

# 1                    **Removing independent noise in systems neuroscience data using** 2    **DeepInterpolation**

3                    Jérôme Lecoq<sup>1</sup>, Michael Oliver<sup>1</sup>, Joshua H. Siegle<sup>1</sup>, Natalia Orlova<sup>1</sup>, Christof Koch<sup>1</sup>

4                    <sup>1</sup>Allen Institute for Brain Science, Seattle, WA, USA.

5                    **Progress in nearly every scientific discipline is hindered by the presence of independent noise in**  
6                    **spatiotemporally structured datasets. Three widespread technologies for measuring neural activity—**  
7                    **calcium imaging, extracellular electrophysiology, and fMRI—all operate in domains in which shot**  
8                    **noise and/or thermal noise deteriorate the quality of measured physiological signals. Current**  
9                    **denoising approaches sacrifice spatial and/or temporal resolution to increase the Signal-to-Noise**  
10                    **Ratio of weak neuronal events, leading to missed opportunities for scientific discovery.**

11                    **Here, we introduce *DeepInterpolation*, a general-purpose denoising algorithm that trains a spatio-**  
12                    **temporal nonlinear interpolation model using only noisy samples from the original raw data.**  
13                    **Applying DeepInterpolation to *in vivo* two-photon Ca<sup>2+</sup> imaging yields up to 6 times more segmented**  
14                    **neuronal segments with a 15 fold increase in single pixel SNR, uncovering network dynamics at the**  
15                    **single-trial level. In extracellular electrophysiology recordings, DeepInterpolation recovered 25%**  
16                    **more high-quality spiking units compared to a standard data analysis pipeline. On fMRI datasets,**  
17                    **DeepInterpolation increased the SNR of individual voxels 1.6-fold. All these improvements were**  
18                    **attained without sacrificing spatial or temporal resolution.**

19                    **DeepInterpolation could well have a similar impact in other domains for which independent noise is**  
20                    **present in experimental data.**

21

## 22 **Introduction**

23 Independent noise is a major impediment to any data acquisition effort. Experimental systems neuroscience  
24 is acutely impacted given the weakness of neuronal signals. For example, *in vivo* imaging of fluorescent  
25 reporters of cellular activity (voltage and calcium) are typically operating in low-photon-count regimes  
26 where shot noise largely dominates the recorded signal. Similarly, thermal and shot noise present in  
27 electronic circuits independently drown out *in vivo* electrophysiological recordings causing lower  
28 sensitivity, impairing the detection of individual spikes. Functional Magnetic Resonance Imaging (fMRI)  
29 also suffers from thermal noise present in recording circuits, as relevant changes in Blood-Oxygen-Level  
30 Dependent (BOLD) signal are typically just a few percent. The presence of independent noise in all these  
31 techniques hampers our ability to measure authentic biological signals, causing irreproducible research or  
32 inducing biases related to particular choices of denoising filters<sup>1</sup> or statistical thresholds<sup>2</sup>.

33 Often, nearby spatio-temporal samples share signals but are independently affected by noise. This  
34 relationship allows the creation of filters to remove noise when applied to the data. Indeed, it is common  
35 practice to apply median or Gaussian filters as a first pre-processing step. Sometimes these averaging filters  
36 are designed in the Fourier domain to efficiently separate noise from the signal. Although this approach has  
37 been successful for many applications, optimal denoising filters can be complicated to build when these  
38 data relationships span multiple dimensions (e.g., time and space) or are intrinsically non-linear.  
39 Importantly, most of these approaches also impact the signal, causing reductions in spatial or temporal  
40 resolution.

41 The use of machine learning with large datasets permits very complex hierarchical relationships between  
42 data points to be learned<sup>3</sup>. In this framework, learned statistical relationships between samples are exploited  
43 to reconstruct a noiseless version of the signal, rather than applying filters to remove noise from the signal.  
44 This approach has been particularly successful for learning denoising filters or upsampling filters, including

45 for fluorescence imaging<sup>4,5</sup> but was limited until recently to cases where ground truth is readily available.  
46 For example, pre-acquired structural data at various spatial resolutions made it possible to train neural  
47 network models to upsample datasets and limit the amount of data necessary to reconstruct a full resolution  
48 image<sup>4</sup>. However, when dealing with the types of dynamic signals at the heart of systems neuroscience,  
49 denoised datasets or ground truth data are not readily available or may be impossible to obtain.

50 A recent approach called Noise2Noise<sup>6</sup> demonstrated that deep neural networks can be trained to perform  
51 image restoration without access to any clean ‘ground-truth’ data, with performance comparable or  
52 exceeding training using cleaned data. Theoretical work extended this framework to remove pixel-wise  
53 independent noise<sup>7,8</sup>. Here, we adapted these recent developments in machine learning to the problem of  
54 denoising dynamical signals at the heart of systems neuroscience, thereby building a general-purpose  
55 denoising algorithm we call *DeepInterpolation*. Here, we describe and demonstrate the use of  
56 *DeepInterpolation* for two-photon *in vivo* Ca<sup>2+</sup> imaging data, *in vivo* electrophysiological recordings, and  
57 whole brain fMRI data. Importantly, our approach is applicable to many experimental fields with minimal  
58 modifications.

## 59 **Results**

60 The Noise2Noise<sup>6</sup> study demonstrated that deep neural networks can be trained to perform image  
61 restoration without access to any clean ‘ground-truth’ data. At the heart of this approach is the insight that  
62 gradients derived from a loss function calculated on noisy samples are still partially aligned toward the  
63 correct solution. Provided enough noisy samples are available to train—and the noisy loss is an unbiased  
64 estimate of the true loss—the correct denoising filters are still learned<sup>6</sup>. However, unlike the data used in  
65 the Noise2Noise publication<sup>6</sup>, systems neuroscience data does not contain pairs of samples with identical  
66 signals but different noise.

67 To address this limitation, we followed a similar approach proposed recently in the Noise2Self<sup>7</sup> and  
68 Noise2Void<sup>8</sup> frameworks. In the absence of training pairs, we sought to solve an interpolation problem in  
69 order to learn the spatio-temporal relationships present in the data. That is, we do not seek to denoise an  
70 image but rather to learn the underlying relationship between an entire frame and spatio-temporally nearby  
71 samples by optimizing the reconstruction loss calculated on noisy instances of the center frame. Second,  
72 the noise present in the target (center) samples needs to be independent from the input (adjacent) samples;  
73 otherwise our relationship model would overfit the noise we seek to remove. We eliminated any chance of  
74 overfitting by (1) omitting the target center frame from the input, and (2) presenting training samples only  
75 once during training. While this noisy center frame is the most informative about its own value, it serves  
76 the key function of the target during training. During inference, the dataset is simply streamed through the  
77 trained network to reconstruct a noiseless version of the signal. The quality of the reconstruction depends  
78 on whether enough information about the center frame was present in spatio-temporally nearby samples  
79 and if it was appropriately learned by the training procedure. Given the extensive leverage of complex  
80 relationships present in the data, spanning multiple spatio-temporal dimensions, we will refer to this general  
81 framework as *DeepInterpolation*.

## 82 **Application of DeepInterpolation to *in vivo* two-photon Ca<sup>2+</sup> imaging**

83 For two-photon imaging, we constructed our denoising network following simple principles: First, given  
84 the presence of neuropil signals distributed throughout the background, a single pixel can share information  
85 with other pixels throughout the entire field of view. Second, the decay dynamics of calcium-dependent  
86 fluorescent indicators (GCaMP6f  $\tau_{\text{peak}} = 80 \pm 35$  ms;  $\tau_{1/2} = 400 \pm 41$  ms<sup>9</sup>) suggest frames as far as 1  
87 second away can carry meaningful information. Therefore, we chose to train a neuronal network to  
88 reconstruct a center frame from the data present in  $N_{\text{pre}}$  prior frames and  $N_{\text{post}}$  subsequent frames. The center  
89 frame was omitted from the input so as to not provide any information about the independent pixel-wise  
90 noise due to shot noise (**Fig. 1A**). The final value of the meta-parameter  $N_{\text{pre}}$  and  $N_{\text{post}}$  were selected by

91 comparing the validation loss during training (**Fig. 1B**, see **Methods** and **Appendix I** for details on loss  
92 calculation). Our denoising network utilized a UNet inspired encoder-decoder architecture with 2D  
93 convolutional layers<sup>1011</sup>.

94 The Allen Brain Observatory datasets (<http://observatory.brain-map.org/visualcoding>) offered a unique  
95 opportunity to train these denoising networks as we had collected more than 1,300 hours of highly  
96 standardized and curated datasets<sup>12</sup>. For each Gcamp6 reporter line (Ai93, Ai148), we had access to more  
97 than 100 million frames carefully curated with standard Quality Control (QC) metrics. We leveraged this  
98 database by presenting each training sample only once during the entire training process, completely  
99 eliminating any chance of noise overfitting.

100 Within the limits of our computational infrastructure, we showed that using a larger number of both prior  
101 and subsequent frames as input yielded smaller final validation reconstruction losses (see **Fig. 1B**). Training  
102 required 225,000 samples pulled randomly from 1144 separate one hour long experiments to stop improving  
103 (see **Fig. 1B**). Since the loss is dominated by the independent noise present in the target image, the  
104 reconstruction loss could not converge to zero. This was supported by simulation of ideal reconstruction  
105 losses with known ground truth (**Supp. Fig. 1**). In fact, even small improvement in the loss related to visible  
106 improvements in the reconstruction quality of the signal (**Fig. 1B**).

107 Applying this trained network to denoise held-out datasets (**Supp. Video 1, 2**) yielded remarkable  
108 improvements in imaging quality. The same trained network generalized well to many movie examples (see  
109 4 examples on **Supp. Video 2**). Shot noise was visibly eliminated from the reconstruction (see single frames  
110 in **Fig. 1C** and associated calcium trace) while calcium dynamics were preserved, yielding a 15 fold increase  
111 in SNR (mean raw pixel SNR = 2.4 +/- 0.01, mean pixel SNR after DeepInterpolation = 37.2 +/- 0.2,  
112 N=9966 pixels). While the movies we used during training were motion corrected, a small amount of

113 motion remained and close inspection of the reconstructions showed that the algorithm also corrected these  
114 remaining motion artifacts (**Supp. Video 1, 2**).

