In the format provided by the authors and unedited.

Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs

Tristan Bepler^{1,2}, Andrew Morin^{2,3}, Micah Rapp^{4,5}, Julia Brasch^{4,5}, Lawrence Shapiro⁴, Alex J. Noble^{5*} and Bonnie Berger^{2,3*}

¹Computational and Systems Biology, MIT, Cambridge, MA, USA. ²Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA. ³Department of Mathematics, MIT, Cambridge, MA, USA. ⁴Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. ⁵National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA. *e-mail: anoble@nysbc.org; bab@mit.edu



Normalization methods on EMPIAR-10261.

Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10261 micrographs downsampled 4x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. (a) Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. Affine normalized micrographs are washed out when there are dark grid regions present in the micrographs. (b) Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. GMM normalization correctly centers the pixel intensities around the high intensity peak. Results are consistent across >20 micrographs examined.



Normalization methods on EMPIAR-10234.

Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10234 micrographs downsampled 8x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. (a) Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. The left-most micrograph contains light carbon grid that is correctly removed by GMM normalization. (b) Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. Results are consistent across >20 micrographs examined.



Normalization methods on EMPIAR-10096.

Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10096 micrographs downsampled 8x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. (a) Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. No grid is present in these micrographs, so affine and GMM normalization given nearly identical results. (b) Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. Results are consistent across >20 micrographs examined.



Toll receptor example micrograph with picks.

Example micrograph for the Toll receptor dataset with the unlabeled micrograph **(top left)**, training picks **(top right)**, DoG Picker picks **(middle left)**, Topaz picks **(middle right)**, FindEM picks **(bottom left)**, and crYOLO picks **(bottom right)**. This result is consistent over >100 micrographs examined.



Toll receptor 3DFSC curves and crYOLO reconstruction.

Toll receptor 3DFSC plots and 3D reconstruction of the Toll receptor using 131,300 particles picked using crYOLO. (a) 3DFSC plots for Toll receptor structures solved using particles from Topaz, DoG, Template, and crYOLO. (b) Density map of the crYOLO structure. The crYOLO structure reaches a resolution of 6.83 Å at FSC0.143 with a sphericity of 0.734.



Sparse picking in aggregates: Topaz vs crYOLO.

Example micrograph with a large amount of aggregate but still a number of real particles, comparing Topaz with crYOLO after following crYOLO's sparse picking procedure. Topaz **(top left)** picks 40 particles while avoiding aggregation and ice contamination at the default threshold of 0.0. crYOLO picks 1 particle at its default threshold of 0.3. Lowering the threshold to 0.07 **(top right)** yields 26 particles with a small cluster of picks in the aggregation near the bottom right hand corner of the image. Further decreasing the threshold to 0.06 **(bottom left)** yields even more particles, but now it is clear the network is picking a large number of pixels within the aggregation. At a threshold of 0.05 **(bottom right)**, crYOLO is no longer able to avoid aggregation or ice contamination. Note that the particle picking threshold increment in crYOLO is 0.01. We found these results to be consistent across >100 micrographs examined.



Sparse picking with low SNR: Topaz vs. crYOLO.

Example micrograph with real particles but thicker ice and thus lower contrast, comparing Topaz with crYOLO after following crYOLO's sparse picking procedure. Topaz (**top left**) picks 127 particles at the default threshold of 0.0. crYOLO picks 0 particle at its default threshold of 0.3. Lowering the threshold to 0.07 (**top right**) yields 37 particles. Further decreasing the threshold to 0.06 (**bottom left**) yields more particles, but the network is still missing real particles while starting to select some background pixels. At a threshold of 0.05 (**bottom right**), there are clear artifacts at the edges of the micrograph. We found these results to be consistent across >100 micrographs examined.



T20S proteasome example micrographs.

Two example micrographs for the T20S proteasome dataset (EMPIAR-10025) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. We found this result to be consistent across >100 micrographs examined.



80S ribosome example micrographs.

Two example micrographs for the 80S ribosome dataset (EMPIAR-10028) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. We found this result to be consistent across >100 micrographs examined.





Rabbit muscle aldolase example micrographs.

Two example micrographs for the rabbit muscle aldolase dataset (EMPIAR-10215) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. In this aldolase dataset, the particles are tightly packed, but Topaz correctly identifies and centers the particles. We found this result to be consistent across >100 micrographs examined.

