Supplementary Material for "OnACID: Online Analysis of Calcium Imaging Data in Real Time"

Andrea Giovannucci, Johannes Friedrich, Matthew Kaufman, Anne K. Churchland, Dmitri Chklovskii, Liam Paninski, and Eftychios A. Pnevmatikakis

A Algorithmic description

Here we present in pseudocode the various steps of the online processing pipeline. For ease of exposition, some details and speedup tricks used in the actual implementation have been omitted.

Algorithms S1 and S2 describe the simple greedy procedure for partitioning the components into disjoint groups where the elements of each group do not overlap spatially with each other. This procedure is used for updating the traces of the neurons in vector form (Alg. S3) leading to substantial speed benefits. Algorithm S4 describes the procedure of detecting and screening possible new components. Finally, Algorithm S5 describes the process of updating the shapes, similar to [3].

Algorithm S1 DetermineGroups

Require: Spatial components matrix A, number of components K

$\mathcal{G} = \emptyset$
for $i = 1 \to K$ **do**
  $\mathcal{G} \leftarrow$ JoinGroups$(A[:, 1:i-1], \mathcal{G}, i-1, \mathbf{a}_i)$
end for
return $\mathcal{G}$

Algorithm S2 JoinGroups

Require: Spatial components matrix A, current groups $\mathcal{G}$, number of components K, new component $\mathbf{a}$

$N_G = |\mathcal{G}|$ ▷ number of groups
repeat = **True**
$g \leftarrow 1$
while repeat **do**
  if $g \leq N_G$ **then**
    if $\mathbf{a}^\top \mathbf{a}_l = 0, \forall l \in G_j$ **then** ▷ Test for overlap with current group
      $G_g \leftarrow G_g \cup \{K+1\}$
      repeat = **False**
    else
      $g \leftarrow g + 1$
    end if
  else
    $N_G \leftarrow N_G + 1$
    $G_{N_G} = \{K+1\}$ ▷ Create a new group
    $\mathcal{G} \leftarrow \{\mathcal{G}, G_{N_G}\}$ ▷ Add to list of groups
    repeat = **False**
  end if
end while
return $\mathcal{G}$

Algorithm S3 UpdateTraces

Require: Spatial footprints matrix $\tilde{A} = [A, \mathbf{b}]$, current value of temporal traces $\tilde{\mathbf{c}} = [\mathbf{c}; f]$, current data frame $\mathbf{y}$, groups $\mathcal{G}$, tolerance level ε.

$\mathbf{u} = \tilde{A}^\top \mathbf{y}$
$V = \tilde{A}^\top \tilde{A}$
$\mathbf{V} = \text{diag}\{V\}$
$\tilde{c}_{old} \leftarrow 0$
while $\|\tilde{c} - \tilde{c}_{old}\| \geq \varepsilon \|\tilde{c}_{old\|}$ **do**
  $\tilde{c}_{old} \leftarrow \tilde{c}$
  for $i = 1 \to |\mathcal{G}|$ **do**
    $\tilde{c}[G_i] = \max\left(\tilde{c}[G_i] + \frac{\mathbf{u}[G_i] - V[G_i,:]\tilde{c}}{\mathbf{v}[G_i]}, 0\right)$ ▷ (Division is pointwise)
  end for
end while
return $\tilde{c}$

Algorithm S4 DetectNewComponents

Require: Spatial footprints matrix $[A, \mathbf{b}]$, temporal traces matrix $[C; \mathbf{f}]$, current number of components K, current state of groups $\mathcal{G}$, current residual buffer R_{buf}, current data frame $\mathbf{y}$, sufficient statistics W, M. Parameters: radius of Gaussian kernel τ, threshold for correlation in space θ_s, threshold for correlation in time r_t.

repeat = **True**
$R_{\text{buf}} \leftarrow [R_{\text{buf}}[:, 1 : l_b - 1], \mathbf{y} - [A, \mathbf{b}][C; \mathbf{f}][:, \text{end}]]$ ▷ Update residual buffer
$M_d = \text{Median}(R_{\text{buf}})$
$R_{\text{buf}} \leftarrow R_{\text{buf}} - M_d$ ▷ Subtract median along time for every pixel
$V \leftarrow \text{Filter}(R_{\text{buf}}, \text{GaussianKernel}(\tau))$ ▷ Filter residual in space
$E \leftarrow \sum_i V[:, i].^2$ ▷ Compute energy value for each pixel
while repeat **do**
  $(i_x, i_y) = \arg\max E$ ▷ Find the point of maximum variance
  $N_{(i_x,i_y)} = \{(x, y) : |x - i_x| \leq \tau, |y - i_y| \leq \tau\}$ ▷ Define a neighborhood around (i_x, i_y)
  $[\mathbf{a}_{\text{new}}, \mathbf{c}_{\text{new}}] = \text{NNMF}(R_{\text{buf}}[N_{(i_x,i_y)}, :], 1)$ ▷ Perform a local rank-1 NMF
  $r = \text{Corr}(\mathbf{a}_{\text{new}}, \text{Mean}(R_{\text{buf}}))$ ▷ Compute correlation coefficient in space
  $o = \text{Find}(\mathbf{a}_{\text{new}}^\top A[N_{(i_x,i_y)}, :] > 0$ ▷ Find components that overlap
  if $\exists j \in o : \text{Corr}(\mathbf{c}_{\text{new}}, C[j, t - l_b + 1 : t]) > r_t$ **then**
    $r \leftarrow 0$ ▷ Detect possible duplicates and stop procedure
  end if
  if $r > \theta_s$ **then** ▷ New component is accepted
    Zero-pad $\mathbf{a}_{\text{new}}$ and $\mathbf{c}_{\text{new}}$ to match dimensionality
    $K \leftarrow K + 1$
    $\mathcal{G} \leftarrow \text{JoinGroups}(A, \mathcal{G}, \mathbf{a}_{\text{new}})$
    $A \leftarrow [A, \mathbf{a}_{\text{new}}]$
    $C \leftarrow [C; \mathbf{c}_{\text{new}}]$
    $R_{\text{buf}} \leftarrow R_{\text{buf}} - \mathbf{a}_{\text{new}} \mathbf{c}_{\text{new}}$
    $V \leftarrow V - \mathbf{a}_{\text{new}}^2 \|\mathbf{c}_{\text{new}}\|^2$
    $W, M \leftarrow \text{UpdateSuffStatistics}(W, M, \mathbf{y}_t, \mathbf{c}_{\text{new}})$ ▷ Equation (5)
  else
    repeat = **False**
  end if
end while
return $[A, \mathbf{b}], [C, \mathbf{f}], K, \mathcal{G}, R_{\text{buf}}, W, M$

Algorithm S5 UpdateShapes

Require: Sufficient statistics W, M, current value of spatial footprints $\tilde{A} = [A, b]$, list of components to be updated l, maximum number of iterations iterations $\mathrm{m_{iter}}$

iter $\leftarrow 0$
while iter $< \mathrm{m_{iter}}$ **do**
  for $i \in l$ **do**
    $\mathbf{p} = \mathrm{find}(\tilde{A}[:, i] > 0)$ ▷ Find the pixels where component i can be non-zero
    $\tilde{A}[\mathbf{p}, i] = \max\left(\tilde{A}[\mathbf{p}, i] + \frac{W[\mathbf{p}, i] - \tilde{A}[\mathbf{p}, :]M[:, i]}{M[i, i]}, 0\right)$
  end for
  iter $\leftarrow$ iter $+ 1$
end while
return $\tilde{A}$

B Dataset Details

Parietal cortex dataset: Data was obtained from the parietal cortex of a transgenic GCaMP6f-expressing mouse during a behavioral task. Field of view was approximately 500×500 μm^2 (512×512 pixels) in size and at depth 125μm below the dura surface. Horizontal scans of the laser were performed using a resonant galvanometer, resulting in a frame acquisition rate of 30Hz. More details can be found in [2].