115 We compared our denoising approach with a recent algorithm called Penalized Matrix Decomposition  
116 (PMD)<sup>13</sup>. While PMD properly reconstructed somatic activities, it rejected most of the variance present in  
117 background pixels, contrary to DeepInterpolation (see **Supp. Video 3**). A key underlying assumption of  
118 PMD is that motion artifacts are absent from the data. Contrary to DeepInterpolation movies, we observed  
119 clear artifacts in movies denoised with PMD due to small residual motion artifacts (see arrow in **Supp.**  
120 **Video 3**).

121 To compare these two approaches given this limitation, we created entirely simulated datasets devoid of  
122 motion artifacts, for which we had knowledge of the underlying ground truth. For this, we used a recent  
123 approach called *in silico* Neural Anatomy and Optical Microscopy (NAOMi)<sup>14</sup>. NAOMi combines a  
124 detailed anatomical model of a volume of tissue with models of light propagation and laser scanning so as  
125 to generate realistic Ca<sup>2+</sup> imaging datasets (see **Methods**). Both DeepInterpolation and PMD performed  
126 well with datasets generated by NAOMi (see **Supp. Fig. 2 B,C** and **Supp. Video 4**). Because  
127 DeepInterpolation used information from nearby frames, it provided visibly smoother calcium traces with  
128 reconstruction losses 1.2 to 2 times smaller than with PMD (see **Supp. Fig. 2**).

129 Having established that DeepInterpolation compares favorably to the state of the art, we investigated how  
130 removing shot noise impacts our ability to analyze the activity of neuronal compartments.

### 131 **Segmentation of active compartments during Ca<sup>2+</sup> imaging is transformed by DeepInterpolation.**

132 We first sought to evaluate the impact of DeepInterpolation on the segmentation process itself. Existing  
133 approaches to calcium imaging segmentation rely on the analysis of correlation between pixels to create  
134 ROIs<sup>15,16</sup>. After denoising, single pixel pairwise correlations greatly increased from near zero (averaged  
135 Pearson correlation =  $9.0 \cdot 10^{-4} \pm 2.3 \cdot 10^{-6}$  s.e.m,  $n = 4 \cdot 10^8$  pairs of pixels across 4 experiments) to a

136 significant positive value (averaged Pearson correlation =  $0.10 \pm 1.210^{-5}$  s.e.m,  $n = 4e8$  pairs of pixels  
137 across 4 experiments, KS test comparing raw with DeepInterpolation:  $p = 9 \times 10^{-71}$ ,  $n = 1,000$  pixels  
138 randomly sub-selected, **Supp. Fig 3**). We expected that this increase in averaged correlation would improve  
139 the quality and the number of segmented regions. To that end, we leveraged the sparse mode algorithm  
140 available in Suite2p<sup>16</sup>.

141 The improvement in correlation structure benefited this algorithm considerably. Suite2p extracted a large  
142 number of additional active compartments in the field of view, including small sets of pixels covered by  
143 axons or dendrites perpendicular to the imaging plane (**Fig. 2A**). In the absence of shot noise and single  
144 frame motion artifacts, cell filters were surprisingly well defined (**Fig. 2B**). Automatically extracted ROIs  
145 showed precise details of each compartment, sometimes encompassing dendrites or axons connected to a  
146 local soma. In some cases, active attached buttons were clearly visible in the extracted filter of an horizontal  
147 dendrite (**Fig. 2B**). In many movies, a large number of smaller, punctated compartments - likely a mixture  
148 of horizontal sections of apical dendrites or axons - were detected, each with clear calcium transients (**Fig.**  
149 **2C**).

150 Although those numerous compartments were not detected in the original movie, we applied the weighted  
151 masks to both denoised and raw movies to quantify the improvement in SNR of the associated temporal  
152 trace. ROIs associated with somatic compartments typically yielded a mean SNR of  $21.0 \pm 0.5$  (s.e.m,  $n$   
153 = 240 in 3 experiment) in the original movie (**Fig. 2D**); after denoising, the majority of those ROIs had an  
154 SNR above 40 (mean  $73.8 \pm 0.5$ ). Similarly, the mean SNR across all compartments, including the  
155 smallest apical dendrites, went from a mean of  $13.2 \pm 0.1$  ( $n = 3385$  in 3 experiments) to  $73.8 \pm 0.5$ .  
156 Assuming that a single spike causes a calcium rise of about 10%<sup>9</sup>, an SNR above 20 is necessary for those  
157 events to be twice as large as the noise. Although this is a necessary but not sufficient condition, this  
158 suggests that removing shot noise will increase our capability to detect the presence of single spikes.

159 The average number of active detected ROIs with SNR above 20 went from 178 +/- 135 (s.e.m., in 3  
160 experiments) to 1122 +/- 371 in a 400x400  $\mu\text{m}^2$  field of view (**Fig. 2D**). That is, as much as six times more  
161 neuronal compartments are now available from a single movie for analysis following DeepInterpolation,  
162 thanks to a more reliable segmentation process.

### 163 **DeepInterpolation improves the analysis of correlated $\text{Ca}^{2+}$ imaging activity**

164 To illustrate this new opportunity, we analyzed the response of all compartments to 10 repeats of a natural  
165 movie stimulation. We found clear examples of large and smaller responses to natural movies in both  
166 somatic and non-somatic compartments (**Supp. Fig. 4A**). The trial by trial response reliability (see  
167 **Methods**) increased substantially from raw traces to traces after DeepInterpolation (see, **Fig. 2E**), both for  
168 somatics ROIs (raw: mean 0.10 +/- 0.01, DeepInterpolation: mean 0.20 +/- 0.01; s.e.m,  $n = 240$  ROIs in  $n$   
169 = 3 experiments) and across all ROIs (raw: mean 0.029 +/- 0.001, DeepInterpolation: mean 0.113 +/- 0.002;  
170 s.e.m,  $n = 3385$  ROIs in  $n = 3$  experiments). When considering the distribution of these responses across a  
171 single field of view, both somatic and non-somatic compartments had a similar distribution for their  
172 preferred movie frame, the maximum DF/F response at this preferred stimulus, as well as the reliability of  
173 their response to natural movies (see **Supp. Fig. 4B , C, D**).

174 Neuronal representations are inherently noisy even when sensory stimuli and the behavioral state are kept  
175 constant. This key property of neuronal networks has fueled a large number of studies on the relationship  
176 between the averaged signal representation in individual neurons and the trial by trial fluctuation of these  
177 responses<sup>17,18</sup>. Experimental noise directly impacts the measure of neuronal relationships as it modulates  
178 trial-by-trial responses.

179 For each pair of ROIs in a single experiment, we extracted their shared signal and noise correlation (see  
180 **Methods**). As expected, we found that both signal as well as noise correlation increased after denoising  
181 (signal: from  $0.06725 \pm 0.00006$  to  $0.25 \pm 0.0001$ ; noise from  $0.02240 \pm 0.00002$  to  $0.12351 \pm 0.00008$ ;



182 both for 2,329,554 pairs of ROIs from three experiments; **Fig. 2F**). This result was not just due to the  
183 improved segmentation as it was preserved using ROIs only detected on raw noisy data (see **Supp. Fig. 5**).  
184 To further illustrate how these improved measures of neuronal fluctuations impacted the analysis of  
185 functional interactions, we compared the spatial distribution of pairwise correlations between neurons  
186 during natural movie stimulation. DeepInterpolation largely increased the number of strong pairwise  
187 interactions for both somatic and non somatic pairs (see **Fig. 2G**, all ROIs: 43 to 1721 pairs). These  
188 correlated pairs of units were almost invisible in the original data, illustrating the improved biological  
189 insight gained through denoising.

## 190 **Application of DeepInterpolation to *in vivo* electrophysiological recordings**

191 Electrophysiological recordings from high-density silicon probes have similar characteristics to two-photon  
192 imaging movies: information is shared across nearby pixels (electrodes), as well as across time. We  
193 therefore reasoned that a similar architecture should perform well for denoising electrophysiology data, in  
194 particular for removing strictly independent thermal and shot noise present in the recordings. For training  
195 data, we used Neuropixels data that were similarly standardized and curated as the two-photon recordings<sup>19</sup>.  
196 Because a single action potential lasts ~1 ms, we constructed our interpolating filter to predict the voltage  
197 profile at any given moment from 30 preceding and 30 following time steps acquired at 30 kHz. Contrary  
198 to two-photon imaging, we found that omitting one sample before and after the center sample better rejected  
199 the background noise while keeping the signal reconstruction of the highest quality. The architecture of the  
200 network was similar to the one used for two-photon imaging, except that the input layer was reshaped to  
201 match the approximate layout of the Neuropixels recording sites (**Fig. 3A**). Because electrophysiological  
202 recordings have a 1,000-fold higher sampling rate than two-photon imaging, we had access to essentially 3  
203 orders of magnitude more training data per experiment and could train the network using only 3 experiments  
204 in total.

205 After inference, a qualitative inspection of the recordings showcased excellent noise rejection (**Fig. 3B**).  
206 We compared spike profiles before and after denoising and found that the spike waveforms were more  
207 clearly visible in the denoised data, especially for units with lower amplitude spikes (**Fig. 3B**, inset; **Fig.**  
208 **3C**). The residual of our reconstruction showed weak spike structure but spectral analysis revealed a mostly  
209 flat frequency distribution (**Supp. Fig. 6A**). Importantly, the overall shape (amplitude and decay time) of  
210 individual action potentials was preserved, while the median channel RMS noise was decreased 1.7-fold  
211 ( $6.96 \mu\text{V}$  after denoising, see **Fig. 3D**). We ran a state-of-the-art spike sorting algorithm (Kilosort2, version  
212 downloaded April 8, 2019) and removed any detected units with waveforms not likely to be associated with  
213 action potentials. After denoising, we found  $25.5 \pm 14.5\%$  more high-quality units per probe (ISI violations  
214 score  $< 0.5$ , amplitude cutoff  $< 0.1$ , presence ratio  $> 0.95$ ; see Methods for details). The number of detected  
215 units was higher after DeepInterpolation regardless of the quality metric thresholds that were chosen (**Supp.**  
216 **Fig. 6B**). The majority of additional units had low-amplitude waveforms, with the number of units with  
217 waveform amplitudes above  $75 \mu\text{V}$  remaining roughly the same (1532 before vs. 1511 after), while the  
218 number with amplitudes below  $75 \mu\text{V}$  increased dramatically (381 before vs. 910 after) (**Fig. 3E**). For units  
219 that were clearly matched before and after denoising (at least 70% overlapping spikes), average SNR  
220 (defined as the ratio of the waveform amplitude to the RMS noise on the peak channel) increased from 11.2  
221 to 14.5 (**Fig. 3F**).

