

1 **Reinforcing neuron extraction and spike inference in calcium**
2 **imaging using deep self-supervised learning**

3 Xinyang Li^{1,2,3†}, Guoxun Zhang^{1,3†}, Jiamin Wu^{1,3}, Yuanlong Zhang^{1,3}, Zhifeng Zhao^{1,3},

4 Xing Lin^{1,3,4}, Hui Qiao^{1,3}, Hao Xie^{1,3}, Haoqian Wang^{2,3*}, Lu Fang^{3,5*} & Qionghai Dai^{1,3*}

5 *¹Department of Automation, Tsinghua University, Beijing 100084, China*

6 *²Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen,*
7 *518055, China*

8 *³Institute for Brain and Cognitive Science, Tsinghua University, Beijing 100084, China.*

9 *⁴Beijing Innovation Center for Future Chips, Tsinghua University, Beijing 100084,*
10 *China*

11 *⁵Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

12 *[†]These authors contributed equally to this work.*

13 **Correspondence: wanghaoqian@tsinghua.edu.cn; fanglu@tsinghua.edu.cn;*

14 *qh dai@mail.tsinghua.edu.cn*

15 **ABSTRACT**

16 **Calcium imaging is inherently susceptible to detection noise especially when imaging with**
17 **high frame rate or under low excitation dosage. We developed DeepCAD, a self-**
18 **supervised learning method for spatiotemporal enhancement of calcium imaging without**
19 **requiring any high signal-to-noise ratio (SNR) observations. Using this method, detection**
20 **noise can be effectively suppressed and the imaging SNR can be improved more than**
21 **tenfold, which massively improves the accuracy of neuron extraction and spike inference**
22 **and facilitate the functional analysis of neural circuits.**

23 Calcium imaging enables parallel recordings of large neuronal ensembles in living animals¹⁻⁴
24 and offers a new possibility for deciphering information propagation, integration, and
25 computation in neural circuits⁵. To obtain accurate neuron extraction and spike inference for
26 downstream neuroscience analysis, high-SNR calcium imaging is desired. However, due to the
27 paucity of fluorescence photons caused by low peak accumulations and fast dynamics of *in*
28 *vivo* calcium transients^{6,7}, calcium imaging is easy to be contaminated by detection noise (*i.e.*
29 photon shot noise and electronic noise), especially in functional imaging where high temporal
30 resolution is particularly important for analyzing neural activities⁸.

31 To capture sufficient fluorescence photons for high-SNR calcium imaging, the most direct
32 way is to use high excitation dosage, but concurrent photobleaching, phototoxicity^{9,10}, and
33 tissue heating¹¹ are detrimental for sample health and photosensitive biological processes,
34 which limits the maximal excitation power for long-term *in vivo* imaging¹². More effective

35 strategies include using brighter calcium indicators^{7,13} and more sensitive photoelectric
36 detectors¹⁴, but their performances are still largely restricted in photon-limited conditions such
37 as dendritic imaging and deep-tissue imaging. Apart from these physical or biological
38 approaches, data-driven methods are promising to offer an alternative solution to recover
39 faithful signals from degraded recordings and reduce the photon budget of calcium imaging.
40 As an intelligent signal processing technique, deep learning has been adopted by microscopists
41 and achieved impressive performance in fluorescence imaging¹⁵⁻¹⁸. However, calcium
42 transients are highly dynamic, non-repetitive activities and a firing pattern cannot be
43 captured twice. Previous schemes for obtaining ground-truth images (*i.e.* clean images
44 without noise contamination or high-SNR images with the same underlying scene) by
45 extending integration time or averaging multiple noisy frames are no longer feasible,
46 posing an entrenched obstacle for conventional supervised learning methods.

47 In this paper, we present DeepCAD, a self-supervised learning method for calcium imaging
48 denoising by over tenfold SNR improvement without requiring any high-SNR observations for
49 training. DeepCAD is based on the insight that a deep neural network can converge to a mean
50 estimator even the target image used for training is another corrupted sampling of the same
51 scene¹⁹. When looking at calcium imaging data, we explored the temporal redundancy of
52 pervasive video-rate imaging and found that any two consecutive frames can be regarded as
53 two independent samplings of the same underlying firing pattern, which can be used for
54 training of denoising models. Furthermore, the input and output data are designed to be
55 3D volumes rather than 2D frames to fully exploit spatiotemporal information in the

56 time-lapse stack. We show that such a 3D self-supervised method is extremely effective for
57 calcium imaging denoising and even the subtlest calcium fluctuations induced by a single
58 action potential (AP) can be restored from severely corrupted images. Finally, a Fiji-based
59 plugin along with a pre-trained model were released to make our method easy to access and
60 convenient to use.

61 The general principle of DeepCAD is schematized in Fig. 1a. For network architecture, we
62 employed 3D U-Net²⁰ to aggregate spatiotemporal information in multiple frames using
63 3D convolutional layers (Supplementary Fig. 1, Methods), which endows DeepCAD with
64 better denoising capability than 2D architecture or classical methods (Supplementary Fig. 2).
65 Benefiting from the self-supervised strategy, a single low-SNR stack of ~3500 frames is
66 sufficient to be a complete training set. To generate the training set, two sub-stacks consisting
67 of interlaced frames were split from the original low-SNR stack and 3D tiles were extracted
68 from these sub-stacks for training (Supplementary Fig. 3). They contain approximate identical
69 calcium transients when the original stack was imaged at near video rate, which is common for
70 commercial or customized microscopes. After proper training, interpretable features can be
71 learned (Supplementary Fig. 4) and the model can be applied to subsequent acquisitions
72 without extra training (Fig. 1b). Although the network was trained on specified spatial and
73 temporal resolution, we found that it had non-inferior performance on various frame rates
74 (Supplementary Fig. 5) and magnifications (Supplementary Fig. 6), indicating the great
75 scalability and generalization for versatile applications of DeepCAD.