Hippocampal dataset: Data was obtained from the hippocampus of a transgenic GP2.11 (Thy1-GCaMP3) mouse generated by the Janelia Farms GENIE Project (Jackson Labs, C57BL/6J-Tg(Thy1-GCaMP3)GP2.11Dkim/J). FOV was approximately 500×500 μm^2, of size 512 × 512 pixels, cropped to 483 × 492 pixels after rigid registration and removal of empty border lines. Horizontal scans of the laser were performed using a resonant galvanometer, resulting in a frame acquisition rate of 30Hz. More details can be found in [1].

C Supplementary Movie

Evolution of the OnACID algorithm on toy simulated data: Top. Raw movie (left). Denoised movie reconstructed from all components (middle). Noiseless ground truth (right). Bottom. Residual movie (left). Inferred (middle) and ground truth (right) spatial components. A 64×64 pixel FOV containing 35 artifical neurons was simulated for this example. Movie is truncated in time for space reasons.

D Simulation details

We generated a dataset of size 256×256 pixels and duration $T = 2000$ frames containing $N = 400$ neurons. The neural centers $\{\mathbf{c}\}_1^N$ were generated using a Halton sequence to cover the space uniform pseudo-randomly.

The unnormalized neural shapes were modeled as the difference of two 2D-Gaussians.

$$\tilde{a}(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{c})^\top \mathrm{diag}\left(\sigma_x^{-1}, \sigma_y^{-1}\right)(\mathbf{x}-\mathbf{c})\right) \tag{1}$$

$$- k \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{c})^\top \mathrm{diag}\left((0.75\sigma_x)^{-1}, (0.75\sigma_y)^{-1}\right)(\mathbf{x}-\mathbf{c})\right) \tag{2}$$

where $\mathbf{x}$ denotes the position of the considered pixel. To incorporate heterogeneity the standard-deviation σ_x and σ_y in x- and y-direction of the wider Gaussian was drawn i.i.d. uniform randomly from the interval [2.5, 3.5]. These values were multiplied by 0.75 to obtain the standard-deviation of the smaller subtracted Gaussian. The magnitude k of the subtracted Gaussian was drawn i.i.d. uniform randomly from the interval $[0.2, 0.8]$ for each neuron.

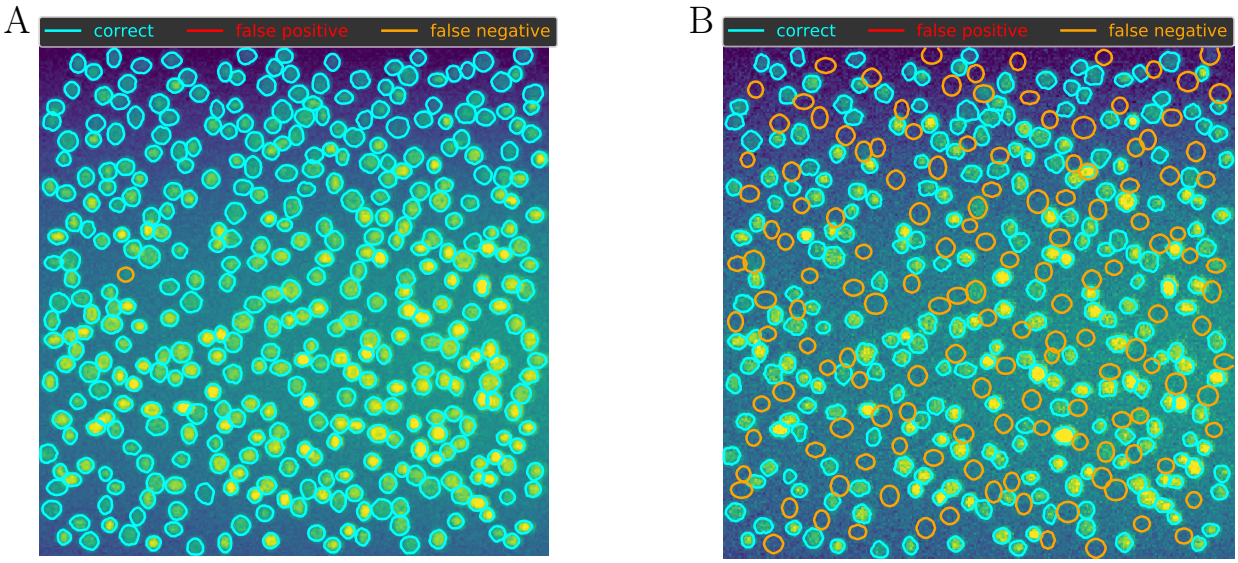

Figure S1: Simulated data. Detected components in batch mode. A) Full data B) Short initial batch.

The spike train $\mathbf{s}$ of each neuron was drawn from a homogeneous Poisson process. The neural firing rate was 0.5 Hz and the frame rate 30 Hz. The calcium traces C were obtained by convolving the spike trains S with an exponentially decaying kernel with time constant 1 s.

The background B was modeled as rank 1 term, where the temporal and spatial component were each drawn from a Kronecker Gaussian process with RBF kernel. The temporal length scale was 300 frames and the spatial length scale 50 pixels. Finally, the simulated raw data is the sum of background B and neural contribution AC corrupted by Gaussian noise, $Y \sim \mathcal{N}(B + AC, 0.2^2)$.

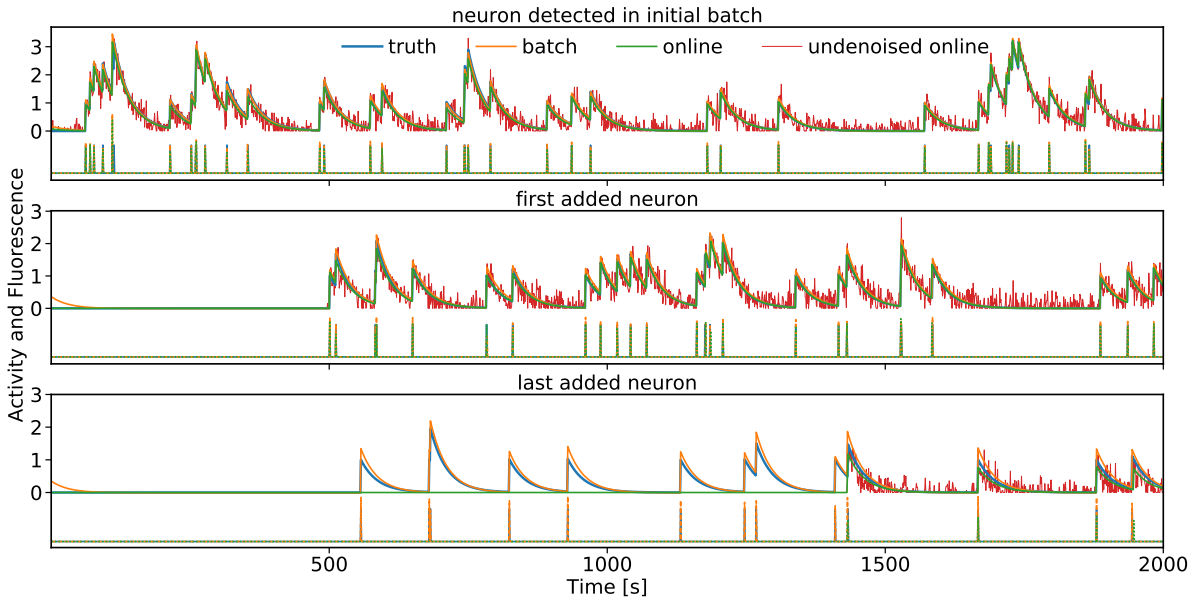

Figure S2: Simulated data. Traces of three neurons selected by time of detection. Upper traces show demixed (red), denoised (orange, green) and ground truth (blue) calcium fluorescence. Lower traces show deconvolved neural activity using the same coloring scheme.

E Detailed comparison between OnACID and manual annotations for the hippocampal 2-photon dataset

In the following pages, Figures S3-S8 show the detailed matches and mismatches between OnACID and the two manual annotations, as well as the two manual annotations against each other.

