \mathbf{G}^{\mathrm{f}}\boldsymbol{s}_{k}^{\mathrm{f}} \geq 0, k = 1, \cdots, K$$

$$\|\mathbf{X}^{\mathrm{f}}(i,:) - \mathbf{R}^{\mathrm{f}}(i,:)\mathbf{S}^{\mathrm{f}^{(\ell-1)}} - \boldsymbol{b}^{\mathrm{f}}(i)\boldsymbol{f}^{\mathrm{f}^{(\ell-1)\top}}\| \leq \sigma_{i}\sqrt{T}$$

9

790   Unlike in CNMF, where the spatial footprints are serially updated and subtracted from

791   the preceding residuals, we extract spatial footprints from the original data that does not

792   depend on preceding iterations. Therefore, the information loss/duplication is reduced and

793   the optimization procedures can be parallelized in our method.

794   **Data availability**   The data that support the findings of this study are available from the

795   corresponding authors upon request.

796   **Code availability**   The codes will be freely available upon publication, and the beta version

797   is available for reviewers upon request.

10