76 To quantitatively evaluate the performance of DeepCAD, we first validated it on simulated
77 calcium imaging data of different imaging SNRs (Supplementary Figs. 7-8 and Supplementary
78 Notes 1-2), which contains synchronous noise-free recordings as the ground truth for
79 comparison. The constrained nonnegative matrix factorization (CNMF) algorithm²¹ was
80 used for downstream neuron extractions (Methods). After the enhancement of
81 DeepCAD, more active neurons can be detected, especially when imaging SNR is low
82 (Fig. 1c). The accuracy of neuron extraction was also quantified with F1 score and
83 significant improvement was observed across a wide range of intersection-over-union (IoU)
84 thresholds (Fig. 1d,e). For a typical IoU threshold of 0.7, the segmentation accuracy was
85 improved by 2.4 folds (0.84 contrast to 0.35). Benefiting from the improved imaging
86 quality, calcium traces extracted from the denoised data possess higher fidelity. To
87 investigate the temporal enhancement of DeepCAD, we extracted calcium traces of all
88 neurons from both raw noisy data and the enhanced counterpart. The Pearson
89 correlation with the clean traces was significantly improved after denoising (Fig. 1f).
90 Even the slightest calcium transients can be restored from the original noisy data (Fig.
91 1g and Supplementary Fig. 9). These facts suggest that the spatiotemporal enhancement
92 of DeepCAD can improve the accuracy of neuronal localization and trace extraction
93 and largely facilitate the analysis of neural circuits.

94 To verify the effectiveness and reliability of DeepCAD on neuroscience research, we then
95 demonstrated its performance on two-photon calcium imaging based on data released by
96 Svoboda lab⁷. In this dataset, simultaneous cell-attached electrophysiological recordings (Fig.

97 2a) are synchronized with two-photon imaging and serve as the reference of calcium transients
98 and the ground truth of spike inference. Contaminated by detection noise, both the spatial
99 footprint and temporal traces of the neuron were severely corrupted in the original data (Fig.
100 2b). After we applied DeepCAD to enhance these data, the annular cytoplasm became
101 recognizable and calcium traces were liberated from noise (Fig. 2c and Supplementary Video
102 1). Even the most imperceptible calcium transients evoked by one AP, two APs, and three APs
103 were clearly distinguished and still maintain their original dynamics (Fig. 2d-g), which
104 otherwise would be submerged in noise. For further comparison, we extracted single-pixel
105 fluorescence from cytoplasmic pixels and found that calcium transients can be unveiled at a
106 single-pixel scale (Supplementary Fig. 10). Moreover, we performed spike inference (Methods)
107 on traces extracted from the original data as well as the corresponding denoised data. Owing to
108 the improvement of imaging SNR, the error rate of spike inference was consequently decreased
109 (Fig. 2h and Supplementary Fig. 11). Among 107 independent calcium traces, 86% of them
110 were observed to have lower error rates.

111 Next, we employed DeepCAD for noise removal of calcium imaging of large neuronal
112 populations in awake mice. To obtain high-SNR recordings for validation of our method, we
113 designed and built a two-photon imaging system with the capability of simultaneous low-SNR
114 and high-SNR recording (Supplementary Fig. 12 and Methods). The high-SNR detection path
115 was strictly synchronized with the low-SNR detection path but with about 10-fold higher
116 imaging SNR (Supplementary Fig. 13), which can be used as the reference for our denoising
117 results. We first imaged spontaneous neuropil activities in cortical layer 1 of a transgenic mouse

118 expressing GCaMP6f and found that calcium fluctuations indiscernible in original low-SNR
119 recordings can be effectively recovered by DeepCAD (Fig. 3a-c and Supplementary Video 2).
120 The imaging SNR was improved more than 10 folds considering that the SNR of enhanced
121 recordings even surpasses corresponding high-SNR reference. Fluorescence traces of dendritic
122 pixels can be accurately resolved and keep high consistency with the high-SNR reference (Fig.
123 3d-e and Supplementary Fig. 14). We also applied DeepCAD to enhance calcium imaging of
124 somatic signals. After denoising, neuronal distribution and circuit dynamics can be recognized
125 from a single frame (Fig. 3f-h and Supplementary Video 3). Using CNMF as the downstream
126 source extraction method, 52.6% (229 contrast to 150) more active neurons can be extracted
127 (Fig. 3i,j and Supplementary Fig. 15) and the trace peak SNR of extracted neurons was also
128 improved more than two folds (9.9 contrast to 4.8, median value) (Fig. 3k), indicating that the
129 functional analysis of large neuronal populations can be effectively strengthened due to
130 improved SNR.

131 In summary, we demonstrate DeepCAD, a deep self-supervised learning-based method for
132 spatiotemporal enhancement of calcium imaging. Quantitative evaluation on both simulated
133 and experimental data shows that the accuracy of neuron extraction and spike inference can be
134 largely reinforced after denoising. To fully evaluate the capability and reliability of our method,
135 a customized two-photon microscope was built to capture synchronized low-SNR and high-
136 SNR recordings, which indicates that DeepCAD enables a more than tenfold improvement in
137 imaging SNR. To maximize its accessibility, we released an open-source Fiji plugin
138 (Supplementary Fig. 16 and Supplementary Notes 3) and a pre-trained DeepCAD model for

139 two-photon imaging of neuron populations. Our method can be efficiently configured on a
140 common desktop and achieve comparable performance on different imaging systems
141 regardless of objectives and detectors (Supplementary Fig. 17 and Supplementary Video 4).
142 Although DeepCAD is currently investigated only on two-photon microscopy, it can be easily
143 extended to other imaging modalities such as wide-field microscopy and light-sheet
144 microscopy. We anticipate that this method could serve as a general processing step for calcium
145 imaging in photon-limited conditions and promote long-term and high-fidelity recording of
146 neural activities.