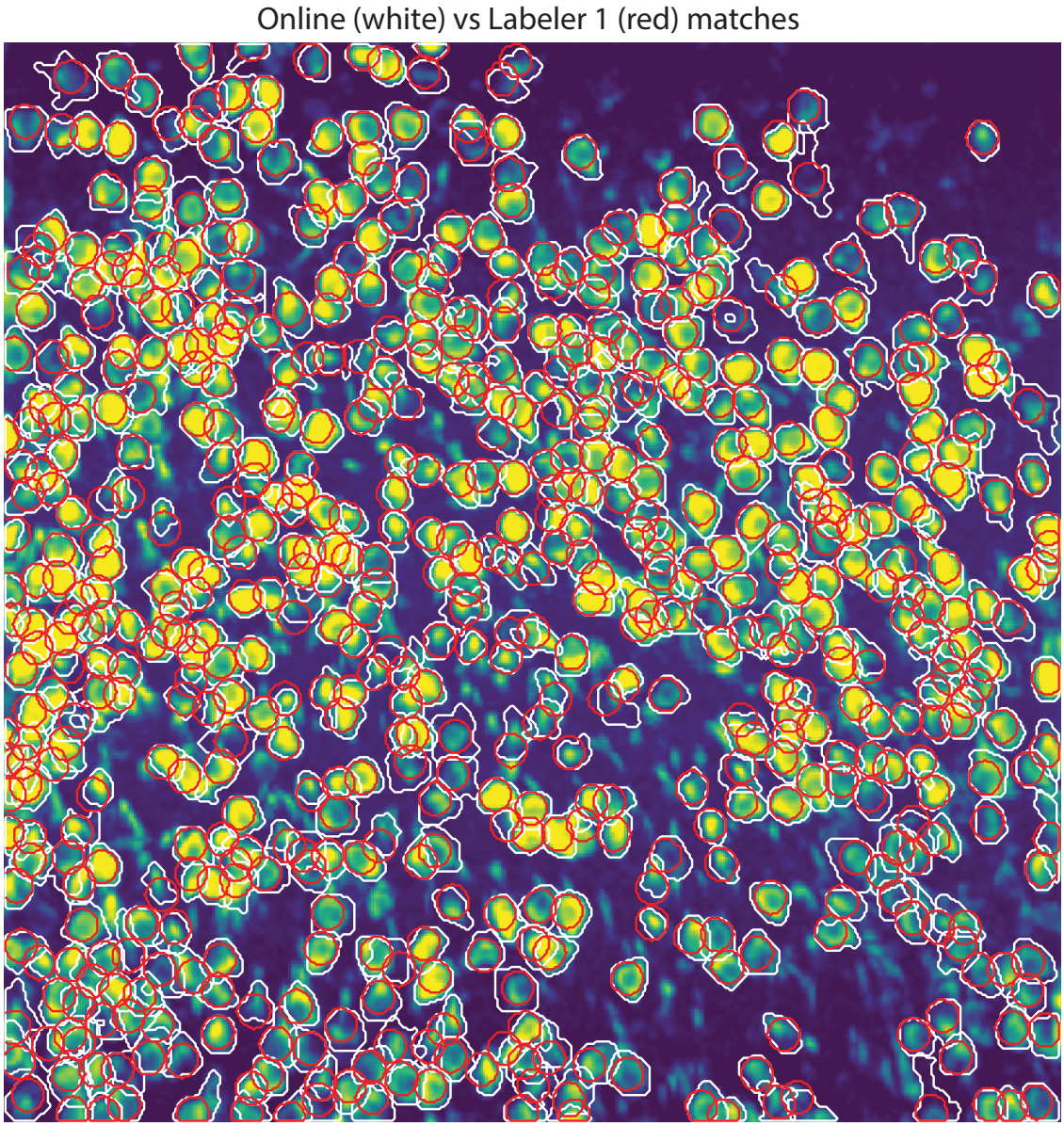

Figure S3: Matches between OnACID (white) and Labeler 1 (red).

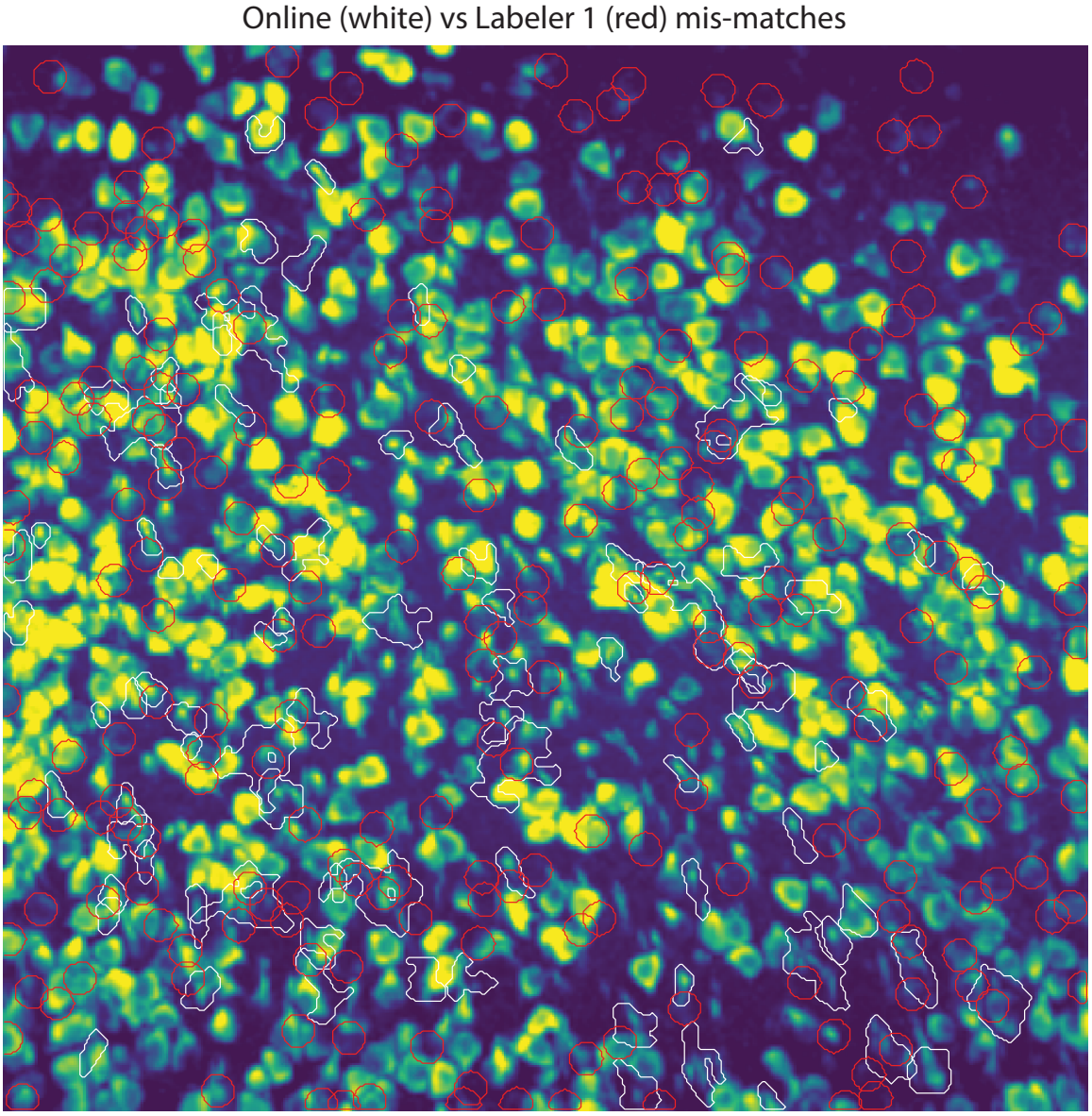

Figure S4: Mismatches between OnACID (white) and Labeler 1 (red).