222 By searching for spatiotemporally overlapping spikes, we determined that  $6.2 \pm 2.9\%$  of units in the original  
223 recording were no longer detected after applying DeepInterpolation ( $< 50\%$  matching spikes). However,  
224 this was counterbalanced by the addition of  $20.2 \pm 4.8\%$  units that had fewer than 50% matching spikes in  
225 the original data (example unit in **Supp. Fig. 6C**). This is less than the 25.5% increase cited above, because  
226 it accounts for a minority of units that were detected but merged together in the original data. To validate  
227 that these novel units were likely to correspond to actual neuronal compartments, we analyzed their  
228 responses to natural movie stimulation, using the same movie as in the two-photon imaging study. We

229 found approximately 9% more reliably visually modulated units in the visual cortex after denoising (**Supp.**  
230 **Fig. 6D**), and found clear examples of stimulus-modulated units that were previously undetected (**Supp.**  
231 **Fig. 6E**).

232 Applying DeepInterpolation to extracellular electrophysiology data increased the unit yield across all brain  
233 regions tested, including visual cortex, hippocampus, and thalamus (**Supp. Fig. 7A**). Additional units  
234 tended to have firing rates in the middle of each region's average range and waveform amplitudes below  
235 100  $\mu$ V (**Supp. Fig. 7B**). In all regions except for dentate gyrus, DeepInterpolation preserved the original  
236 waveform amplitude, while detecting more low-amplitude units. The downward shift in waveform  
237 amplitude observed in dentate gyrus suggests that it may be necessary to provide additional training data  
238 from this region in order to optimize performance of the denoising step. In the visual cortex, additional  
239 units were primarily "regular-spiking" (waveform peak-to-trough duration > 0.4 ms), rather than "fast-  
240 spiking." This suggests that these may be previously undetectable regular-spiking interneurons with small  
241 cell bodies; however, applying DeepInterpolation to datasets with ground truth cell type labels obtained via  
242 optotagging<sup>20</sup> will be required to confirm this hypothesis.

### 243 **Application of DeepInterpolation to functional Magnetic Resonance Imaging (fMRI)**

244 We sought to evaluate how DeepInterpolation could help the analysis of volumetric datasets like fMRI.  
245 fMRI is very noisy as the BOLD response is typically just a 1-2 % change of the total signal amplitude<sup>21</sup>.  
246 Thermal noise is present in the electrical circuit used for receiving the MR signal. A typical fMRI processing  
247 pipeline involves averaging nearby pixels and successive trials to increase the SNR<sup>22,23</sup>. We reasoned that  
248 our approach could replace smoothing kernels with more optimal local interpolation functions without  
249 sacrificing spatial or temporal resolution.

250 Because of the low sampling rate (0.3 to 0.5 Hz), a single fMRI recording session could only provide several  
251 hundreds full brain volumes, while each volume typically contains up to 10 millions voxels. To augment

252 our training datasets, we decided to alter our approach: Instead of learning a full brain interpolation model,  
253 we sought to train a more local interpolation function.

254 To reconstruct the value of a brain sub-volume, we fed a neural network with a consecutive temporal series  
255 of a local volume of  $7 \times 7 \times 7$  voxels. The entire target volume was omitted from the input (**Fig. 4A**). To  
256 allow our interpolation network to be robust to edge conditions, input volumes on the edge of the volume  
257 were fed with zeros for missing values. For inference, we convolved the denoising network through all  
258 voxels of the volume, across both space and time, using only the center pixel of the output reconstructed  
259 volume to avoid any potential volume boundaries artifacts.

260 We trained a single denoising network using 140 fMRI datasets acquired across 5 subjects or 1.2 billion  
261 samples. Each training sample was presented only once to avoid any chance of overfitting. After training,  
262 we applied the network to held-out test data, showing excellent denoising performance (**Fig. 4C**).

263 We extracted the temporal traces from individual voxels and found that tSNR (see **Methods**) increased  
264 from  $61.6 \pm 20.6$  to  $100.40 \pm 38.7$  ( $n = 100,001$  voxels). We noticed that the background noise around the  
265 head was clearly excluded and only present in the residual image (**Fig. 4C and Supp. Video 5**). Surrounding  
266 soft tissues became clearly visible after denoising (see **Fig. 4C**, Raw data: background voxel std =  $7.59 \pm$   
267  $0.01$  s.e.m; brain voxels std =  $15.95 \pm 0.08$  s.e.m.,  $n = 10,000$  voxels; DeepInterpolation data: background  
268 std =  $2.24 \pm 0.01$  s.e.m.,  $n = 10,000$ ; brain voxels std =  $9.72 \pm 0.05$  s.e.m.,  $n = 10,000$  voxels; Residual:  
269 background std =  $7.91 \pm 0.01$ ; brain voxels std =  $15.79 \pm 0.08$ ,  $n = 10,000$  voxels). The residual movie  
270 showed no visible structure except for occasional blood vessels, corrected motion artifacts and head  
271 mounting hardware (**Fig. 4C, Supp. Video 7**). We compared movies extracted with DeepInterpolation to  
272 fully pre-processed movies that included a Gaussian denoising kernel (**Supp. Video 6**): the original 3D  
273 resolution was fully maintained in DeepInterpolation while the processed data with Gaussian kernel clearly  
274 diffused the voxel dynamic across all nearby voxels.

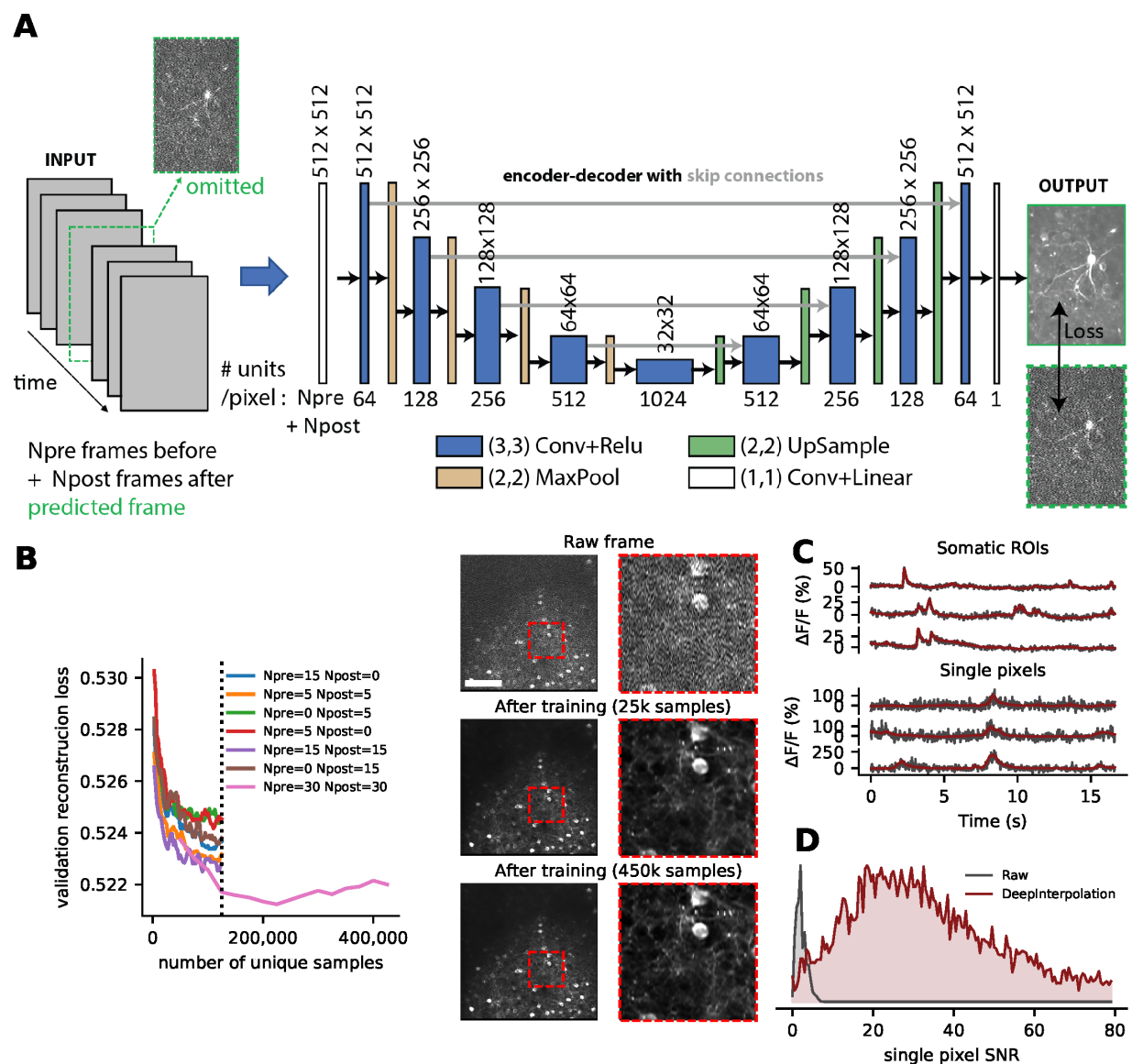
## 275 **Discussion**

276 Here, we demonstrated a methodology, *DeepInterpolation*, related to the approach outlined in the  
277 *Noise2Self*<sup>7</sup> and *Noise2Void*<sup>8</sup> framework to reconstruct noiseless versions of dynamic datasets, without any  
278 requirement for ground truth. While we share the key separation of noise and signal with these techniques,  
279 we developed our framework independently. As a result, several key differences are notable. First, instead  
280 of working with single fluorescence frames with pixel-wise omissions, we adapted our approach to complex  
281 multi-dimensional biological datasets at the heart of systems neuroscience. Second, we trained our models  
282 on large databases, demonstrating the impact of richer denoising models on existing neuroscience scientific  
283 data and data workflows like cell segmentations or unit detection. Third, we contributed practical solutions:  
284 network topology, training regimen and a validation process to extract complex multi-dimensional models.  
285 When applied to two-photon Ca<sup>2+</sup> imaging datasets, *DeepInterpolation* increased the SNR of single pixels  
286 15-fold. On electrophysiological data, with little architectural change, *DeepInterpolation* uncovered 25%  
287 more high quality neuronal units. Similarly, applying *DeepInterpolation* led to a 1.6-fold increase in the  
288 SNR of fMRI voxels, effectively removing thermal noise from the recordings. The successful application  
289 of *DeepInterpolation* to these three experimental data modalities supports the general relevance of our  
290 method.