147 **Methods**

148 **Optical setup.** A two-photon imaging system was designed to capture strictly
149 synchronized low-SNR and high-SNR calcium recordings for validation of our method.
150 Our system was based on a standard two-photon laser scanning microscope (2PLSM)
151 and the detection path was specially designed to split the fluorescence in a ratio of 1:10.
152 All components of our imaging system are commercially available or easy to fabricate.
153 The schematic of the custom-built two-photon microscope is shown in Supplementary
154 Fig. 12. At the forefront of the optical path, a titanium-sapphire laser system with tunable
155 wavelength (Mai Tai HP, Spectra-Physics) was used as the illumination source to emit
156 the linearly polarized, femtosecond-pulsed Gaussian excitation beam (920 nm central
157 wavelength, pulse width <100 fs, 80 MHz repetition rate). A half-wave plate
158 (AQWP10M-980, Thorlabs) was used to adjust the polarization of the laser beam. Then
159 the laser beam went through an electro-optic modulator (350-80LA-02, Conoptics) to
160 modulate the excitation power and the half-wave plate was rotated to make the electro-
161 optic modulator have maximal extinction ratio. A 4f system composed of two
162 achromatic lenses (AC508-200-B, Thorlabs) with the same focal length was followed
163 to collimate the laser beam. Another 4f system (AC508-100-B and AC508-400-B,
164 Thorlabs) with a fourfold magnification was used to expand the laser beam and guide
165 the beam into a galvo-resonant scanner (8315K/CRS8K, Cambridge Technology) for
166 fast optical scanning. The scanner mount was optimally designed for reliable and
167 distortion-free scanning. Then the beam went through a scan lens (SL50-2P2, Thorlabs)

168 and a tube lens (TTL200MP, Thorlabs) and converged into a tight focus through a high
169 numerical aperture (NA) water-dipping objective (25×/1.05 NA, XLPLN25XWMP2,
170 Olympus). A high-precision piezo actuator (P-725, Physik Instrumente) was
171 additionally used to drive the objective for fast axial scanning. The beam size at the
172 back aperture of the objective was further restricted with an iris set behind the beam
173 expander (L4) to keep the back aperture of the objective underfilled. The effective
174 excitation NA was about 0.5 in our imaging experiments.

175 For the detection path, fluorescence excited by the Gaussian focus was first
176 collected by the objective. High-NA detection is helpful to detect more fluorescence
177 photons and improve the signal intensity. A long-pass dichroic mirror (DMLP650L,
178 Thorlabs) was used to separate fluorescence by reflecting the fluorescence signals and
179 transmitting the excitation light. A 1:9 (reflectance: Transmission) non-polarizing plate
180 beam splitter (BSN10, Thorlabs) was then placed in the detection path. All fluorescence
181 going through the beam splitter will be split into a 10% component (low-SNR path) and
182 a 90% component (high-SNR path), propagating in two orthogonal directions and
183 detected by two photomultiplier tubes (PMT1001, Thorlabs). A pair of fluorescence
184 filters (MF525-39, Thorlabs; ET510/80M, Chroma) was configured in front of each
185 PMT to fully block wavelengths outside the emission passband of green fluorescent
186 protein (GFP). To improve detection efficiency, we conjugated the back aperture of the
187 objective to the sensor planes of the two PMTs using two 4f systems (TTL200-A and
188 AC254-050-A, Thorlabs). The two detection paths recorded synchronized fluorescence

189 signals but with quite different imaging SNR. Although the high-SNR recording still
190 suffers from noise, it can be used as the reference to identify underlying structures and
191 calcium fluctuations. The field-of-view (FOV) of our two-photon imaging system is
192 about 600 μm and the frame rate is about 30 Hz.

193 **System calibration.** To confirm the fluorescence intensity ratio between the high-SNR
194 detection path and the low-SNR detection path, we imaged 1 μm green-fluorescent
195 beads (G0100, ThermoFisher) for system calibration. The beads suspension was first
196 diluted and embedded in 1.0% agarose and then mounted on a microscope slide to form
197 a single beads layer composed of sparse beads. A specified region was continuously
198 scanned to acquire 500 consecutive frames. These frames can be regarded as
199 independent samplings of the same underlying scene. To reduce the impact of detection
200 noise, we averaged these frames to obtain the noise-free image of each path
201 (Supplementary Fig. 13). All beads in the FOV were manually segmented and the
202 intensity of each bead was calculated by averaging all pixels inside its segmentation
203 mask. According to our statistical analysis, the fluorescence intensity of the high-SNR
204 detection path was approximately tenfold higher than that of the low-SNR detection
205 path.

206 **Mouse preparation and calcium imaging.** All experiments involving mice were
207 performed in accordance with institutional guidelines for animal welfare and have been
208 approved by the Institutional Animal Care and Use Committee (IACUC) of Tsinghua
209 University.

210 Adult transgenic mice (Ai148D/Rasgrf2-dCre) at 8-12 postnatal weeks were
211 anesthetized with 1.5% isoflurane and craniotomy surgeries were conducted using a
212 stereotaxic instrument (68018, RWD Life Science) under a bright-field binocular
213 microscope (77001S, RWD Life Science). A custom-made coverslip fitting the shape
214 of the cranial window (~6 mm in diameter) was embedded and cemented to the skull.
215 A biocompatible titanium headpost was then cemented to the skull for stable head
216 fixation. The edge of the cranial window was enclosed with dental cement to hold the
217 immersion water of the objective. After the surgery, 0.25mg/g body weight of
218 Trimethoprim (TMP) was intraperitoneally injected to induce the expression of
219 GCaMP6f genetically encoded calcium indicator (GECI) in layer 2/3 neurons across
220 the whole brain. To reduce potential inflammation, 5 mg/kg body weight of Ketoprofen
221 was injected subcutaneously. Each mouse was housed in a separate cage for 1-2 weeks
222 of postoperative recovery.

223 Imaging experiments were carried out when the cranial window became clear and
224 no inflammation occurred. Mice were first rapidly anesthetized with 3.0% isoflurane
225 and then fixed onto a custom-made holder with the headpost. The mouse holder was
226 mounted on a precision translation stage with three motorized axes (M-VP-25XA-
227 XYZL, Newport) to find the region of interest (ROI) for imaging. The correction ring
228 of the objective was adjusted to compensate for the coverslip thickness and eliminate
229 spherical aberrations. The excitation power after the objective was kept below 140 mW
230 in all experiments to avoid potential laser-induced tissue damage. Gaseous anesthesia

231 was switch off and the mice kept awake during the whole imaging process.