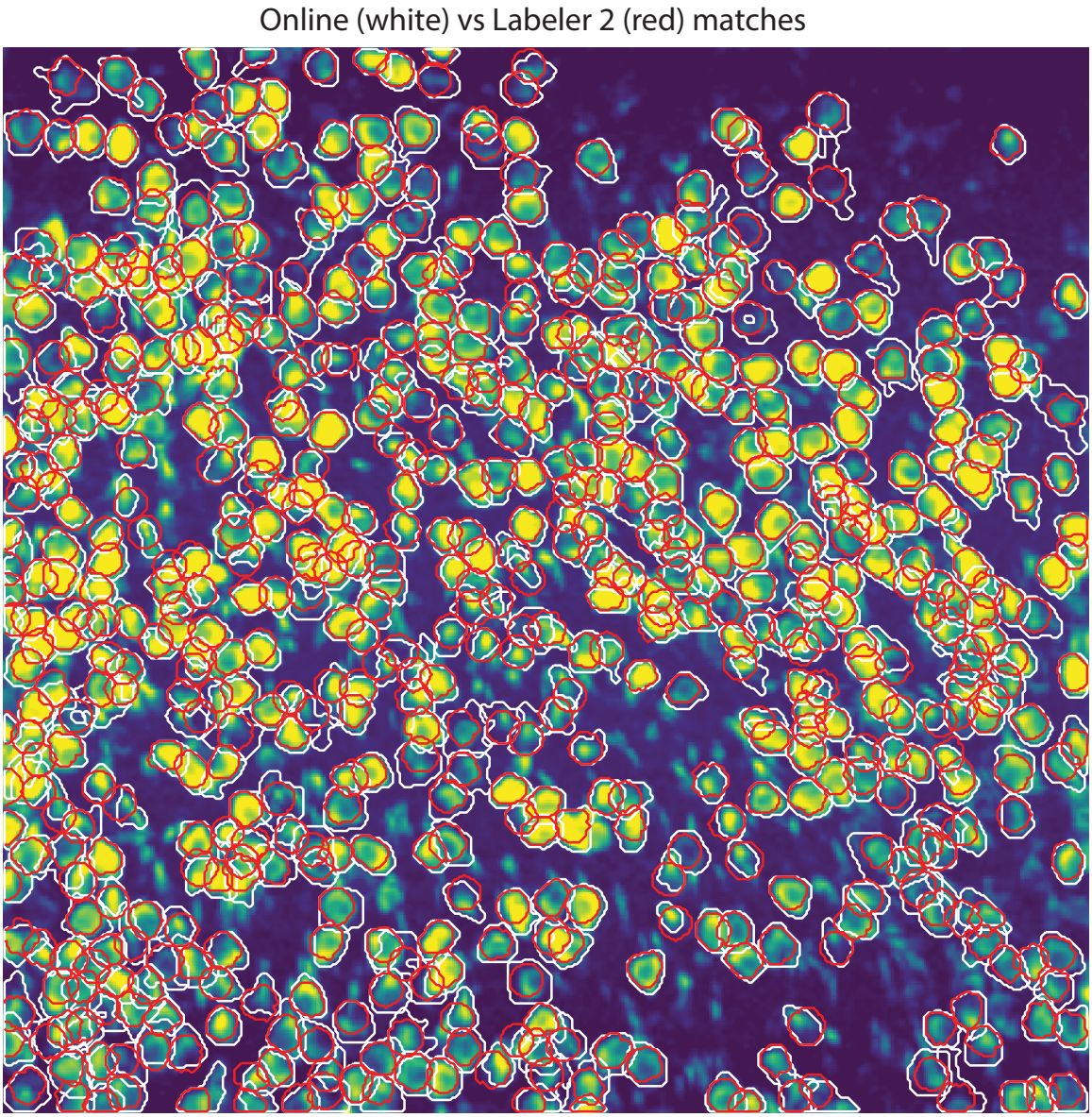

Figure S5: Matches between OnACID (white) and Labeler 2 (red).

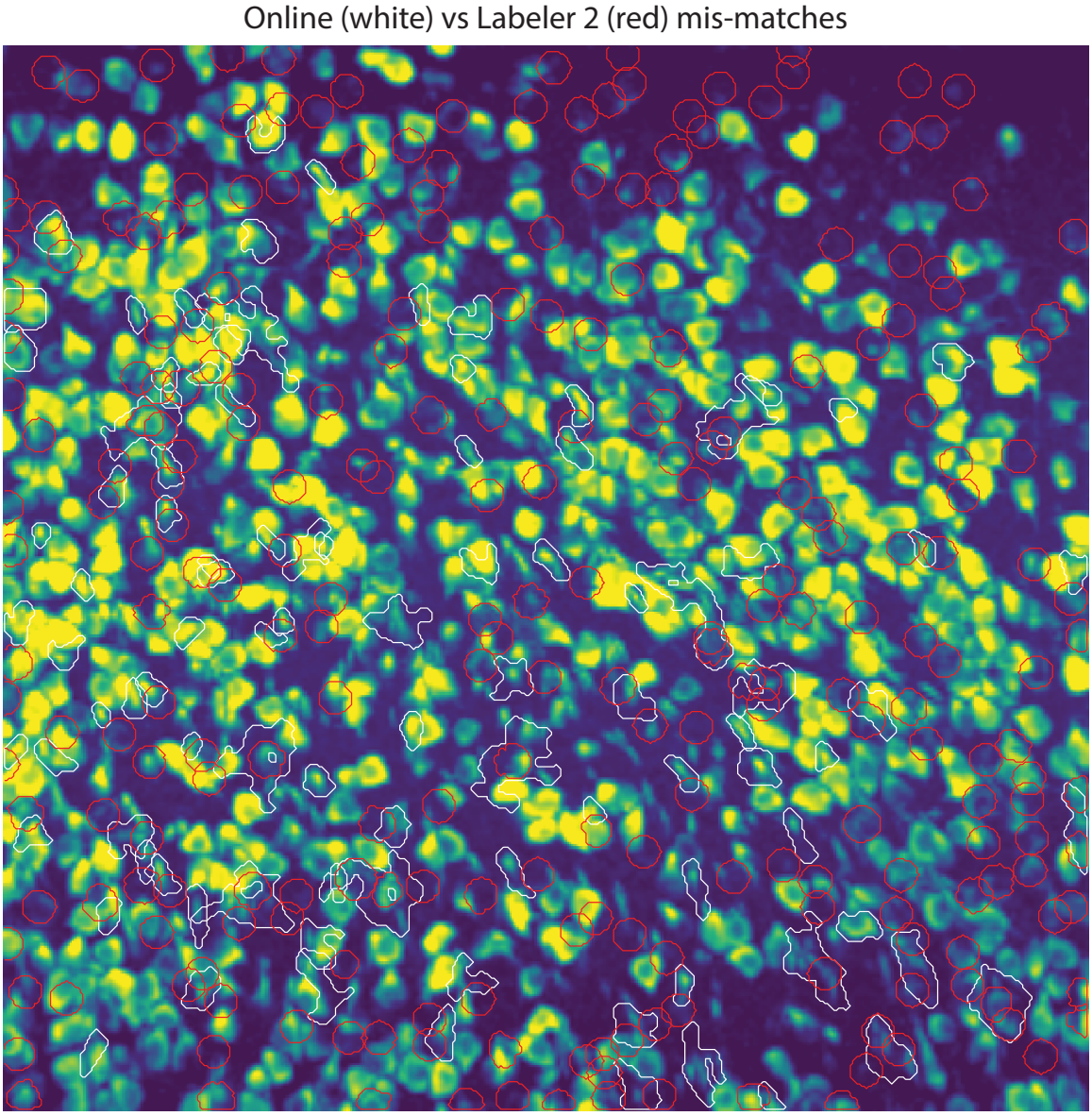

Figure S6: Mismatches between OnACID (white) and Labeler 2 (red).