291 Given the increased leverage of the information present between samples, we anticipate *DeepInterpolation*  
292 to be instrumental for the development of large scale voltage-imaging of neuronal activity<sup>24,25</sup> as well as  
293 facilitate high-speed fMRI where thermal noise dominates the BOLD signal. Importantly, the pixel-wise  
294 framework we used for fMRI increased our training dataset, offering the opportunity to apply the same  
295 methodology to many different modalities, regardless of whether or not vast quantities of training data are  
296 available. Our approach should permit scientists in a variety of fields to re-analyze their existing datasets  
297 but now with independent noise removed.

298 Speeding up and/or parallelizing the inference step could facilitate real-time denoising of spatiotemporal  
299 signals. This could enable a process whereby a short dataset would be acquired at the onset of any one  
300 experiment, similar to a calibration routine. This dataset could then be used to fine-tune a pre-trained  
301 denoiser that removes independent noise. Whether such training needs to be specific for each instrument,  
302 each type of instrument, each brain area, each transgenic mouse line or even each subject remains a topic  
303 for future research.

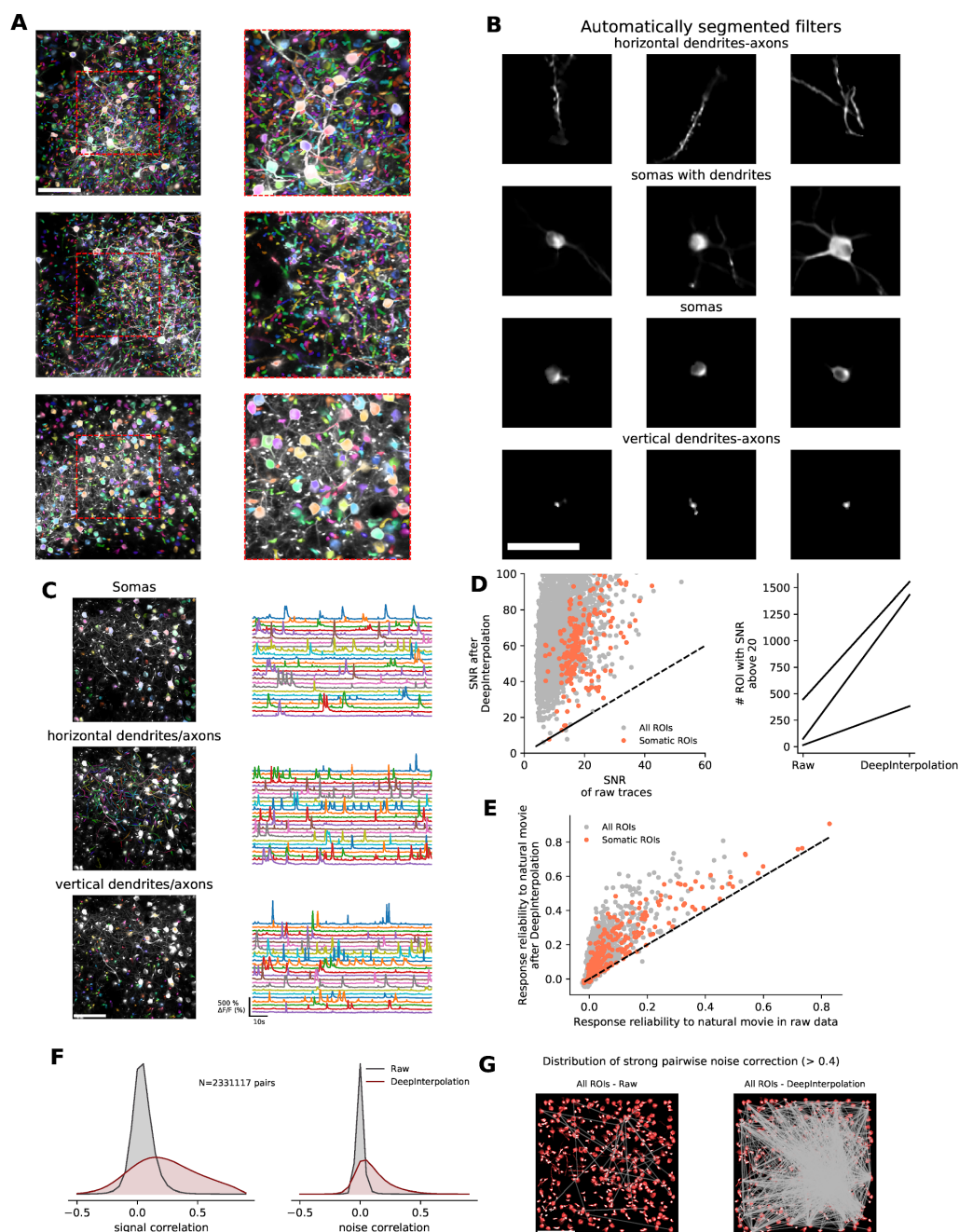
304 We here demonstrated a general purpose denoising methodology to reconstruct dynamic signals at the heart  
305 of systems neuroscience without the contamination of independent noise. Applying DeepInterpolation to  
306 Ca<sup>2+</sup> imaging, electrophysiological recordings and fMRI showed its wide impacts on SNR and its ability to  
307 uncover previous hidden neuronal dynamics. We anticipate that future work in all 3 modalities we presented  
308 here will revise key subsequent processing steps to better leverage clearer correlation structures now  
309 measured in the data.



310  
 311 **Fig. 1 | Training DeepInterpolation networks for denoising two-photon  $\text{Ca}^{2+}$  imaging.** (A) An encoder-  
 312 decoder deep network with skip connections was trained to denoise two-photon imaging data. The training  
 313 procedure minimized the L1 loss between the predicted frame and a raw frame contaminated with shot  
 314 noise. The encoder-decoder network utilized Npre and Npost frames, acquired at 30 Hz, respectively, before  
 315 and after the target frame (labelled in green) to predict the central, target frame. (which was omitted from  
 316 the input). (B) To validate training performance, the validation loss was monitored for convergence as well  
 317 as the quality of the image reconstruction throughout training (see inset). Y axis is the mean absolute

318 difference between a predicted frame and a noisy sample across 2500 samples. Individual data samples  
319 were z-scored using a single estimate of mean and standard deviation per movie (see Methods). Different  
320 combinations of (Npre, Npost) values were tested during training. Dashed vertical line indicates early  
321 stopping (due to the extreme computational demands during training) to evaluate the best set of parameters.  
322 We continued training the model with Npre=30 frames and Npost=30 frames and used this set of parameters  
323 for the rest of this study. Right panels illustrate the denoising performance of this model compared to a  
324 single raw frame (top). Notice the improved contrast between 25,000 and 450,000 samples training: the  
325 smallest neuronal compartments are more clearly visible in the latter. Scale bar is 100  $\mu\text{m}$ . Each red inset  
326 is 100  $\mu\text{m}$ . (C) Six example traces extracted from a held-out denoised movie before (black) and after (red)  
327 denoising with DeepInterpolation. The top three traces are extracted from a somatic ROI, while the bottom  
328 three traces are extracted from a single pixel. (D) Distribution of SNR (mean over standard deviation, see  
329 Methods) for 10,000 voxels (randomly selected across N=19 denoised held-out test movies) before and  
330 after DeepInterpolation.  
331





332

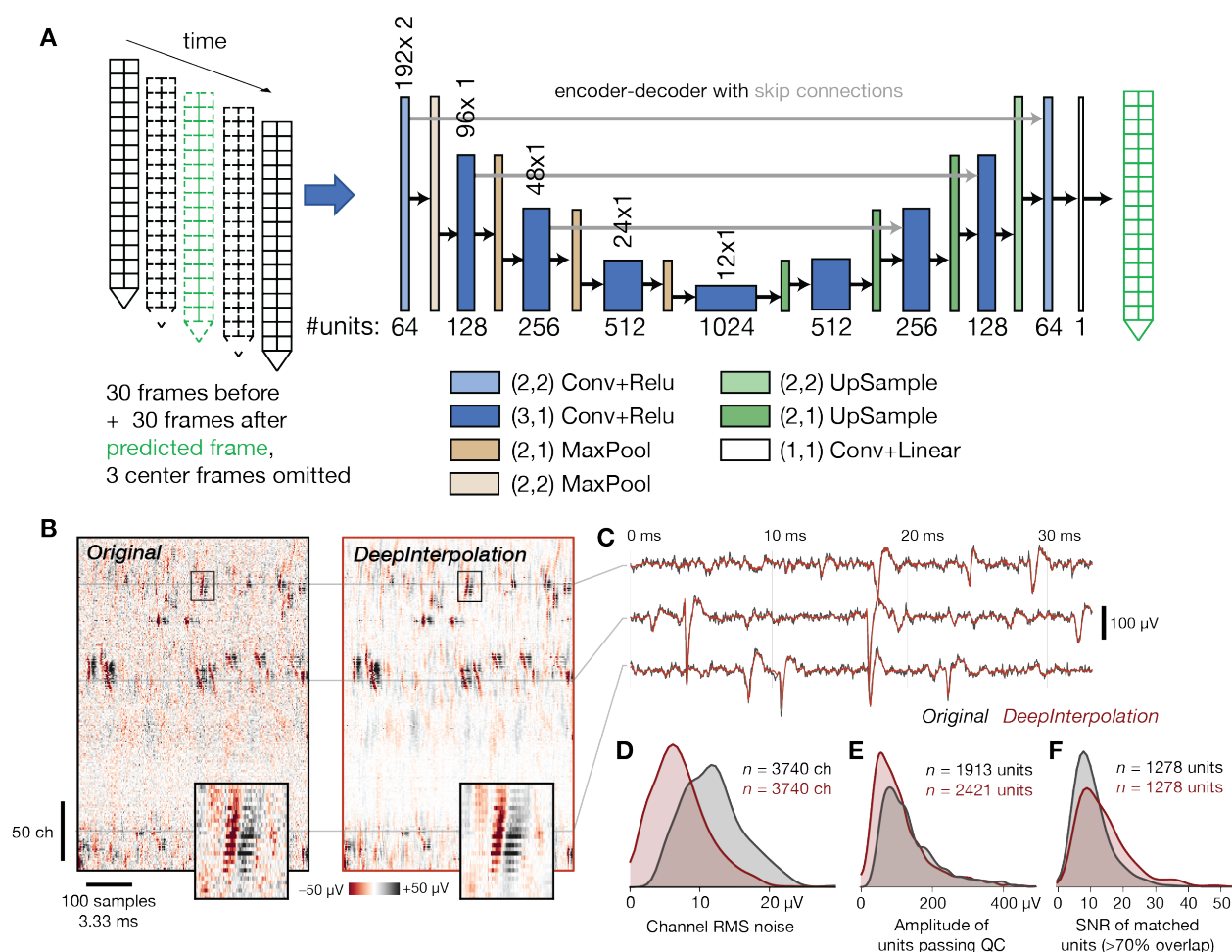
333 **Fig. 2 | Applying DeepInterpolation to Ca<sup>2+</sup> imaging reveals many more ROIs and rich network**