232 **Network architecture and training details.** The network architecture of DeepCAD
233 employs 3D U-Net, which is reported to have superior performance on the segmentation
234 of volumetric data²⁰. In general, the network is composed of a 3D encoder module (the
235 contracting path), a 3D decoder module (the expanding path), and three skip
236 connections from the encoder module to the decoder module (Supplementary Fig. 1). In
237 the 3D encoder module, there are three encoder blocks. Each block consists of two
238 $3\times 3\times 3$ convolutional layers followed by a leaky rectified linear unit (LeakyReLU) and
239 a $2\times 2\times 2$ max pooling with strides of two in three dimensions. In the decoder module,
240 there are three decoder blocks, each of which contains two $3\times 3\times 3$ convolutional layers
241 followed by a LeakyReLU and a 3D nearest interpolation. A group normalization²²
242 layer is configured after each convolutional layer. The skip connections link low-level
243 features and high-level features by concatenating their feature maps. All operations
244 (convolutions, max poolings, and interpolations) in the network are in 3D to aggregate
245 spatial information and temporal information. For the loss function, we used the
246 arithmetic average of a L1-norm loss term and a L2-norm loss term. The model was
247 trained on 3D tiles with a spatial size of 64×64 pixels and a temporal size of 300 frames.
248 Small spatial size can lower memory requirements and reduce the training time, and
249 large temporal size is helpful to make full use of temporal information.

250 Adam optimizer²³ was used for network training with a learning rate of 0.00005 and
251 exponential decay rates of 0.5 for the first moment and 0.9 for the second moment. We

252 used Graphics Processing Units (GPU) to accelerate the training and test process. It
253 took about 12 hours to train our model for 20 epochs on a typical training set (about
254 1200 3D tiles) with a single GPU (Nvidia TITAN RTX, 24 GB memory). Training time
255 can be further shortened by using a more powerful GPU or parallelizing the training
256 process on multiple GPUs.

257 The full 3D architecture of DeepCAD makes it easy to overfit because 3D
258 convolutions usually involve more parameters than the 2D counterpart. The best
259 denoising performance is only achieved at the point where there is neither underfitting
260 nor overfitting. To screen out the model with the best generalization ability, we saved
261 the network snapshot after each training epoch and evaluated its performance on a
262 holdout validation set. We fed the validation data into each model and calculated the
263 standard deviation projection of the output stack of each model. Then, the average pixel
264 intensity was calculated on a small dark region (*e.g.* blood vessels or a small region
265 without neural activity during the recording) of all standard deviation projections. The
266 best model was selected to be the one with the smallest dark standard deviation.

267 **Data simulation.** Our simulation program includes a step for synthesizing the noise-
268 free video (ground truth) and a step for adding the Mixed Poisson-Gaussian (MPG)
269 noise (Supplementary Notes 1-2). Firstly, to generate realistic simulated calcium imaging
270 data, we constructed a neuron library containing the spatial profiles of 517 neurons.
271 These neurons were extracted using the constrained nonnegative matrix factorization
272 algorithm²¹ (CNMF) from an experimentally obtained two-photon calcium imaging

273 data of a virus-transfected wild-type mouse expressing GCaMP6f (layer 2/3 at the
274 primary somatosensory cortex). For the spatial component that defines the location of
275 each neuron, 120 neurons were randomly selected from the library to keep the sparsity
276 of neurons. For the temporal component that defines the fluorescence fluctuations of
277 each neuron, MLspike²⁴ was employed to generate calcium traces with GCaMP6f
278 kinetics. Then, these two components were reshaped into 2D matrices and the simulated
279 noise-free data (1 μm /pixel spatial sampling rate, 30 Hz frame rate) was synthesized as
280 the product of the spatial matrix and the temporal matrix. The noise-contaminated
281 counterpart was ultimately generated by adding the content-related MPG noise. Data
282 with different imaging SNRs were simulated with different relative photon numbers.
283 Their relationship was investigated in Supplementary Fig. 7. All images were saved as
284 uncompressed tif files with the format of unsigned 16-bit integer (uint16). More details
285 of data simulation and related mathematical models are described in Supplementary
286 Notes 1-2.

287 **Single-neuron recordings.** The data of simultaneous two-photon imaging and
288 electrophysiological recordings of single-neuron activities were released by the
289 Svoboda lab²⁵ and were downloaded from the Collaborative Research in Computational
290 Neuroscience (CRCNS) platform. Only recordings of GCaMP6f neurons were used in
291 this study. The image stacks were fourfold downsampled to reduce the sampling rate
292 and some outlier recordings with very sparse spikes and low electrophysiological SNR
293 were excluded. Fluorescence traces were extracted from temporal stacks using

294 manually annotated cytoplasmic masks. For spike inference, we used the MLspike
295 algorithm²⁴, which was reported to rank first in the *Spikefinder* challenge²⁶. All traces
296 were divided by their mean values for normalization before fed into the spike inference
297 pipeline. Recommended model parameters for GCaMP6f indicator were used to ensure
298 optimal performance of spike inference.

299 **Data analysis of neuronal populations.** Calcium imaging data of large neuronal
300 populations were first registered with a non-rigid motion correction method²⁷ and the
301 black edges of registered images were clipped. Then, CNMF²¹ was employed as the
302 source extraction method for neuron segmentation and trace extraction. A spatial matrix
303 and a temporal matrix can be obtained from each video, storing the spatial footprints
304 and corresponding calcium traces of all active neurons, respectively. The same set of
305 parameters was used for the original low-SNR recording and corresponding DeepCAD
306 enhanced counterpart, as well as the high-SNR recording. Simulated data were analyzed
307 following the same pipeline except motion correction. Along with automatic neuron
308 extraction, we also performed manual annotations to inspect our results. High-SNR
309 recordings were tenfold downsampled along the time axis by averaging each
310 consecutive ten frames, which reduced the disturbance of detection noise and was
311 helpful to improve annotation accuracy. Boundaries of all active components were
312 annotated using the ROI Manager toolbox of Fiji. The final segmentation masks were
313 generated through subsequent morphological operations of images and connected
314 domain extraction implemented with customized MATLAB scripts.