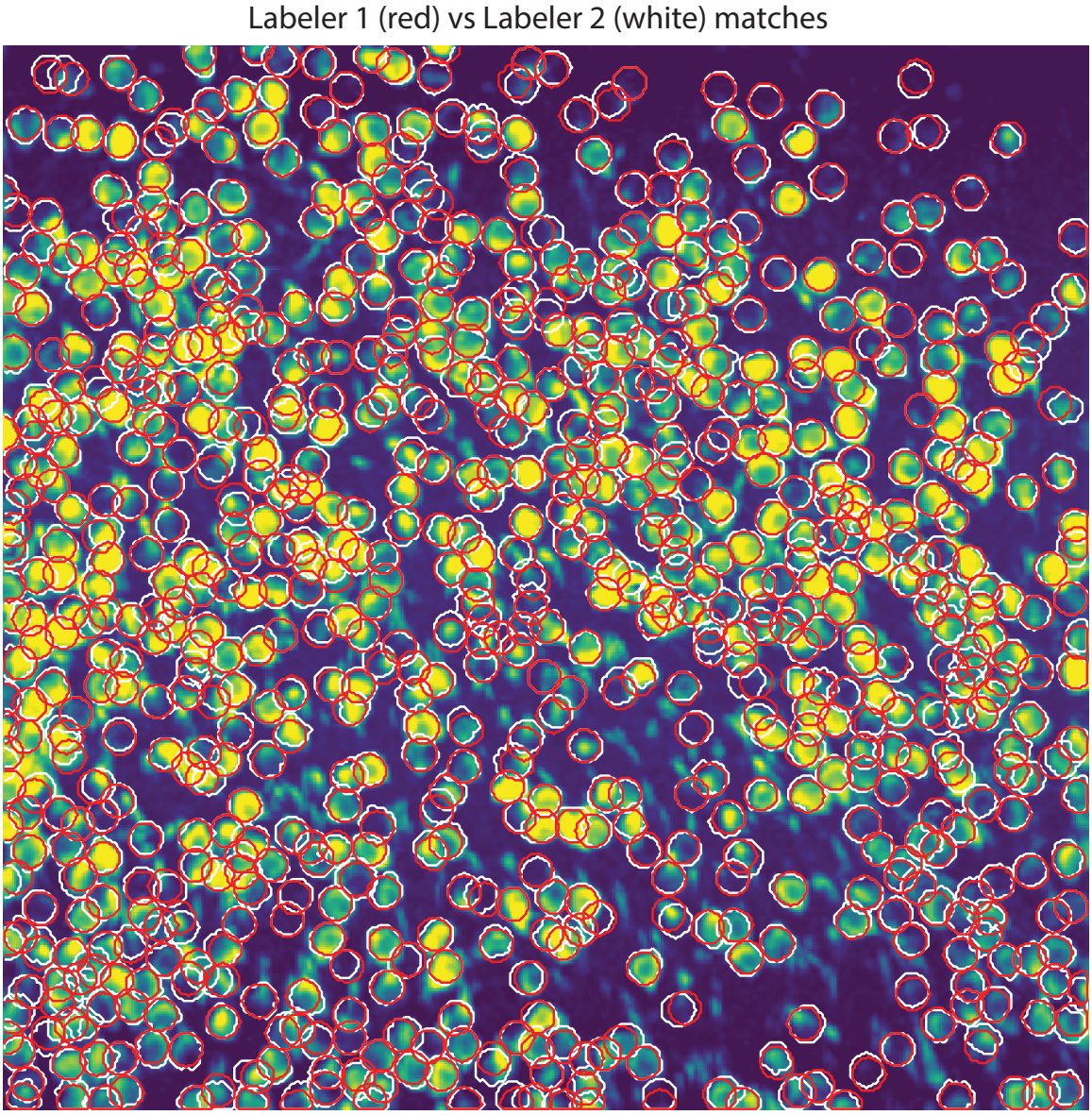

Figure S7: Matches between Labeler 1 (red) and Labeler 2 (white). The two labelers annotated the dataset independently and have a high degree of matching ($F_1 = 0.89$). The contour shapes are also similar for both annotators, as expected from the labeling process using the ImageJ Cell Wand tool.

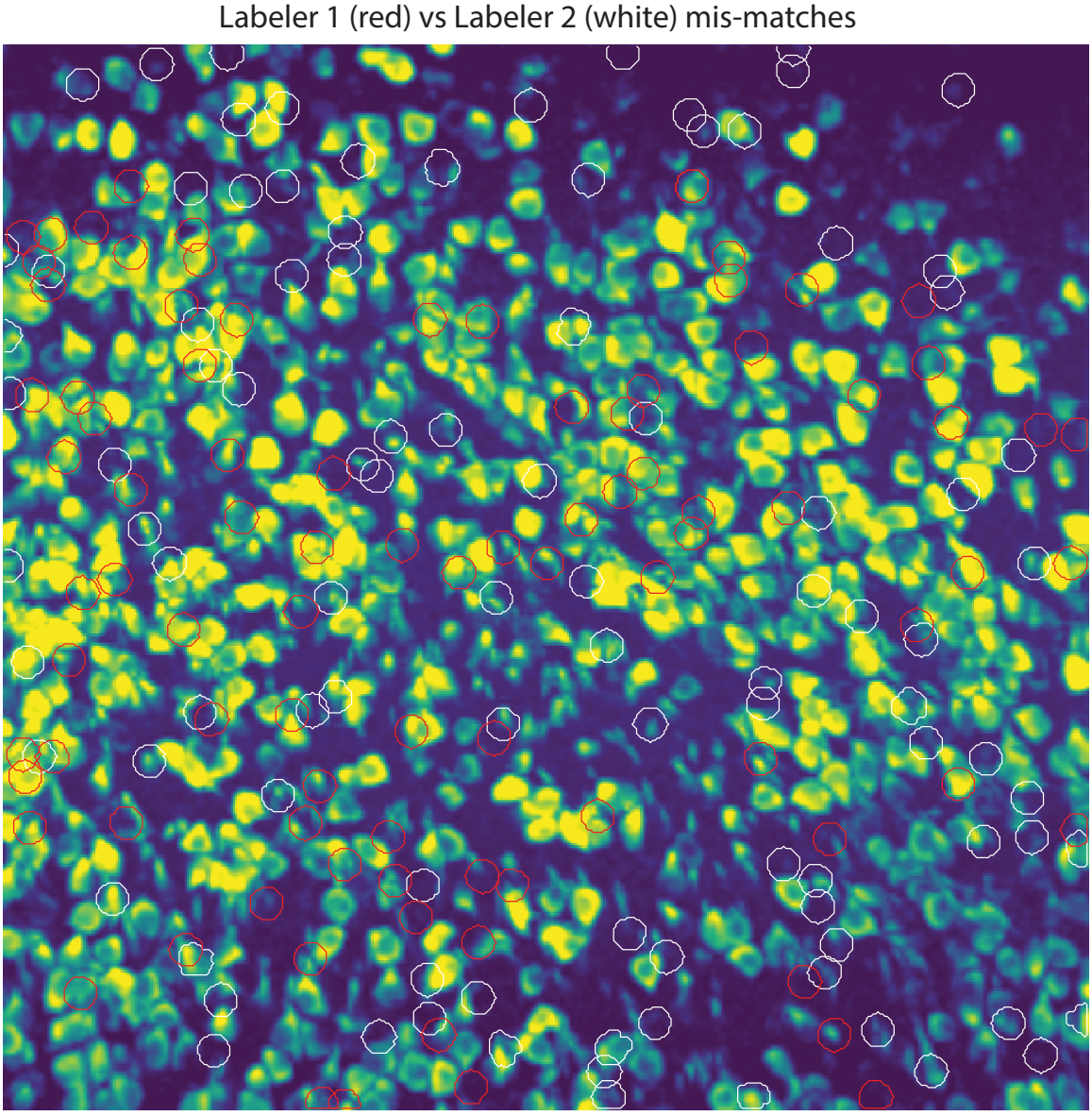

Figure S8: Mismatches between Labeler 1 (red) and Labeler 2 (white).

F Detailed comparison between OnACID and manual annotations for the parietal cortex 2-photon dataset

In the following pages, Figures S9-S14 show the detailed matches and mismatches between OnACID and the two manual annotations, as well as the two manual annotations against each other.