334 **physiology across small and large neuronal compartments. (A)** Three examples showing overlaid

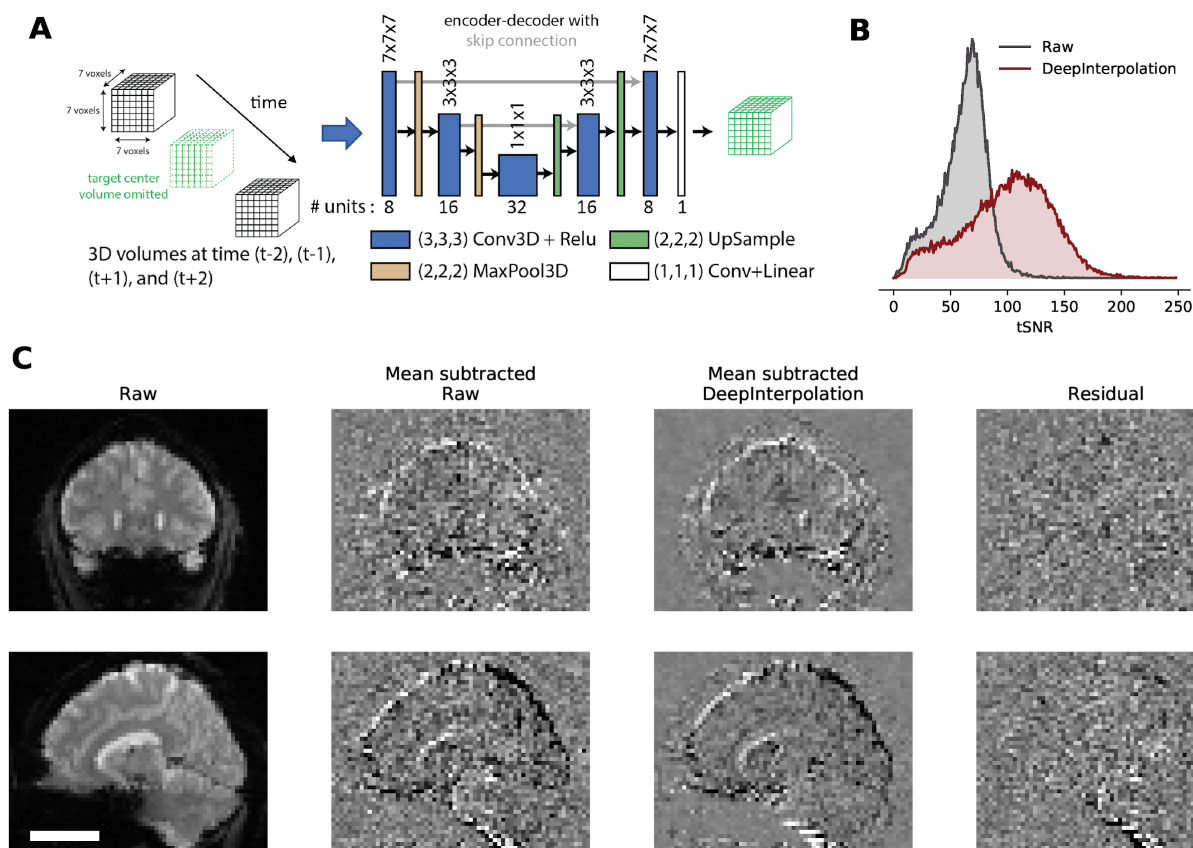
335 colored segmentation masks on top of a maximum projection image of calcium imaging. Improved

336 correlation structure (see **Supp. Fig 4**) allows the detection of a large number of ROIs using the “sparsity”

337 segmentation mode in Suite2p<sup>16</sup>. Red dashed boxes show zoomed-in view over smaller segmented  
338 compartments. Apical and proximal dendrites are cleanly segmented, regardless of whether they were  
339 imaged along their length or through a short vertical section. Scale bar is 100  $\mu\text{m}$ . **(B)** Example individual  
340 weighted segmented filters showcasing dendrites, isolated somas, somas with attached dendrites, and axons  
341 and small sections of dendrites/axons. Scale bar is 50  $\mu\text{m}$  **(C)** Manually sorted ROIs from one experiment  
342 showcasing the calcium traces extracted from each individual type of neuronal compartments (from top to  
343 bottom - cell body, horizontal processes, processes perpendicular to imaging plane). After  
344 DeepInterpolation, calcium events are recorded with high SNR in all three compartment types, regardless  
345 of their size. **(D, left)** Quantification of SNR for all detected ROIs (gray dots) and somatic ROIs (red dots)  
346 with and without DeepInterpolation. The majority of ROIs have an SNR above 20 after DeepInterpolation.  
347 Dashed line represents the identity line. **(D, right)** Comparison of the number of ROIs with an SNR above  
348 20 with and without DeepInterpolation. DeepInterpolation allows to reliably record the activity of more  
349 than a thousand compartments in a small  $(400 \mu\text{m})^2$  field of view ( $n = 3$  experiments). **(E)** Quantification  
350 of the response reliability of individual ROIs across 10 trials of a natural movie visual stimulus. Both  
351 somatic and non-somatic ROIs see increased response reliability to a visual stimulus. Dashed line represents  
352 the identity line. **(F, left)** signal correlation (average correlation coefficient between the average temporal  
353 response of a pair of neurons) for all pairs of ROI in **(D)** for both raw and denoised traces. **(F, right)** noise  
354 correlation (average correlation coefficient at all time points of the mean-subtracted temporal response of  
355 a pair of neurons) for all pairs of ROIs in **(D)**. **(G)** For an example experiment, pairs of ROIs with high  
356 noise correlation ( $>0.4$ ) are connected with a straight line for both original two-photon data and after  
357 DeepInterpolation. Scale bar is 100  $\mu\text{m}$ .  
358



359  
 360 **Fig. 3 | Applying DeepInterpolation to electrophysiological recordings decreases background noise**  
 361 **and yields more neuronal units.** (A) Structure of the DeepInterpolation encoder-decoder network for  
 362 electrophysiological data recorded from a Neuropixels silicon probe, with two rows of 192 electrodes each,  
 363 spaced 20  $\mu$ m apart. Data is acquired at 30 kHz. (B) Side-by-side comparison of original (left) and denoised  
 364 (right) data, plotted as a 2D heatmap with time on the horizontal axis and channels along the vertical axis.  
 365 Insets show a close-up of a single action potential across 22 contiguous channels (the box is 3.6 ms wide).  
 366 (C) Denoised time series (red) overlaid on the original time series (gray) for three channels from panel B.  
 367 (D) Histogram of RMS noise for all channels from 10 experiments, before and after applying  
 368 DeepInterpolation. (E) Histogram of waveform amplitudes for all high-quality units from 10 experiments.  
 369 (F) Histogram of waveform SNRs for all units that were matched before and after denoising.



370  
 371 **Fig. 4 | Applying DeepInterpolation to fMRI removes thermal noise.** (A) Structure of the  
 372 DeepInterpolation encoder-decoder network for fMRI data. Instead of predicting a whole brain volume at  
 373 once, the network reconstructs a local  $7 \times 7 \times 7$  cube of voxels. (B) tSNR (see Methods) for 10,000 voxels  
 374 randomly distributed in the brain volume in raw data and after DeepInterpolation. (C) Exemplar  
 375 reconstruction of a single fMRI volume. First row is a coronal section while the second row is a sagittal  
 376 section through a human brain. Scale bar is 5 cm. In the second column, the temporal mean of 3D scan was  
 377 removed to better illustrate the presence of thermal noise in the raw data. The local denoising network was  
 378 processed throughout the whole 3D scan for denoising. The impact of DeepInterpolation on thermal noise  
 379 is illustrated in the 3rd column. The 4th column shows the residual of the denoising process. Notice how  
 380 the residual captured independent noise without any signal structure. Occasional large blood vessels are  
 381 visible in the residual (see bottom).

## 382 **Bibliography**

- 383 1. Botvinik-Nezer, R. *et al*. Variability in the analysis of a single neuroimaging dataset by many teams.  
384 *Nature* **582**, 84–88 (2020).
- 385 2. Mesa, N., Waters, J. & de Vries, S. E. J. The effect of inclusion criteria on the functional properties  
386 reported in mouse visual cortex. doi:10.1101/2020.05.12.091595.
- 387 3. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks.  
388 *Science* **313**, 504–507 (2006).
- 389 4. Weigert, M. *et al*. Content-aware image restoration: pushing the limits of fluorescence microscopy.  
390 *Nature Methods* **15**, 1090–1097 (2018).
- 391 5. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for  
392 fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).
- 393 6. Lehtinen, J. *et al*. Noise2Noise: Learning Image Restoration without Clean Data. arXiv preprint  
394 arXiv:1803.04189 (2018).
- 395 7. Batson, J. & Royer, L. Noise2Self: Blind Denoising by Self-Supervision. arXiv preprint  
396 arXiv:1901.11365. (2019).
- 397 8. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - Learning Denoising from Single Noisy Images.  
398 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2129-2137  
399 (2018).
- 400 9. Chen, T.-W. *et al*. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**,  
401 295–300 (2013).
- 402 10. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder  
403 Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495  
404 (2017).
- 405 11. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely Connected Convolutional