315 **Performance metrics.** Two types of metrics were used for quantitative evaluation of
316 the spatial and temporal performance of DeepCAD. For synthetic calcium imaging data,
317 corresponding clean images and ground-truth calcium traces were available. SNR and
318 PSNR were used as the spatial metric to evaluate pixel-level similarity between
319 DeepCAD enhanced images and ground-truth images. Pearson correlation coefficient
320 (R) was used as the temporal metric to reflect the similarity between enhanced traces
321 and ground-truth traces. The Pearson correlation between signal x and the reference
322 signal y is defined as

$$323 \quad R = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}$$

324 where μ_x and μ_y are the mean values of signal x and y , respectively; σ_x and σ_y are the
325 standard deviations of signal x and y , respectively; E represents arithmetic mean.

326 Furthermore, we also evaluated the performance of DeepCAD based on more
327 complex downstream tasks such as neuron extraction and spike inference, which are
328 the most crucial prerequisites in functional analysis of neural circuits from calcium
329 imaging data. We considered neuron extraction as an instance segmentation problem
330 and adopted an object-level metric to evaluate segmentation performance²⁸. Different
331 intersection-over-union (IoU, defined as the intersection area divided by the union area
332 of two objects) thresholds were selected to determine correctly segmented objects. For
333 a specified IoU threshold, the segmentation accuracy (F1 score) was defined as the
334 harmonic mean of sensitivity and precision, which can be formulated as

335
$$F1 = \frac{2TP}{2TP + FP + FN} .$$

336 Here, TP, FP, and FN are the number of true positives, false positives, and false
337 negatives, respectively. When applied CNMF as the source extraction method, the SNR
338 of calcium traces was quantified with the peak SNR automatically calculated by the
339 CaImAn toolbox²⁹ with infinite outliers eliminated. For spike inference, we used the
340 error rate (ER) to quantify the performance of spike inference, which is defined as ER
341 = 1- F1. Spikes detected from simultaneous electrophysiological recordings were used
342 as the ground truth for ER calculation. The evaluation process was implemented with
343 customized MATLAB scripts. SNR, PSNR, Pearson correlation coefficient, and IoU
344 were computed using built-in functions.

345 **Data availability**

346 Our data will be made publicly available post peer-review.

347 **Code availability**

348 Our python code and Fiji plugin will be made publicly available post peer-review.

349 **Acknowledgements**

350 We would like to acknowledge Y. Tang and Y. Yang at the School of Medicine of Tsinghua
351 University for providing transgenic mice for imaging and the mesoscope imaging data for
352 cross-system validation. We thank the Svoboda lab at Janelia Research Campus for releasing

353 their data of simultaneous electrophysiology and two-photon recording. This work was
354 supported by the National Natural Science Foundation of China (62088102, 61831014,
355 61531014 and 6181001011) and the Shenzhen Science and Technology Project under Grant
356 (ZDYBH201900000002 and JCYJ20180508152042002).

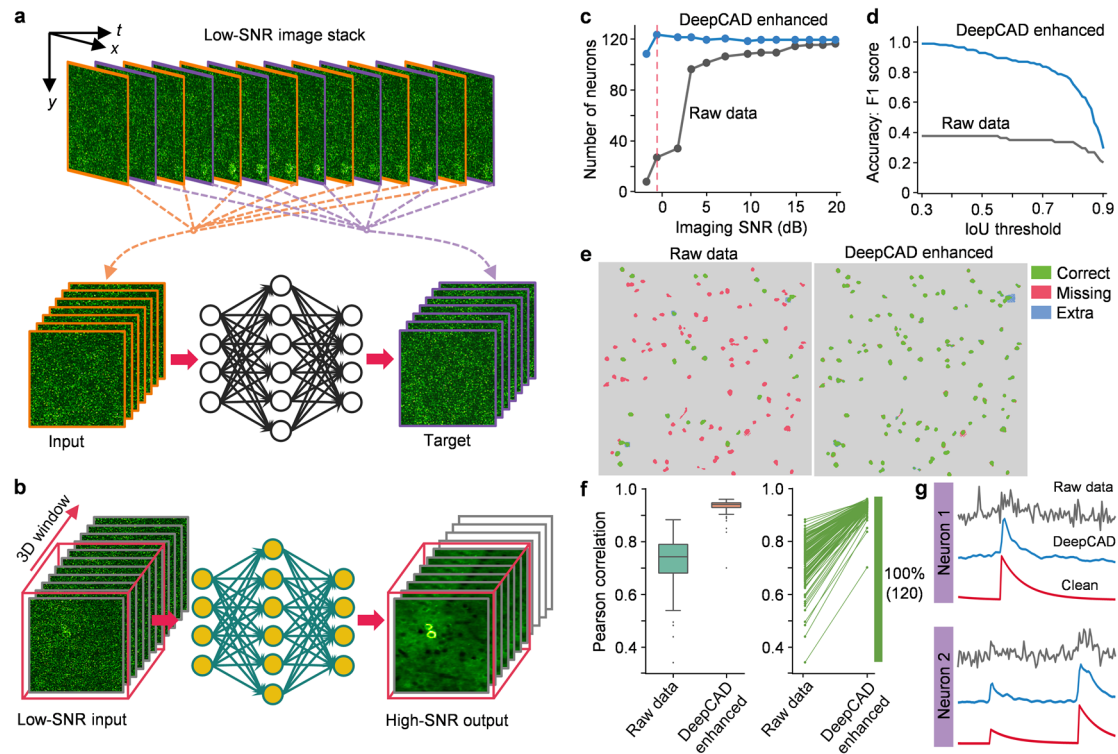
357 **Author Contributions**

358 Q. D., H. W., L. F. and XY. L. conceived this project. Q. D., H. W. and L. F. supervised this
359 research. XY. L. and G. Z. designed detailed implementations and processed the data. XY. L.
360 designed and set up the imaging system. XY. L. and G. Z. conducted the experiments. G. Z.
361 developed the python code and the Fiji plugin. J. W., Y. Z., and X. L. directed the experiments
362 and data analysis. L. F., Y. Z., Z. Z., H. Q. and H. X. gave critical support on system setup and
363 imaging procedure. J. W., L. F., Y. Z., X. L., H. Q., H. X., H. W. and Q. D. gave critical
364 discussions on the results. All authors participated in the writing of the paper.