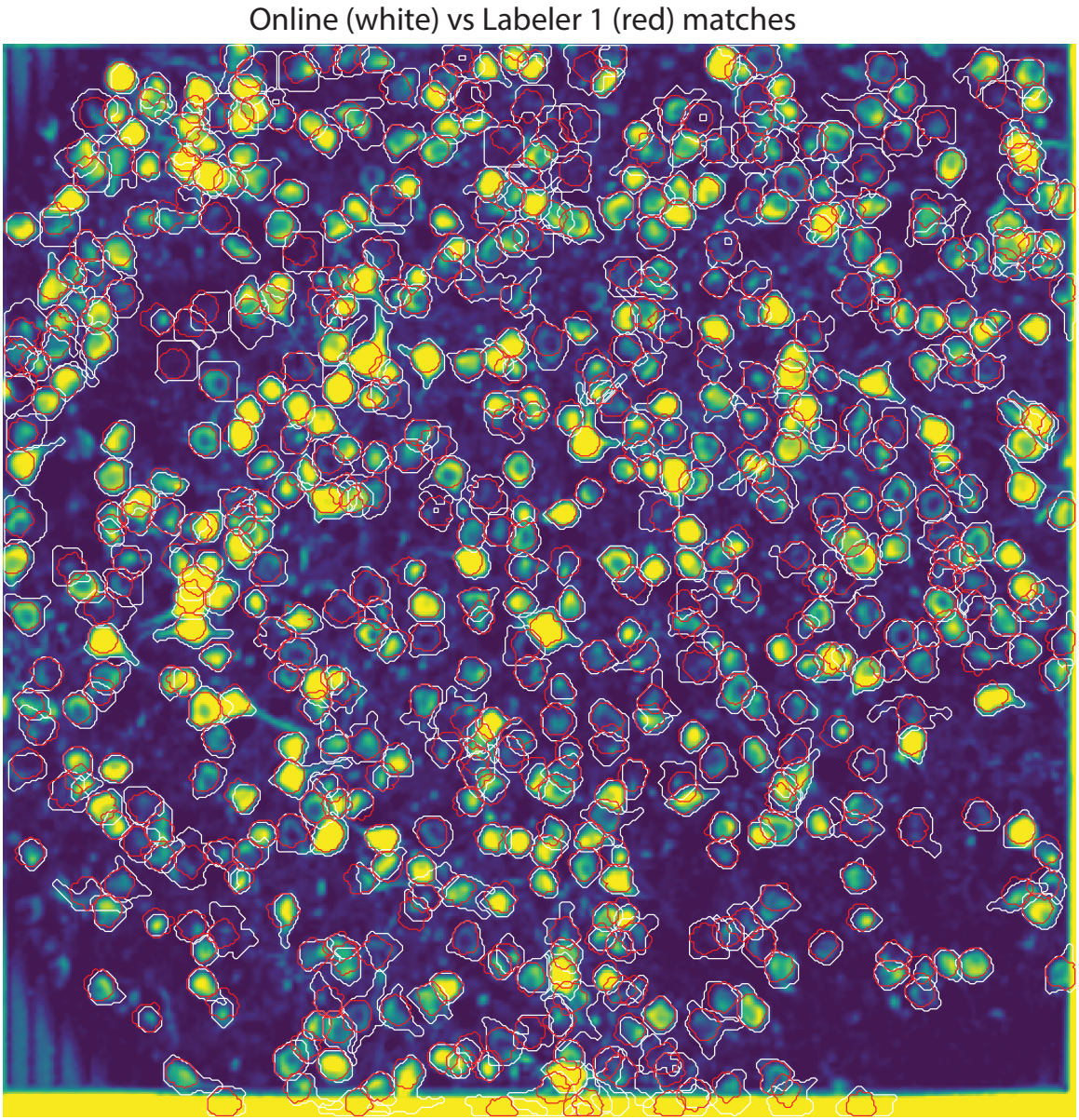

Figure S9: Matches between OnACID (white) and Labeler 1 (red).

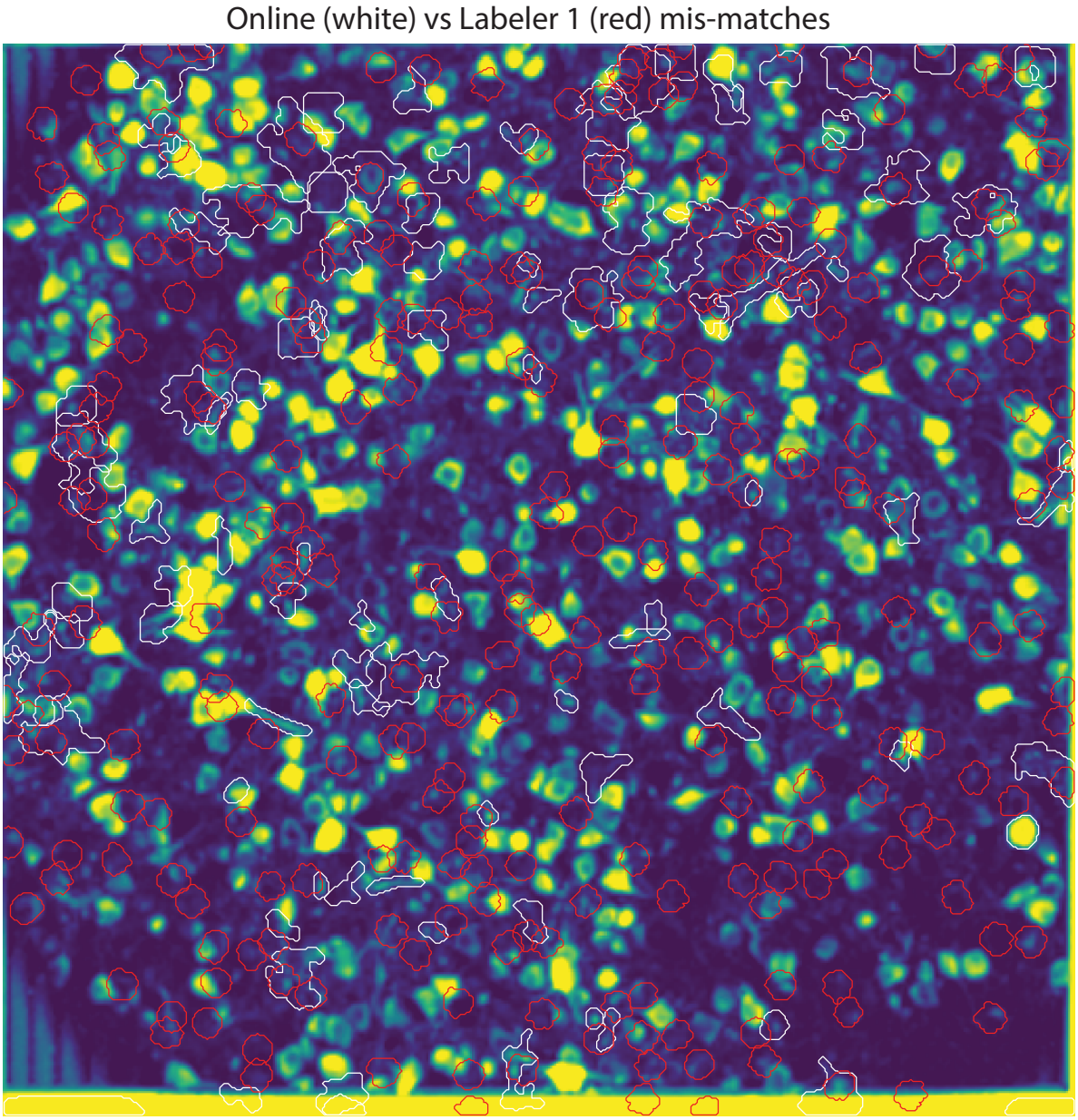

Figure S10: Mismatches between OnACID (white) and Labeler 1 (red).