- 406 Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017)  
407 doi:10.1109/cvpr.2017.243.
- 408 12. de Vries, S. E. J. *et al*. A large-scale standardized physiological survey reveals functional  
409 organization of the mouse visual cortex. *Nat. Neurosci.* **23**, 138–151 (2020).
- 410 13. Buchanan, E. K. *et al*. Penalized matrix decomposition for denoising, compression, and improved  
411 demixing of functional imaging data. doi:10.1101/334706.
- 412 14. Charles, A. S., Song, A., Gauthier, J. L., Pillow, J. W. & Tank, D. W. Neural Anatomy and Optical  
413 Microscopy (NAOMi) Simulation for evaluating calcium imaging methods. doi:10.1101/726174.
- 414 15. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from  
415 large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).
- 416 16. Pachitariu, M. *et al*. Suite2p: beyond 10,000 neurons with standard two-photon microscopy.  
417 doi:10.1101/061507.
- 418 17. Rummyantsev, O. I. *et al*. Fundamental bounds on the fidelity of sensory cortical coding. *Nature* **580**,  
419 100–105 (2020).
- 420 18. Cohen, M. R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–  
421 819 (2011).
- 422 19. Siegle, J. H. *et al*. A survey of spiking activity reveals a functional hierarchy of mouse  
423 corticothalamic visual areas. *Neuroscience* 889 (2019).
- 424 20. Lima, S. Q., Hromádka, T., Znamenskiy, P. & Zador, A. M. PINP: A New Method of Tagging  
425 Neuronal Populations for Identification during In Vivo Electrophysiological Recording. *PLoS One* **4**,  
426 e6099 (2009).
- 427 21. Caballero-Gaudes, C. & Reynolds, R. C. Methods for cleaning the BOLD fMRI signal. *NeuroImage*  
428 vol. 154 128–149 (2017).
- 429 22. Park, B.-Y., Byeon, K. & Park, H. FuNP (Fusion of Neuroimaging Preprocessing) Pipelines: A Fully  
430 Automated Preprocessing Software for Functional Magnetic Resonance Imaging. *Front.*



- 431 *Neuroinform.* **13**, 5 (2019).
- 432 23. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**,
- 433 111–116 (2019).
- 434 24. Zhang, T. *et al.* Kilohertz two-photon brain imaging in awake mice. *Nat. Methods* **16**, 1119–1122
- 435 (2019).
- 436 25. Wu, J. *et al.* Kilohertz two-photon fluorescence microscopy imaging of neural activity in vivo. *Nat.*
- 437 *Methods* **17**, 287–290 (2020).
- 438 26. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with
- 439 high specificity and performance. *Neuron* **85**, 942–958 (2015).
- 440 27. Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-
- 441 Cell-Type Targeting and Functionality. *Cell* **174**, 465–480.e22 (2018).
- 442 28. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature*
- 443 **551**, 232–236 (2017).
- 444 29. Horikawa, T. & Kamitani, Y. Generic decoding of seen and imagined objects using hierarchical
- 445 visual features. *Nat. Commun.* **8**, 15037 (2017).
- 446 30. Siegle, J. H. *et al.* Open Ephys: an open-source, plugin-based platform for multichannel
- 447 electrophysiology. *J. Neural Eng.* **14**, 045003 (2017).
- 448 31. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & D., H. K. Kilosort: realtime spike-sorting
- 449 for extracellular electrophysiology with hundreds of channels. doi:10.1101/061481.
- 450 32. Hill, D. N., Mehta, S. B. & Kleinfeld, D. Quality metrics to accompany spike sorting of extracellular
- 451 signals. *J. Neurosci.* **31**, 8699–8705 (2011).
- 452 33. Magland, J. *et al.* SpikeForest, reproducible web-facing ground-truth validation of automated neural
- 453 spike sorters. *Elife* **9**, (2020).

454

## 455 **Acknowledgement**

456 We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support. We thank  
457 Alex Song from the Tank Laboratory for help with NAOMi and for providing early access to the codebase  
458 that he created with Adam Charles. We thank Liam Paninski, Saveliy Yusufov, and Ian August Kinsella  
459 for sharing code and support for applying PMD onto comparative datasets. We thank Daniel Kapner for  
460 help with two-photon segmentation as well as for supportive discussions. We thank James Philip Thomson  
461 from the Schnitzer laboratory for discussions related to denoising two-photon movies. We thank Anton  
462 Arkhipov and Nathan Gouwens for fruitful discussions and for supporting early use of DeepInterpolation  
463 toward analyzing two-photon datasets. We thank Mark Schnitzer and Biafra Ahanonu for continuous  
464 support and evaluating early application of DeepInterpolation to one-photon datasets. We thank Todd  
465 Peterson, Myra Imanaka, Steve Lawrenz, Sarah Naylor, Gabe Ocker, Michael Buice, Jack Waters, Peter  
466 Ledochowitsch and Hongkui Zeng for encouragement and support. We thank Doug Ollerenshaw and  
467 Nathan Gouwens for sharing code to standardize figure generation.

## 468 **Funding**

469 Funding for this project was provided by the Allen Institute, as well as by the National Institute Of  
470 Neurological Disorders And Stroke of the National Institutes of Health under Award Number  
471 UF1NS107610. The content is solely the responsibility of the authors and does not necessarily represent  
472 the official views of the National Institutes of Health.

## 473 **Author Contributions**

474 J.L. initiated, conceptualized and directed the project. J.L. and M.O. conceptualized the DeepInterpolation  
475 framework. J.L. wrote all training and inference code with guidance from M.O. J.L. performed all two-  
476 photon and fMRI analysis and ran DeepInterpolation on electrophysiological data. J.H.S and J.L. iterated  
477 the electrophysiological denoising process. J.H.S. performed all analysis on electrophysiological data. C.K.



478 guided the characterization of all denoising performance. N.O. ran the two-photon ground truth simulation.

479 J.L., M.O., J.H.S and C.K wrote the paper.

480

481 **Correspondence and requests** for materials should be addressed to J.L. ([jeromel@alleninstitute.org](mailto:jeromel@alleninstitute.org)).

482 **Competing interests**

483 The Allen Institute has applied for a patent related to the content of this manuscript.

484

485 **Supplementary Video Table**

486

Supp. Video 1	<b>Link:</b> <a href="https://www.dropbox.com/s/xrqn0x0o1z46c7d/Supplementary_video_1.mp4?dl=1">https://www.dropbox.com/s/xrqn0x0o1z46c7d/Supplementary_video_1.mp4?dl=1</a> <b>Caption:</b> Left: Example two-photon movie before and after DeepInterpolation. Right: The average movie was subtracted to highlight uncovered fluorescence signals.
Supp. Video 2	<b>Link:</b> <a href="https://www.dropbox.com/s/58rh63zhepjf7cb/Supplementary_video_2.mp4?dl=1">https://www.dropbox.com/s/58rh63zhepjf7cb/Supplementary_video_2.mp4?dl=1</a> <b>Caption:</b> 4 examples of two-photon movies before and after DeepInterpolation. The same DeepInterpolation model was used for all 4 examples.
Supp. Video 3	<b>Link:</b> <a href="https://www.dropbox.com/s/tx6me5gbf6trv23/Supplementary_video_3.mp4?dl=1">https://www.dropbox.com/s/tx6me5gbf6trv23/Supplementary_video_3.mp4?dl=1</a> <b>Caption:</b> Example two-photon movie processed with Penalized-Matrix Decomposition (PMD) and DeepInterpolation. White arrow highlights the impact of motion artifacts on PMD.
Supp. Video 4	<b>Link:</b> <a href="https://www.dropbox.com/s/r0lh9si02dg7s9f/Supplementary_video_4.mp4?dl=1">https://www.dropbox.com/s/r0lh9si02dg7s9f/Supplementary_video_4.mp4?dl=1</a> <b>Caption:</b> Simulated two-photon data without motion artifacts with NAOMi. 1st column is ground truth data, in the absence of shot noise. The 2nd column is the same movie with shot noise. The 3rd column is the output of PMD after processing the movie in the 2nd column. 4th column is the output of DeepInterpolation after processing the movie in the 2nd column.
Supp. Video 5	<b>Link:</b> <a href="https://www.dropbox.com/s/o4v8tisrcy2ho7p/Supplementary_video_5.mp4?dl=1">https://www.dropbox.com/s/o4v8tisrcy2ho7p/Supplementary_video_5.mp4?dl=1</a> <b>Caption:</b> Example fMRI data with and without DeepInterpolation. The 3 columns on the right are horizontal, sagittal and coronal sections of the same movie in the first column. The mean was subtracted to better highlight small changes in signal.
Supp. Video 6	<b>Link:</b> <a href="https://www.dropbox.com/s/m4uqr7vgfga8kkg/Supplementary_video_6.mp4?dl=1">https://www.dropbox.com/s/m4uqr7vgfga8kkg/Supplementary_video_6.mp4?dl=1</a> <b>Caption:</b> Example fMRI data. Left column is the original raw data. Middle column is the same data through the original processing pipeline, containing a gaussian kernel. The right column is the same data through DeepInterpolation.
Supp. Video 7	<b>Link:</b> <a href="https://www.dropbox.com/s/7cttm3107xlpp1/Supplementary_video_7.mp4?dl=1">https://www.dropbox.com/s/7cttm3107xlpp1/Supplementary_video_7.mp4?dl=1</a> <b>Caption:</b> Example fMRI data with and without DeepInterpolation. The residual was calculated to illustrate the type of signal that was removed by DeepInterpolation.

487

488

## 489 **Methods**

### 490 **Description of experimental data collection of datasets used in this study**

491 We trained 4 denoising neuronal networks in this study: one for two-photon imaging experiments using  
492 the Ai93(TITL-GCaMP6f) reporter line, one for the Ai148(TIT2L-GCaMP6f-ICL-tTA2) reporter line<sup>26,27</sup>,  
493 one for Neuropixels recordings using “Phase 3a” probes<sup>28</sup>, and one for fMRI imaging using datasets from  
494 a study published on OpenNeuro.org<sup>29</sup>. All raw datasets used for training are available at the time of  
495 publication through AWS S3 buckets (Amazon.com, Inc., see  
496 <https://openneuro.org/datasets/ds001246/versions/1.2.1> and [https://registry.opendata.aws/allen-brain-](https://registry.opendata.aws/allen-brain-observatory/)  
497 [observatory/](https://registry.opendata.aws/allen-brain-observatory/)). Importantly, each denoising network is tightly coupled to the inherent properties of the  
498 datasets they were trained on (acquisition parameters, sample characteristics,...).

499 Calcium imaging in mice was performed with two different two-photon imaging instruments<sup>12</sup> (either a  
500 Scientifica Vivoscope or a Nikon A1R MP+). Laser excitation was provided by a Ti:Sapphire laser  
501 (Chameleon Vision–Coherent) at 910 nm. 512x512 Movies were recorded at 30 Hz with resonant scanners  
502 over a 400- $\mu$ m field of view .