score > 6	6594 ptcls	5673 ptcls	4386 ptcls	3013 ptcls 5.8 A 1 ess	2728 ptcls	2235 ptcls 7.5 A1 ess	1753 ptcls.	1676 ptc/s 7.2 A1 ess	1456 ptcls 8.9 A.2 ess	1085 ptdis 426.0 A 2 ess
score > 5.5	11120 ptels 5.8 A 1 ess	9839 prots 5.9 A 1 ets	8003 prois 5.9 A 1 ess	5046 ptcts	4985 ptcls 7.2 A 1 ess	4411 ptcis	3347 ptcis	1954 ptcis 6.6 A 2 ess	1639 ptcls 6.9 A 2 ess	1064 ptds 16.8 A 2 exs
score > 5	16182 ptcls 5.8 A 1 ess	14369 ptcis	13194 ptcls 5.7 A 1 ess	8263 ptcls 9,8 A 1 ess	7065 ptcls 3.4 A 1 ess	6970 ptcls	-6362 ptcis: 5.8 A 1 ess	2596 ptcls 426.0 A 2 ess	2540 ptcis 7.7 A 2 ess	2469 ptcls 9.1 A 2 ess
score > 4	28218 ptcls	22624 ptcls 5.7 A 1 ess	21306 ptcls 5.6, A 1 ess	20139 ptcls 426.0 A 1 ess	16328 ptcls 6.0 A 1 ess	16177 ptcls	15936 ptcls	9743 ptcis 8.9 A 2 ess	8160 ptcls 12.6 A 2 ess	7756 ptcls 9.7 A 2 ess
score > 3	48652 ptcis 3.4 A 1 ess	42884 ptcls 5.7 A 1 ess	40071 ptcis 5.7 A 1 ess	39273 ptcls	35691 ptcls 727 3.4 A 1 ess	28950 ptcls 5,8 A 1 ess	22511 ptcls 5.7 A 1 ess	12787 ptcls 3.4 A 2 ess	11709 ptcls 5.7 A 2 ess	11658 ptcls 8.4 A2 ess
score > 2	89856 ptchs 3.4 A 1 ess	71601 ptcls 5.7 A 1 ess	63229 ptcls	55577 ptcls 5.7 A 1 ess	45194 ptchs	30565 ptels 3.4 A 2 ess	29272 ptcls 7.9 A 2 ess	24027 ptcls 3.4 A 2 ess	24007 ptels 3.4.A.2 ess	15051 ptcls 7.6 A 2 ess
score > 1	128874 ptcls	121902 ptcls 3.4 A 1 ess	76261 ptdis 5.8 A 1 ess	74829 ptcis	50903 ptcls 3.4 A 1 ess	37763 ptcls 3.4 A 2 ess	34922 ptcls 3.4 A/2 ess	31795 ptels 3.4 A 2 ess	28304 ptcis -3.4 A 2 ess	22356 ptcls 3.4 A 2 ess
score > 0	116484 ptcls 5.8 A 1 ess	115125 pteis 5.7 A1 ess	105974 ptcis 1.4 A 1 ess	83989 ptcls 3.4 A 1 ess	80599 ptcls 3.4 A 1.ess	74541 ptcls 3.4 A 1 ess	52397 ptcls 7.5 A 2:ess	42857 ptcls 3.4 A/2 ess	42782 ptcls 9.0 A 2 ess	29188 ptcls 3.4 A 2 ess
score > -1	163256 ptcls	160316 ptcls	117510 ptcis 5.7 A 1 ess	96740 ptcls 3.4 A 1 ess	56603 ptcls 3.4 A 2 ess	35847 ptcls. 812 A 2 ess	53578 ptcls 3.4 A 2 ess	50873 ptcls	46016 ptcls 3.4 A 2 ess	41429 ptcls 3,4 A 2 ess
score > -2	220111 ptcls 3,4 A*1 ess	3.4 A3 ess	121487 ptcis	93448 ptcis 3.4 A 1 ess	73661 ptcls 3.4 A 2 ess	72936 ptcls 3.4 A 2 ess	65842.ptcls	54452 ptcls 511 324 A 2 efs	45708 ptcls 5.8 A 2 ess	41261 ptcts 9.8 A 2 ess

EMPIAR-10215 2D class averages.

2D class averages of Topaz particles with decreasing score threshold for the aldolase dataset. Classes identified as false positives for quantification in Figure 5 are indicated by orange boxes.