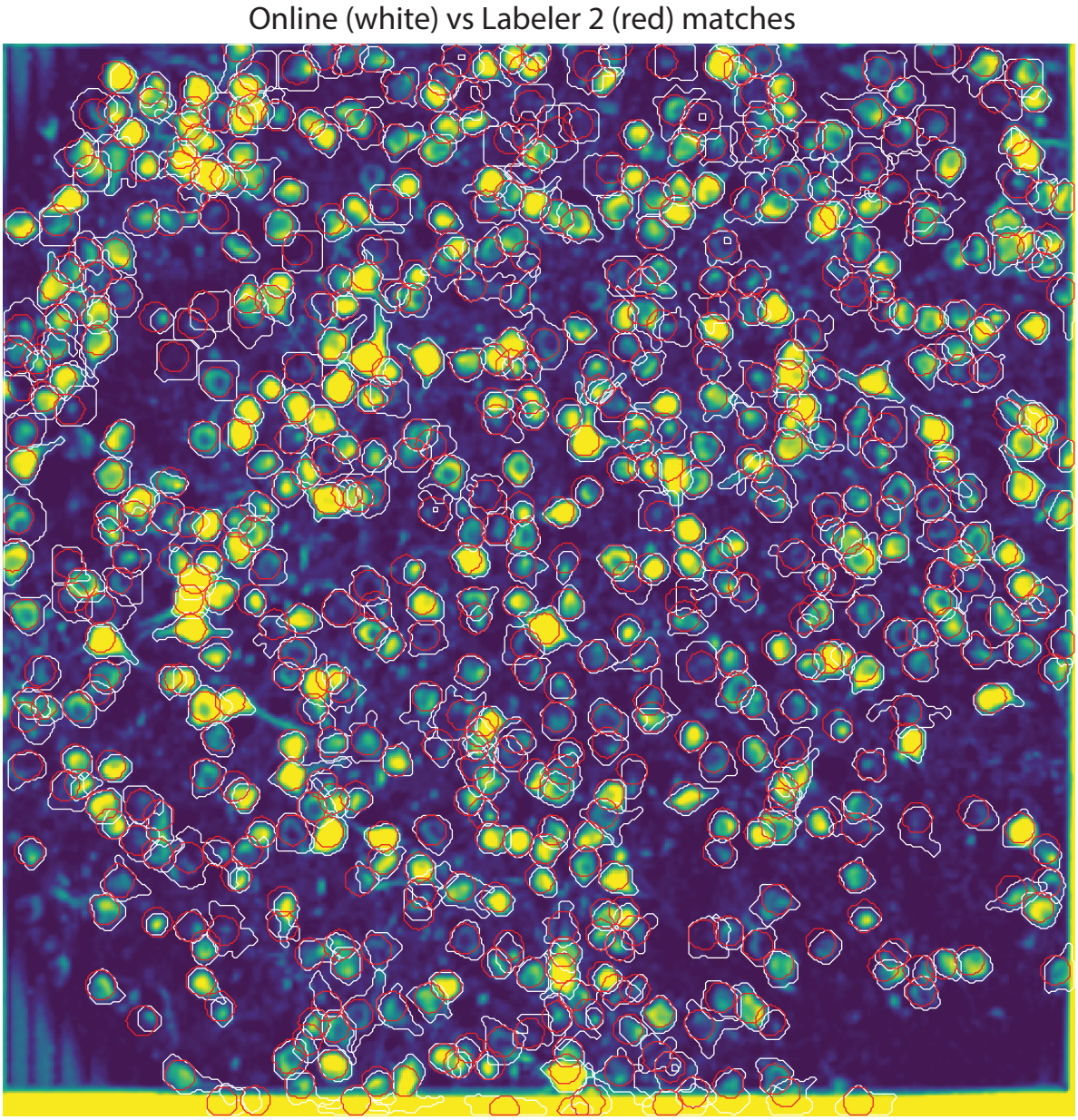

Figure S11: Matches between OnACID (white) and Labeler 2 (red).

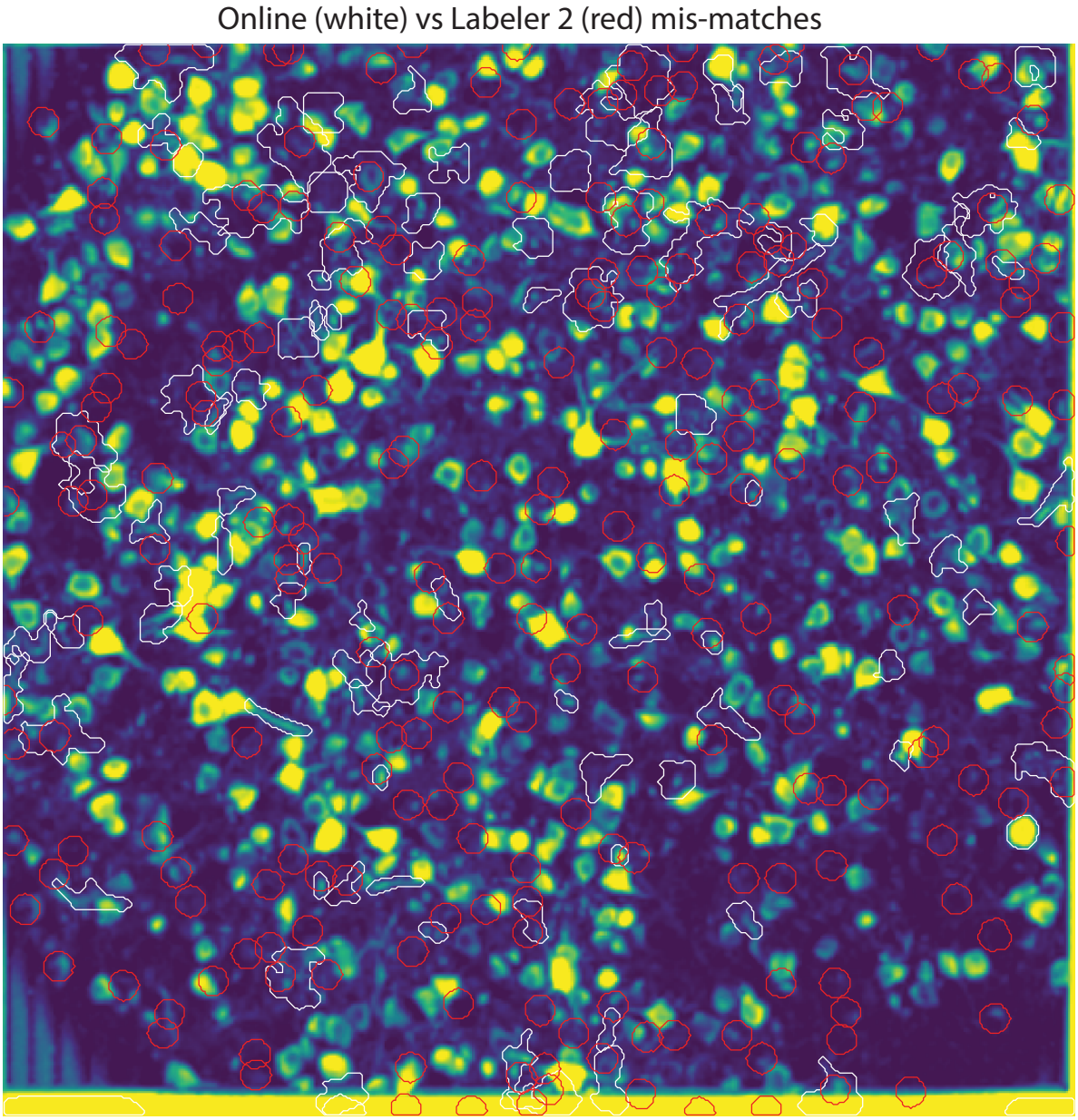

Figure S12: Mismatches between OnACID (white) and Labeler 2 (red).