503 All extracellular electrophysiology recordings were carried out with Neuropixels probes<sup>28</sup> acutely inserted  
504 into the brains of awake mice. Data at each recording site was acquired at 30 kHz (spike band) and 2.5 kHz  
505 (LFP band). The spike band (which was used for denoising) had a hardware gain settings of 500x and a 500  
506 Hz high pass filter. Each probe contained 960 recording sites, with the 374 sites closest to the tip configured  
507 for recording. Neural signals were routed to an integrated base containing filtering, amplification,  
508 multiplexing, and digitization circuitry, and were streamed via an Ethernet link to the Open Ephys GUI<sup>30</sup>  
509 for online visualization and disk writing. Training data consisted of 1-hour of segments extracted from 10  
510 different ~2.5 hour long recordings (see below for information on the specific sessions used).

511 All fMRI datasets were collected using 3.0-Tesla Siemens MAGNETOM Trio A Tim scanner located at  
512 the ATR Brain Activity Imaging Center. Methodology was previously described<sup>29</sup>. Briefly, the functional  
513 images covered the entire brain (TR, 3,000 ms; TE, 30 ms; flip angle, 80°; voxel size, 3 × 3 × 3 mm; FOV,  
514 192 × 192 mm; number of slices, 50, slice gap, 0 mm). The dataset contained fMRI data from five subjects  
515 with 3 types of scanning sessions: “ses-perceptionTraining”, “ses-perceptionTest” and “ses-imageryTest”.  
516 We trained our denoiser on “ses-perceptionTraining” sessions and measured the denoising performance on  
517 “ses-perceptionTest” sessions.

### 518 **Visual stimulation.**

519 Visual stimuli in mice were displayed on an ASUS PA248Q LCD monitor, with 1,920 × 1,200 pixels (see  
520 original data publication for more details<sup>12</sup>). Stimuli were presented monocularly, and the monitor was  
521 positioned 15 cm from the eye and spanned 120° × 95° of visual space. Each monitor was gamma corrected  
522 and had a mean luminance of 50 cd/m<sup>2</sup>. To account for the close viewing angle, spherical warping was  
523 applied to all stimuli to ensure that the apparent size, speed, and spatial frequency were constant across the  
524 monitor as seen from the mouse’s perspective.

### 525 **Pre-processing steps**

526 Two-photon imaging movies were motion-corrected similarly to our previous publication<sup>12</sup>. The motion  
527 correction algorithm relied on phase correlation and only corrected for rigid translational errors. We used  
528 these motion corrected movies for denoising.

529 For Neuropixels recordings, the median across channels was subtracted to remove common-mode noise.  
530 The median was calculated across channels that were sampled  $\left[ \begin{smallmatrix} \text{---} \\ \text{SEP} \\ \text{---} \end{smallmatrix} \right]$  simultaneously, leaving out adjacent  
531 (even/odd) channels that are likely measuring the same  $\left[ \begin{smallmatrix} \text{---} \\ \text{SEP} \\ \text{---} \end{smallmatrix} \right]$  spike waveforms, as well as reference channels  
532 that contain no signal. For each sample, the median value of channels N:24:384, where N = [1,2,3,...,24],  
533 was calculated, and this value was subtracted from the same set of channels. Importantly, this step removes

534 noise that is *correlated* across channels, without affecting the independent noise targeted by  
535 DeepInterpolation. We performed denoising on the spike band after the median subtraction and offline  
536 filtering steps (150 Hz highpass) were applied.

537 For fMRI recordings, we used raw, unprocessed Nifti-1 volume data provided on OpenNeuro.org (see  
538 <https://openneuro.org/datasets/ds001246/versions/1.2.1>).

### 539 **Detection of ROIs in two-photon data**

540 We used two different segmentation algorithms in this study. The first one was described in a previous  
541 publication and leverages a succession of morphological filters to extract binary masks surrounding active  
542 pixels<sup>12</sup>. These masks are publicly available through the AllenSDK (see  
543 <https://allensdk.readthedocs.io/en/latest/>). In all analysis related to **Supp. Fig. 5**, we applied these binary  
544 masks to both raw and DeepInterpolation movies to extract matching calcium traces.

545 In all analysis related to **Fig. 2**, we used the sparse mode in Suite2P<sup>16</sup> to extract all individual filters. We  
546 used default parameter values of this mode except for `threshold_scaling`, which was set to 3. With the  
547 increased SNR achieved with DeepInterpolation, this single change limited the proportion of false positives  
548 in the final set of filters. Using Suite2p sorting GUI, we then manually sorted filters. We excluded filters  
549 present at the edge of the image, filters created by motion artifacts as well as features that did not cover any  
550 neuronal segment present in the max projection image. Somatic and non-somatic ROIs were also manually  
551 sorted based on the presence of a soma-excluded region in the filter. Filters that included a blurred out of  
552 plane soma were considered non-somatic. Once the final set of filters were extracted, we reapplied those  
553 weighted masks either to the original raw movie or the movie after DeepInterpolation to extract individual  
554 traces.

### 555 **Detection of neuronal units in electrophysiological data**

556 Kilosort2 was used to identify spike times and assign spikes to individual units<sup>31</sup>. Traditional spike sorting  
557 techniques extract snippets of the original signal and perform a clustering operation after projecting these

558 snippets into a lower-dimensional feature space. In contrast, Kilosort2 attempts to model the complete  
559 dataset as a sum of spike "templates." The shape and locations of each template is iteratively refined until  
560 the data can be accurately reconstructed from a set of  $N$  templates at  $M$  spike times, with each individual  
561 template scaled by an amplitude,  $a$ .

562 Kilosort2 was applied to the original and denoised datasets using identical parameters (all default  
563 parameters, except for ops.Th, which was lowered from [10 4] to [7 3] to increase the probability of  
564 detecting low-amplitude units). Because the spike detection threshold is relative to the overall noise level  
565 per channel, the absolute value of the threshold was lower following DeepInterpolation.

566 The Kilosort2 algorithm will occasionally fit a template to the residual left behind after another template  
567 has been subtracted from the original data, resulting in double-counted spikes. This can create the  
568 appearance of an artificially high number of ISI violations for one unit or artificially high zero-time-lag  
569 synchrony between nearby units. To eliminate the possibility that this artificial synchrony will contaminate  
570 data analysis, the outputs of Kilosort2 are post-processed to remove spikes with peak times within 5 samples  
571 (0.16 ms) and peak waveforms within 5 channels (~50 microns).

572 Kilosort2 generates templates of a fixed length (2 ms) that matches the time course of an extracellularly  
573 detected spike waveform. However, there are no constraints on template shape, which means that the  
574 algorithm often fits templates to voltage fluctuations with characteristics that could not physically result  
575 from the current flow associated with an action potential. The units associated with these templates are  
576 considered “noise,” and are automatically filtered out based on 3 criteria: spread (single channel, or >25  
577 channels), shape (no peak and trough, based on wavelet decomposition), or multiple spatial peaks  
578 (waveforms are non-localized along the probe axis). A final manual inspection step was used to remove  
579 any additional noise units that were not captured by the automated algorithm.

#### 580 **Quality control for electrophysiological units**

581 Units with action-potential-like waveforms detected by Kilosort2 are not necessarily high quality. To ensure  
582 that units met basic quality standards for further analysis, we filtered them using three different quality

583 metrics, computing with the `ecephys_spike_sorting` Python package  
584 ([github.com/AllenInstitute/ecephys\\_spike\\_sorting](https://github.com/AllenInstitute/ecephys_spike_sorting)):

585 • **ISI violations score < 0.5**. This metric is based on refractory period violations that indicate a unit  
586 contains spikes from multiple neurons. The ISI violations metric represents the relative firing rate  
587 of contaminating spikes. It is calculated by counting the number of violations <1.5 ms, dividing by  
588 the amount of time for potential violations surrounding each spike, and normalizing by the overall  
589 spike rate. It is always positive (or 0), but has no upper bound. See<sup>32</sup> for more details.

590 • **Amplitude cutoff < 0.1**. This metric provides an approximation of a unit’s false negative rate.  
591 First, a histogram of spike amplitudes is created, and the height of the histogram at the minimum  
592 amplitude is extracted. The percentage of spikes above the equivalent amplitude on the opposite  
593 side of the histogram peak is then calculated. If the minimum amplitude is equivalent to the  
594 histogram peak, the amplitude cutoff is set to 0.5 (indicating a high likelihood that >50% of spikes  
595 are missing). This metric assumes a symmetrical distribution of amplitudes and no drift, so it will  
596 not necessarily reflect the true false negative rate.

597 • **Presence ratio > 0.95**. The presence ratio is defined as the fraction of blocks within a session that  
598 include 1 or more spikes from a particular unit. Units with a low presence ratio are likely to have  
599 drifted out of the recording, or could not be tracked by Kilosort2 for the duration of the experiment.

600 Applying these quality metrics removed 54% of detected units in the original data, and 60% of units after  
601 denoising. Spike sorting after DeepInterpolation found more units regardless of the threshold used for each  
602 QC metric (**Supp. Fig. 6B**).

### 603 **Procedure for finding matched vs. new units**

604 Following spike sorting steps, we searched for overlapping spikes between pairs of units detected before  
605 and after denoising. Spikes were considered to overlap if they had a peak occurring within  $\pm 5$  channels and  
606 0.5 ms of one another. The number of overlapping spikes was used to compute three metrics, using the  
607 original spike trains as “ground truth”<sup>33</sup>.

608       • **Precision.** The fraction of denoised spikes that were also found in the original data.

609       • **Recall.** The fraction of original spikes that were also found in the denoised data.

610       • **Accuracy.**  $N_{\text{match}} / (N_{\text{denoised}} + N_{\text{original}} - N_{\text{match}})$

611 Units were considered matched if they had an accuracy exceeding 0.7. Units were considered novel after

612 denoising if they had a total precision (summed over all original units) less than 0.5. Units were considered

613 undetected after denoising if they had a total recall (summed over all denoised units) less than 0.5.

#### 614 **Training of denoising networks**

615 We used 3 different sets of environments for training deep neural networks. Two-photon denoising

616 networks were trained using TensorFlow 1.9.0, keras 2.2.4 and CUDA 9.0. Electrophysiology denoising

617 networks were trained using TensorFlow 2.0 and CUDA 10.0 through their built-in Keras libraries. fMRI

618 denoising networks were trained using TensorFlow 2.2 and CUDA 10.1.

619 We utilized NVIDIA Titan X, Geforce GTX 1080, and Tesla V100-SXD2 GPUs available on the Allen

620 Institute internal computing clusters. The fMRI denoising network was trained on Amazon AWS using the

621 p2.8xlarge and p3.8xlarge instance type depending on availability.