365 **References**

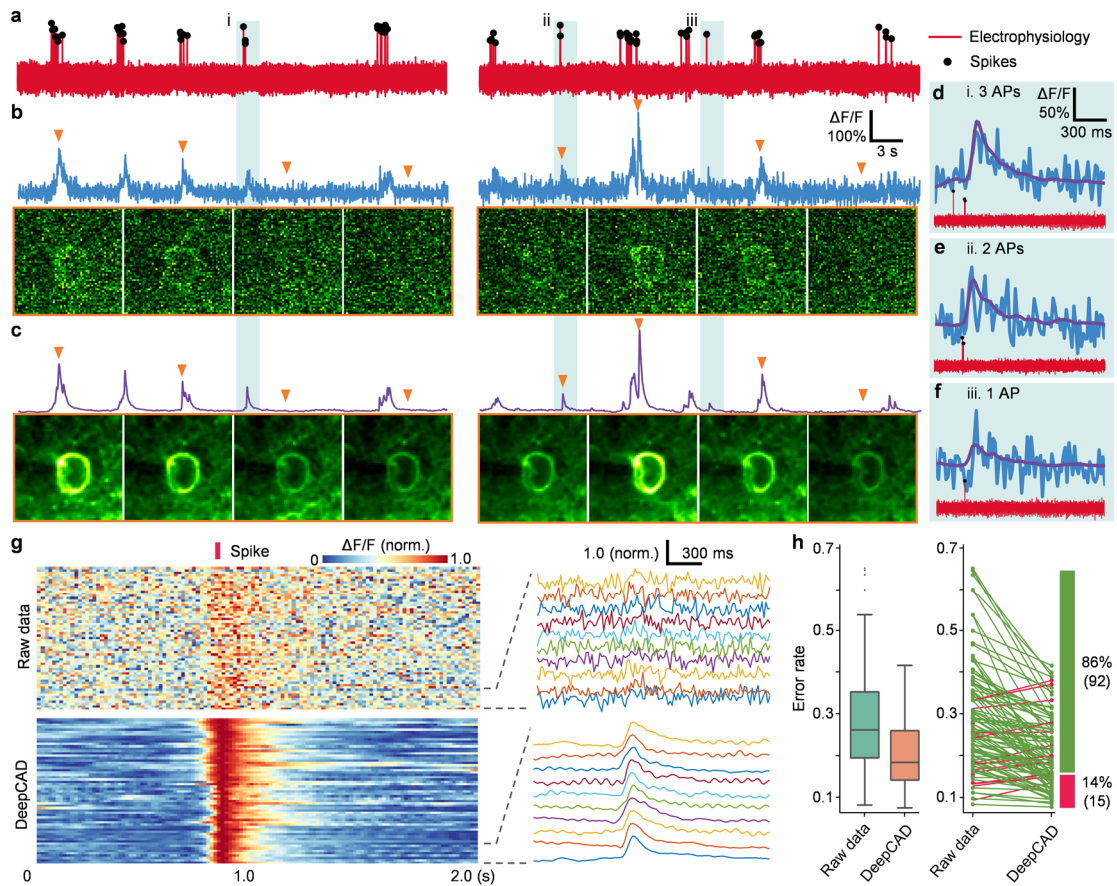
- 366 1. Grienberger, C. & Konnerth, A. Imaging Calcium in Neurons. *Neuron* **73**, 862-
367 885 (2012).
- 368 2. Lu, R. et al. Video-rate volumetric functional imaging of the brain at synaptic
369 resolution. *Nat. Neurosci.* **20**, 620-628 (2017).
- 370 3. Weisenburger, S. et al. Volumetric Ca(2+) Imaging in the Mouse Brain Using
371 Hybrid Multiplexed Sculpted Light Microscopy. *Cell* **177**, 1050-1066 e1014
372 (2019).
- 373 4. Chow, D.M. et al. Deep three-photon imaging of the brain in intact adult
374 zebrafish. *Nat. Methods* **17**, 605-608 (2020).
- 375 5. Calarco, J.A. & Samuel, A.D. Imaging whole nervous systems: insights into
376 behavior from worms to fish. *Nat. methods* **16**, 14-15 (2019).
- 377 6. Sabatini, B.L., Oertner, T.G. & Svoboda, K. The Life Cycle of Ca²⁺ Ions in
378 Dendritic Spines. *Neuron* **33**, 439-452 (2002).
- 379 7. Chen, T.W. et al. Ultrasensitive fluorescent proteins for imaging neuronal
380 activity. *Nature* **499**, 295-300 (2013).
- 381 8. Ji, N., Freeman, J. & Smith, S.L. Technologies for imaging neural activity in
382 large volumes. *Nat. Neurosci.* **19**, 1154-1164 (2016).
- 383 9. Svoboda, K. & Yasuda, R. Principles of two-photon excitation microscopy and
384 its applications to neuroscience. *Neuron* **50**, 823-839 (2006).
- 385 10. Skylaki, S., Hilsenbeck, O. & Schroeder, T. Challenges in long-term imaging
386 and quantification of single-cell dynamics. *Nat. Biotechnol.* **34**, 1137-1144
387 (2016).
- 388 11. Podgorski, K. & Ranganathan, G. Brain heating induced by near-infrared lasers
389 during multiphoton microscopy. *J. Neurophysiol.* **116**, 1012-1023 (2016).
- 390 12. Wang, T. et al. Quantitative analysis of 1300-nm three-photon calcium imaging
391 in the mouse brain. *eLife* **9** (2020).
- 392 13. Dana, H. et al. High-performance calcium sensors for imaging activity in
393 neuronal populations and microcompartments. *Nat. Methods* **16**, 649-657
394 (2019).
- 395 14. Samantaray, N., Ruo-Berchera, I., Meda, A. & Genovese, M. Realization of the
396 first sub-shot-noise wide field microscope. *Light Sci. Appl.* **6**, e17005 (2017).
- 397 15. Weigert, M. et al. Content-aware image restoration: pushing the limits of
398 fluorescence microscopy. *Nat. Methods* **15**, 1090-1097 (2018).
- 399 16. Belthangady, C. & Royer, L.A. Applications, promises, and pitfalls of deep
400 learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215-1225
401 (2019).
- 402 17. Wang, H. et al. Deep learning enables cross-modality super-resolution in
403 fluorescence microscopy. *Nat. Methods* **16**, 103-110 (2019).
- 404 18. Ouyang, W. et al. Deep learning massively accelerates super-resolution
405 localization microscopy. *Nat. Biotechnol.* **36**, 460-468 (2018).