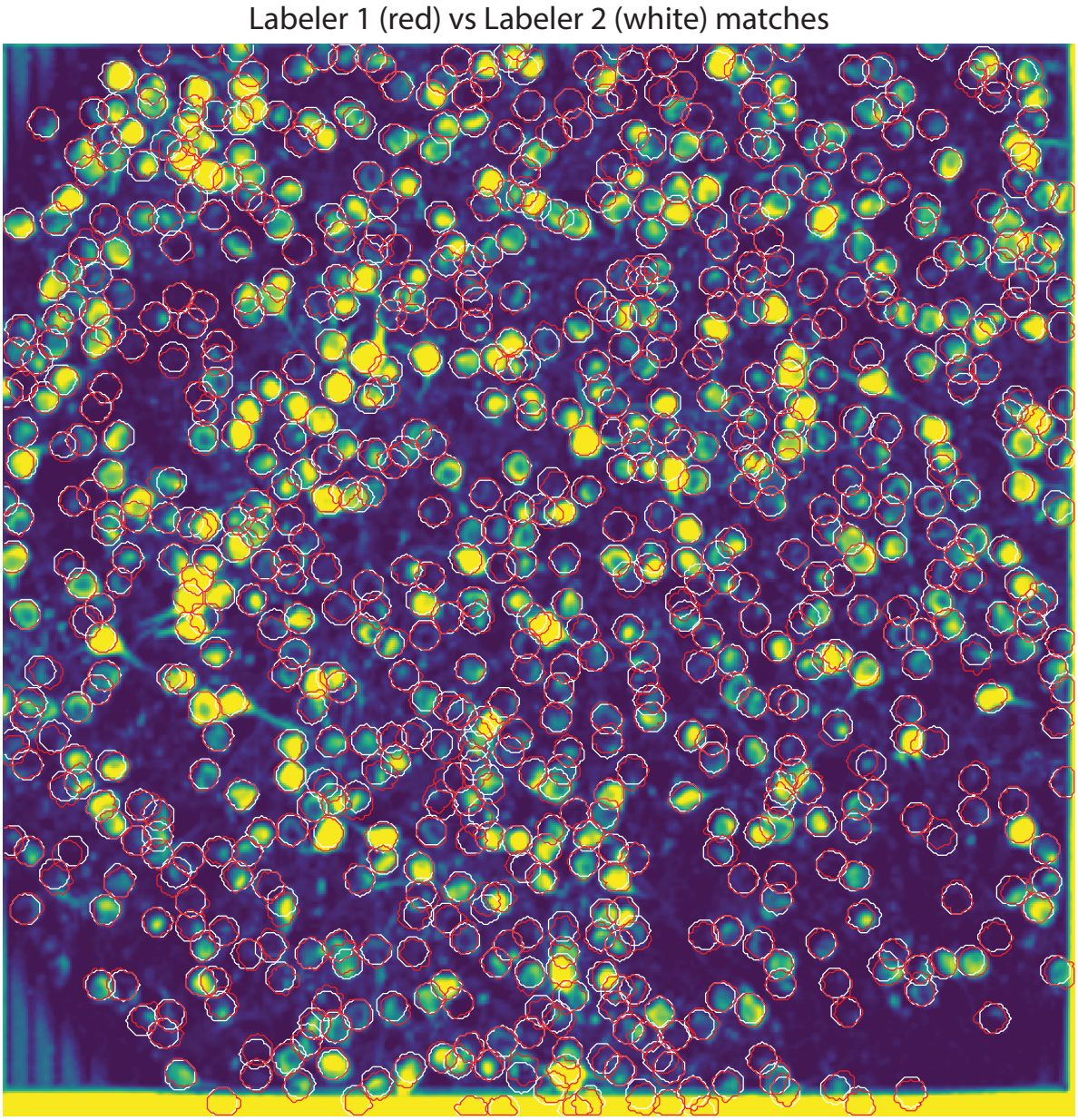

Figure S13: Matches between Labeler 1 (red) and Labeler 2 (white).

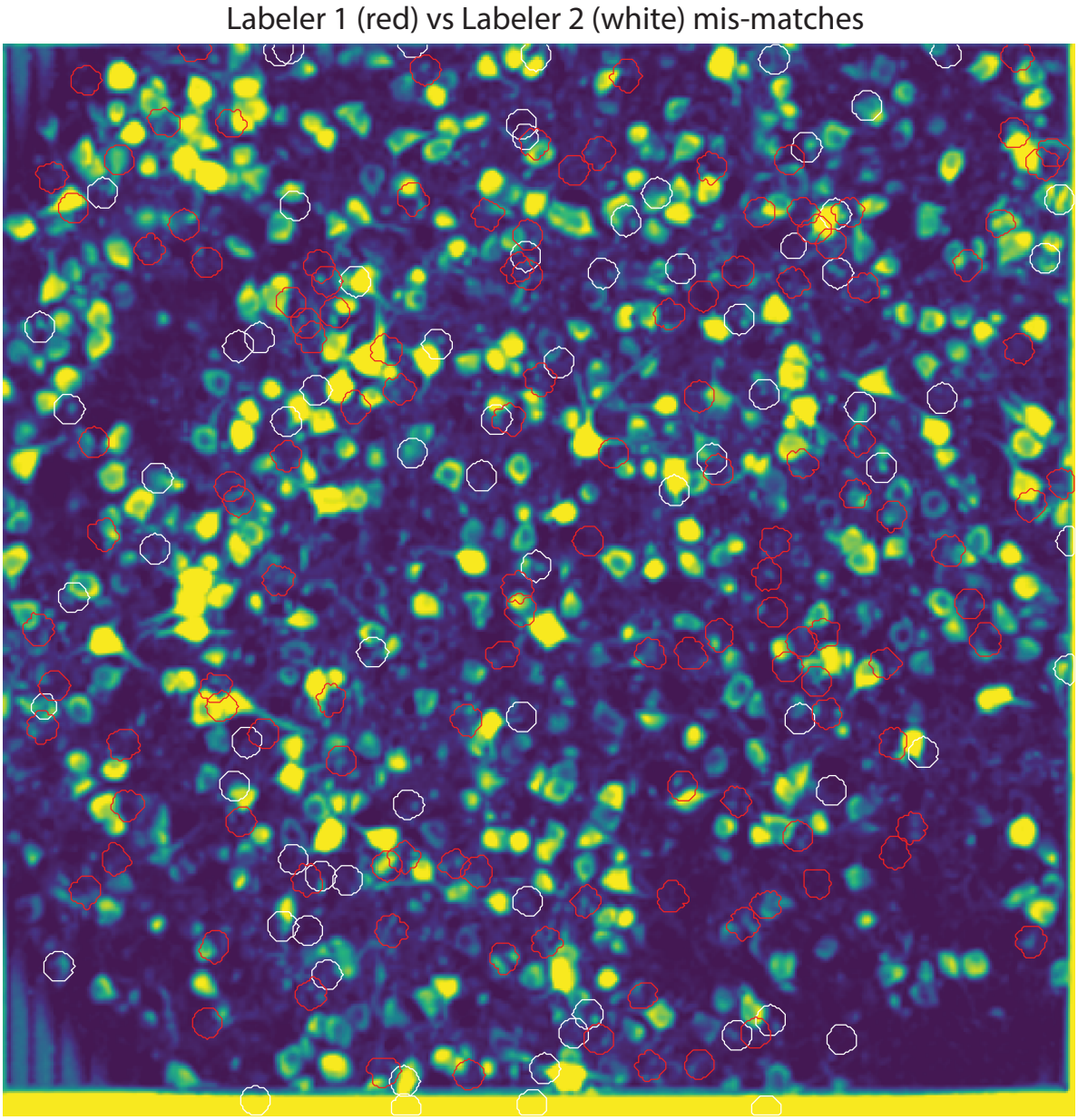

Figure S14: Mismatches between Labeler 1 (red) and Labeler 2 (white).

References

[1] J. Gauthier and D.W. Tank. A subset of ca1 and subiculum neurons selectively encode rewarded locations. In *Computational and Systems Neuroscience Meeting, Cosyne*, 2016.

[2] Sue Ann Koay, Ben Engelhard, Lucas Pinto, Ben Deverett, Stephan Thiberge, Carlos Brody, and David Tank. Neural dynamics in a mouse navigation and accumulation of visual evidence task. In *Society for Neuroscience*, number 739.07, 2016.

[3] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(Jan):19–60, 2010.