622 We used an L1 loss during training for both two-photon imaging and fMRI datasets. Electrophysiological

623 datasets were trained with an L2 loss. All training was done with the RMSProp gradient descent algorithm

624 implemented in keras. Two-photon denoiser was trained with a batch size of 5 so as to fit on available GPU

625 memory. The learning rate was set to  $5 \cdot 10^{-4}$ . The Neuropixels denoiser was trained with a batch size of

626 100, with the learning rate was set to  $10^{-4}$ . The fMRI denoiser was trained with a batch size of 10,000. The

627 larger batch size was allowed by the smaller input-output size. The fMRI denoiser was trained with a

628 learning scheduler, initialized at  $10^{-4}$ , dropping by half every 45 millions samples.

629 To facilitate training, all samples were mean-centered and normalized by a single shared value for each

630 experiment during training. For two-photon movies, the mean and standard deviation was pre-calculated

631 using the first 100 frames of the movie. For Neuropixels recordings, the mean and standard deviation was

632 pre-calculated using the 200,000 samples. For fMRI recordings, we extracted the centered volume of the



633 movie that was  $1/4^{\text{th}}$  of the total movie size and averaged all voxels.

634 A detailed step by step pseudocode description of this process is available in **Supp. Appendix 1**.

### 635 **Inference of final datasets with trained networks**

636 Once a DeepInterpolation network was learned, inference was performed by streaming entire experiments  
637 through the same fixed network. To match the conditions in training, each dataset was mean-centered and  
638 normalized following the same procedure used during training. Output denoised data were brought back to  
639 their original scale after going through the DeepInterpolation model.

### 640 **Quantification of SNR**

641 In the two-photon data, SNR was defined as the ratio of the mean fluorescence value divided by the standard  
642 deviation along the temporal dimension. In ideal photon shot-noise limited conditions, this SNR is  
643 proportional to the square root of N, where N is the photon flux detected per pixel.

644 In the Neuropixels data, SNR for individual units was defined as the ratio of the maximum peak-to-peak  
645 waveform amplitude to the RMS noise on the peak channel.

646 In the fMRI data, tSNR of individual voxels was defined as the ratio of the mean BOLD signal value  
647 divided by standard deviation along the temporal dimension.

### 648 **Analysis of natural movie responses**

#### 649 Noise and signal correlation analysis in two-photon data

650 For each natural movie presentation, we extracted the corresponding ROI traces. The signal correlation was  
651 computed by averaging all 10 presentations of the movie and calculating the Pearson pairwise correlation  
652 of those averaged responses between pairs of ROIs.

653 The noise correlation was calculated by calculating the Pearson correlation of individual trial responses  
654 between a pair of ROIs and then averaging all 10 presentations of the movie.

#### 655 Response reliability in two-photon data

656  
657 The response reliability was calculated by first calculating the pearson correlation matrix of an ROI

658 individual response to a single presentation of a natural movie with other trials. For each ROI, we then  
659 excluded auto-correlation values and averaged out all pairwise combinations. This process yielded a single  
660 reliability measure for each ROI.

#### 661 Significant responses in Neuropixels data

662 When analyzing spike data, the trial-to-trial correlation of the natural movie response depends heavily on  
663 the bin size chosen. To determine responsiveness for Neuropixels data, we instead compared each unit’s  
664 natural movie PSTH (averaged across 25 trials) to a version in which all bins were temporally shuffled. A  
665 Kolmogorov–Smirnov test between the original and shuffled PSTHs was used to determine the probability  
666 that the response could have occurred by chance ( $p < 0.05$  for significantly responsive units; see **Supp. Fig.**  
667 **7B** for an example).

#### 668 **Synthetic data generation**

669 We created realistic synthetic calcium imaging datasets using a recent approach called *in silico* Neural  
670 Anatomy and Optical Microscopy (NAOMi)<sup>14</sup>. The code repository was kindly made available to us by  
671 Alex Song and Adam Charles. Given the computational load of the model, we generated a single dataset  
672 made of 15,000 frames simulating a 400x400  $\mu\text{m}^2$  field of view. Except for the field of view size, all  
673 parameters were set to default values. We did not use a pre-trained network to denoise this dataset and  
674 trained our DeepInterpolation model directly on this simulated data.

#### 675 **Comparison with Penalized Matrix Factorisation**

676 A PMD docker repository was kindly made available to us by the Paninski lab and we collaborated to run  
677 the PMD algorithm in the best conditions<sup>13</sup>. PMD was run on an AWS instance (m5.8xlarge) under a pre-  
678 packaged jupyterhub environment (<https://hub.docker.com/r/paninski/trefide>). All movies were pre-  
679 centered and normalized using code available in the trefide package (psd\_noise\_estimate).

#### 680 **DeepInterpolation code availability**

681 Code for DeepInterpolation and all other steps in our algorithm are available online through a GitHub  
682 repository (<https://github.com/AllenInstitute/deepinterpolation>). Example training and inference tutorial

683 code are available on the repository. Several Docker inference containers are available on Docker Hub (see  
684 list below for more details). They allow inference using the TensorFlow ModelServer framework.  
685 Code and data to regenerate all figures presented in this manuscript is available online through a GitHub  
686 repository ([https://github.com/AllenInstitute/deepinterpolation\\_paper](https://github.com/AllenInstitute/deepinterpolation_paper)).

687

688 Two-photon Ai93 excitatory line DeepInterpolation network:

689 **Key recording parameters:**

- 690 ● 30Hz sampling rate, 400x400  $\mu\text{m}^2$  field of view, 512x512 pixels.
- 691 ● 0.8 NA objective.
- 692 ● 910 nm excitation wavelength.
- 693 ● Gcamp6f calcium indicator.
- 694 ● Ai93 reporter line expressed in excitatory neurons.

695 **Docker hub id :** 245412653747/deep\_interpolation:allen\_400um\_512pix\_30hz\_ai93

696

697 Two-photon Ai148 excitatory line DeepInterpolation network:

698 **Key recording parameters:**

- 699 ● 30 Hz sampling rate, 400x400  $\mu\text{m}^2$  field of view, 512x512 pixels.
- 700 ● 0.8 NA objective.
- 701 ● 910 nm excitation wavelength.
- 702 ● Gcamp6f calcium indicator.
- 703 ● Ai93 reporter line expressed in excitatory neurons.

704 **Pre-processing:** Individual movies were motion corrected. Each movie recording was mean-centered and  
705 normalized with a single pair of value for all pixels

706 **Docker hub id :** 245412653747/deep\_interpolation:allen\_400um\_512pix\_30hz\_ai148

707

708 Neuropixels DeepInterpolation network:

709 **Key recording parameters:**

- 710 ● Neuropixels Phase 3a probes
- 711 ● 374 simultaneous recording sites across 3.84 mm, 10 reference channels
- 712 ● Four-column checkerboard site layout with 20  $\mu\text{m}$  spacing between rows
- 713 ● 30 kHz sampling rate
- 714 ● 500x hardware gain setting
- 715 ● 500 Hz high pass filter in hardware, 150 Hz high-pass filter applied offline

716 **Pre-processing:** Median subtraction was applied to individual probes to remove signals that were  
717 common across all recording sites. Each probe recording was mean-centered and normalized with a single  
718 pair of value for all nodes on the probe.

719 **Docker hub id :** 245412653747/deep\_interpolation:allen\_neuropixel

720

721 fMRI DeepInterpolation network:

722 **Key recording parameters:**

- 723 ● TR, 3,000 ms; TE, 30 ms; flip angle, 80°; voxel size, 3  $\times$  3  $\times$  3 mm; FOV, 192  $\times$  192 mm;
- 724 number of slices, 50, slice gap, 0 mm

725 **Pre-processing:** N/A

726 **Docker hub id :** 245412653747/deep\_interpolation:allen\_3\_3\_3\_tr\_3000\_fmri

727 **Data availability**

728 The two-photon imaging and Neuropixels raw data can be downloaded from the following S3 bucket:  
729 `arn:aws:s3:::allen-brain-observatory`

730

731 Two-photon imaging files are accessed according to Experiment ID, using the following path:

732 visual-coding-2p/ophys\_movies/ophys\_experiment\_<Experiment ID>.h5

733 We used a random subset of 1144 experiments for training the denoising network for Ai93, and 397

734 experiments for training the denoising network for Ai148. The list of used experiments IDs is available in

735 json files (in deepinterpolation/examples/json\_data/) on the DeepInterpolation GitHub repository. The

736 majority of these experiment IDs are available on the S3 bucket. Some experimental data has not been

737 released to S3 at the time of publication.

738 Neuropixels raw data is accessed by Experiment ID and Probe ID, using the following paths:

739 visual-coding-neuropixels/raw-data/<Experiment ID>/<Probe ID>/spike\_band.dat

740 The dat files have the median subtraction post-processing applied, but do not include an offline highpass

741 filter. Prior to DeepInterpolation, we extracted 3600 seconds of data from each of the recordings listed in

742 **Table 1.**

743

Experiment ID	Probe ID	Start time	Brain Structures	Used for training
778998620	792626851	4665.8179 s	VISl (LM), CA1, CA3	No
787025148	792586842	4672.1967 s	VISp (V1), SUB, LP	No
768515987	773549856	4672.2224 s	VISrl (RL), CA1, DG, APN	Yes
794812542	810758787	4665.8332 s	VISrl (RL), CA1, DG, APN	No
778998620	792626853	4665.8179 s	VISal (AL), CA1, CA3, DG, LGv	No
767871931	773462997	4672.2352 s	VISal (AL), CA1, CA3, DG, LT	Yes
779839471	792645501	4672.3841 s	VIS, CA1, CA3, DG, LGd	No
771160300	773621948	4672.3791 s	VISal (AL), CA1, CA3, DG, LGd	Yes
781842082	792586881	4672.1312 s	VISp (V1), SUB, SGN	No
793224716	805124815	4675.2401 s	VISrl (RL), CA1, DG, LP, APN	No

744

745 **Table 1. Details of Neuropixels data used for DeepInterpolation training and inference.**

746 To train the fMRI denoiser, we used datasets that can be downloaded from an S3 bucket:

747 [arn:aws:s3:::openneuro.org/ds001246](https://openneuro.org/ds001246)

748

749 We trained our denoiser on all “ses-perceptionTraining” sessions, across 5 subjects (3 perception training

750 sessions, 10 runs each).