- 406 19. Lehtinen, J. et al. Noise2Noise: Learning Image Restoration without Clean Data.
407 In *Proc. 35th International Conference on Machine Learning* (eds Dy, J. &
408 Krause, A.) 2965–2974 (PMLR, 2018).
- 409 20. Çiçek, Ö. et al. 3D U-Net: learning dense volumetric segmentation from sparse
410 annotation. In *Medical Image Computing and Computer-Assisted Intervention*
411 424–432 (2016).
- 412 21. Pnevmatikakis, E.A. et al. Simultaneous Denoising, Deconvolution, and
413 Demixing of Calcium Imaging Data. *Neuron* **89**, 285-299 (2016).
- 414 22. Wu, Y. & He, K. Group normalization. In *2018 IEEE European Conference on*
415 *Computer Vision (ECCV)* 3–19 (IEEE, 2018).
- 416 23. Kingma, D. P. & Ba, J. L. Adam: a method for stochastic optimization. In
417 *International Conference on Learning Representations* 1–15 (2015).
- 418 24. Deneux, T. et al. Accurate spike estimation from noisy calcium signals for
419 ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat.*
420 *Commun.* **7**, 12190 (2016).
- 421 25. GENIE project, Janelia Farm Campus, HHMI; Karel Svoboda (contact). (2015).
422 Simultaneous imaging and loose-seal cell-attached electrical recordings from
423 neurons expressing a variety of genetically encoded calcium indicators.
424 CRCNS.org. <http://dx.doi.org/10.6080/K02R3PMN>
- 425 26. Berens, P. et al. Community-based benchmarking improves spike rate inference
426 from two-photon calcium imaging data. *PLoS Comput. Biol.* **14**, e1006157
427 (2018).
- 428 27. Pnevmatikakis, E.A. & Giovannucci, A. NoRMCorre: An online algorithm for
429 piecewise rigid motion correction of calcium imaging data. *J. Neurosci.*
430 *Methods* **291**, 83-94 (2017).
- 431 28. Caicedo, J.C. et al. Nucleus segmentation across imaging experiments: the 2018
432 Data Science Bowl. *Nat. Methods* **16**, 1247-1253 (2019).
- 433 29. Giovannucci, A. et al. CaImAn an open source tool for scalable calcium imaging
434 data analysis. *eLife* **8** (2019).



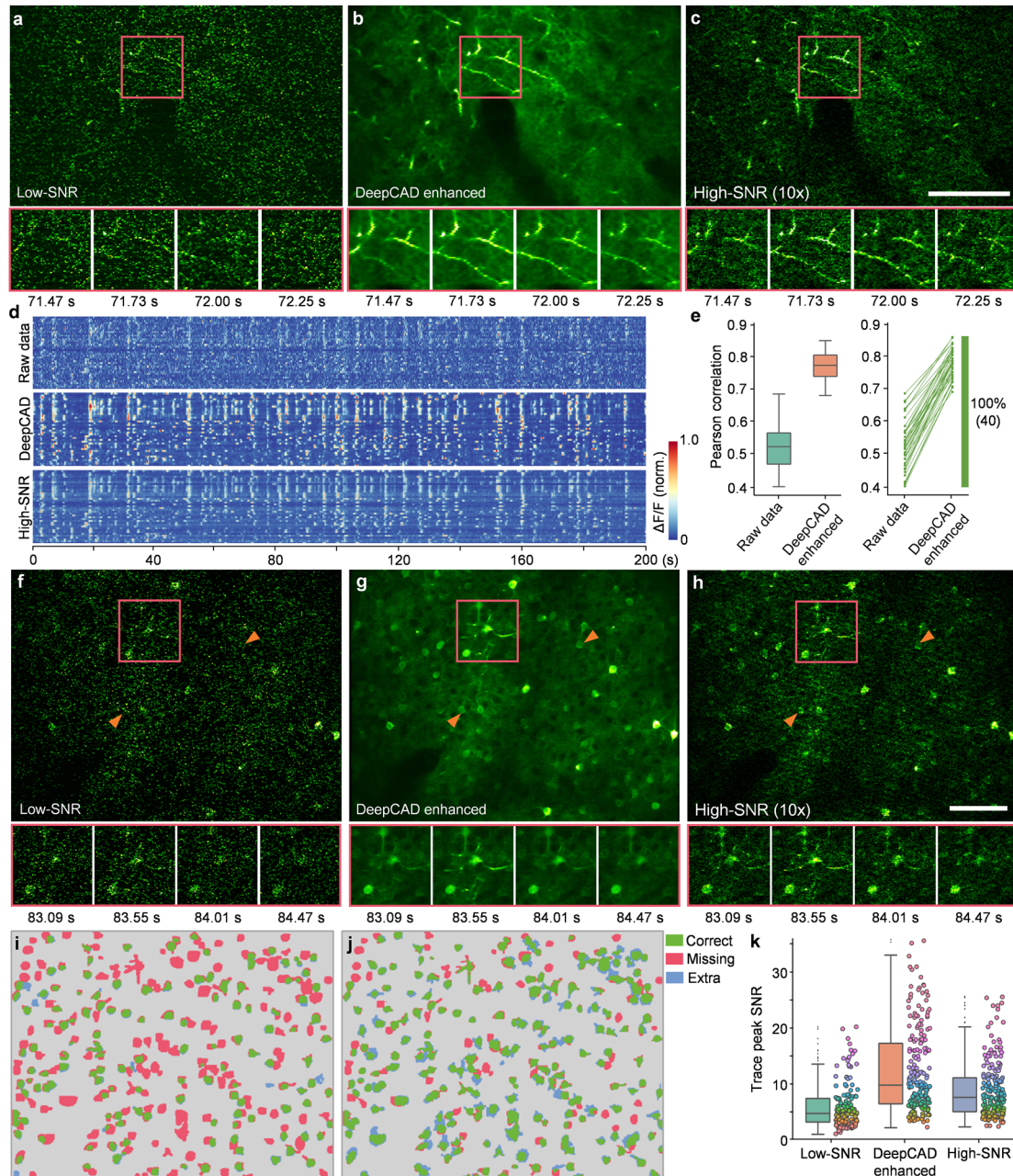
435

436 **Fig. 1 | General principle and validation of DeepCAD.** **a**, Self-supervised training
 437 strategy of DeepCAD. Consecutive frames in the original low-SNR stack are divided
 438 into two sub-stacks, used as the input volume and corresponding target volume to train
 439 a deep neural network (3D U-Net). After training, a denoising model can be established
 440 and memorized in network parameters. **b**, Application of the DeepCAD model. For
 441 subsequent acquisitions, a 3D (x - y - t) window traverses the entire stack and 3D tiles are
 442 sequentially fed into the pre-trained model. Denoised recordings will be obtained after
 443 the processing of the model. **c**, The number of neurons extracted under different
 444 imaging SNRs before and after the enhancement of DeepCAD. $N=120$ active neurons
 445 were simulated in the field of view (FOV). **d**, Accuracy of neuron segmentation
 446 quantified with F1 score at different intersection-over-union (IoU) thresholds (imaging
 447 SNR=-0.7 dB, indicated by the red dashed line in **c**). **e**, Spatial profiles of extracted
 448 neurons (imaging SNR=-0.7 dB). Correctly segmented regions (true positive) are
 449 colored green. Missing (false negative) and extra regions (false positive) are colored
 450 red and blue, respectively. Neuron extraction was implemented with CNMF²¹. **f**, Left:
 451 boxplot showing the distribution of Pearson correlation coefficients with clean traces
 452 before and after denoising ($N=120$). Right: increases of trace correlations. Each line
 453 represents one of 120 calcium traces and correlation coefficients of all neurons were
 454 observed improved. **g**, Calcium transients indiscernible from noise (gray) can be
 455 restored by DeepCAD (blue). Traces without noise contamination (red) serve as the
 456 ground truth for comparison.



457

458 **Fig. 2 | Spatiotemporal enhancement of DeepCAD.** a, Electrophysiological
 459 recordings of neural activities from a single neuron. Detected spikes are marked with
 460 black dots. b, Two-photon calcium imaging of the same neuron synchronized with cell-
 461 attached electrophysiology. Both spatial footprints and temporal traces of the neuron
 462 were severely corrupted in detection noise. Representative frames indicated with
 463 orange triangles are presented below the trace. c, Fluorescence traces and representative
 464 frames after the enhancement of DeepCAD. d-f, The most imperceptible calcium
 465 transients evoked by three APs (d), two APs (e), and one AP (f) can be resolved and
 466 still keep their original dynamics noise removal. g, Calcium fluctuations evoked by 61
 467 isolated action potentials. All spikes were normalized and temporally aligned with the
 468 red bar. Zoom-in traces are shown in the right panel. h, Left: Boxplot showing the
 469 distribution of error rates (lower is better) of spike inference for calcium traces extracted
 470 from enhanced data compared with those extracted from the original data (N=107).
 471 Real spike timings were revealed by simultaneous cell-attached recordings. Right:
 472 decreases of the error rate of spike inference. Each line represents one of 107 recordings,
 473 using green for decreased error rates and red for increased error rates.



474

475 **Fig. 3 | DeepCAD reinforces the recording of large neuronal populations.**
 476 Spontaneous neuropil activities in layer 1 of the mouse cortex captured by the low-SNR
 477 detection path. **b**, Images restored from the low-SNR recording using DeepCAD. **c**,
 478 Synchronized recording acquired by the high-SNR detection path (10-fold imaging
 479 SNR). Magnified views of the boxed regions show calcium transients in a ~0.8 s time
 480 window. Scale bar 100 μ m. **d**, Fluorescence traces extracted from 40 dendritic pixels.
 481 Top: low-SNR recording, Middle: DeepCAD enhanced recording, Bottom: high-SNR
 482 recording. **e**, Pearson correlation coefficients of single-pixel calcium traces before and
 483 after denoising (left). High-SNR traces were used as the reference for correlation
 484 calculation. Improvements were observed in all 40 traces (right). **f**, Low-SNR recording
 485 of somatic signals in cortical layer 2/3. **g**, DeepCAD enhanced recording. **h**,
 486 Synchronized high-SNR recording (10-fold imaging SNR). Orange arrows point to two

487 neurons. Magnified views of the boxed regions show calcium transients in a ~ 1.4 s time
488 window. Scale bar 100 μm . **i**, Neurons extracted from the original low-SNR recording
489 (N=150). **j**, Neurons extracted from the DeepCAD enhanced recording (N=229).
490 Manual annotations served as the ground truth. Correctly segmented regions (true
491 positive) are colored green. Missing (false negative) and extra regions (false positive)
492 are colored red and blue, respectively. **k**, Distribution of peak SNRs of extracted
493 calcium traces. CNMF was used for source extraction and peak SNR estimation.