score > 3	2458 ptcls 10.2 A 1 ess	1717 ptck 10.3 A 1 ess	1552 ptcls 10.5 A 3 ess	1429 ptcls 20.0 A 1 ess	1365 ptds	1339 ptck 14.2 & 1 ess	1265 ptcls	899 ptcls 13.0 A 1 ess	786 ptcls 13.2 A 1 ess	704 ptcls 15.7 A 1 ess
score > 2.5	4425 ptcls 9.1 A 1 ess	4216 ptcls 9,3 A 1 ess	40.32 ptcls	3650 ptck 8.3 A 1 ess	3604 ptcls 20 B A 1 ess	3435 ptcls 8.8 A 1 ess	2874 ptcls 9.7 A 1 ess	2757 ptc.s 13.0 A 1 ess	1966 ptcls 8.1.A.1.ess	1801 pitels
score > 2	10566 ptc/s	8504 ptcls	7033 ptcls	5983 ptcb	5880 ptds	5695 ptcls	5649 ptds	4359 ptc s	3970 pttls	3424 ptcls
	8.6 A 1 ess	8.5 A 1 ess	7.5 A 1 ess	11.2 A 1 ess	7.3 A 1 ess	9.8 A 1 ess	8.1 A 1 ess	9.1 A 1 ess	9.4 A 1 ess	8.1 A 1 ess
score > 1.5	13936 ptcis	12694 ptcls	11571 ptc/s	9957 ptck:	9145 ptd5	6836 ptcls	7101 ptcls	0659 ptc s	6468 ptcls	5277 ptcls
	7.7 A 1 ess	8.1 A 1 ess	6.6 A 1 ess	7.5 All ess	104 A 1 ess	7.1 A 1 ess	11.6 A 1 ess	12.2 A 1 ess	8.3 A 1 ens	8.4 All ess
score > 1	16824 ptc/s	16605 ptcb	15586 ptcis	14863 ptcls	13863 ptcls	12030 ptc/s	9688 ptch	7199 ptc's	5717 ptcls	6538 ptcls
	7.4 A 1 ess	6.5 A 1 ess	10.3 A 1 ess	7.2 A 1 ess	71A1 ess	0.3 A 1 ess	9,9 A 1 ess	6.6 A 1 ess	6 9 A 1 ess	8.8 A 1 ess
score > 0	20781 ptcls	20092 ptdis	16044 ptc/s	15948 ptris	14448 ptds	13965 ptcis	13525 ptcfs	11727 ptcls	11285 ptds	9792 ptcfs
	7.1 A 1 ess	7.3 A 1 ess	13.1 A 1 ess	8.1 A 1 ess	7.0.4 1 ess	7.1 A 1 ess	6.0.A 1-ess	19.1 A 1 ess	9.7 A 1 ess	6.6 A 1 ess
score > -1	22799 ptcls 6.0 A 1 ess	21485 ptcls 10.6 A 1 ess	19777 ptds 7.2 A 1 ess	19407 prois 7.2 A 1 ess	16100 ptcls	15697 ptcls 10.6 A 1 ess	14307 ptcls 9.1 A 1 ess	13899 ptck 11.9 A 1 ess	13590 ptcls 7.0 A 1 ess	7642 ptcls 24 9 A 1 was
score > -2	25082 ptc/s	23095 ptcls	22043 ptcb	20587 ptc/s	20879 ptcis	18240 ptcb	16346 pick	13789 ptds	12045 ptds	8314 ptc/s
	6.4. A 1 ess	7.2 Aless	7.2 A 1 ess	7.5.4 1 ess	13 / Alless	8.6.A.1.ess	7.6 A 1 ess	7.4 A 1 ess	ZZ.1 A.1 ess	74.5 A 1 ess
score > -3	43461 pteis 8,3 A 1 css	26207 ptcis 7.8 A 1 055	26033 ptels 7,4 A 1 ess	25342 pteis 7.6 A 1 css	18594 ptcls	16119 prois 26.2 % 1 css	15427 picis. 20.3 A 1 cs5	12415 ptds 20 2 A 1 ess	9152 pttis 7.3 A 1 ess	8566 ptc s 19.9 A 1 css
score > -4	30504 ptcls	29314 ptcls	26263 ptds	25706 ptcls	25129 ptcls	22642 ptcls	20412 ptc/s	17639 ptds	16720 stals	14129 ptcis
	10.4 A 1 ess	8.1 A 1 ess	6.4 A 1 ess	9.7 A 1 ess	7.1 A 1 ess	7.3 A 1 ess	20.3 A 1 ess	9.3 A 1 ess	25.2 A 1 ess	21.5 A 1 ess

EMPIAR-10028 2D class averages.

2D class averages of Topaz particles with decreasing score threshold for the 80S ribosome. Classes identified as false positives for quantification in Figure 5 are indicated by orange boxes.



Precision-recall and F1 curves for EMPIAR-10025, EMPIAR-10028, and EMPIAR-10215.

Precision-recall curves and threshold vs precision, recall, and F1 score curves for classifiers trained on the EMPIAR-10025, EMPIAR-10028, and EMPIAR-10215 datasets. Curves were calculated by matching the particles predicted by the Topaz models on the test set micrographs of each dataset to the published particle annotations on those micrographs. We note that the precision and average-precision scores are *underestimates* of the true precision and true average-precision of the Topaz models due to incompleteness of the published particle set.

a)





Supplementary Figure 14

EMPIAR-10096 and EMPIAR-10234 example micrographs with Topaz picks.

Representative micrographs from (a) the EMPIAR-10096 and (b) the EMPIAR-10234 test sets. For EMPIAR-10096, curated particles from EMPIAR (blue) and particles predicted by Topaz (red) are circled. For EMPIAR-10234, manually selected (blue) and predicted (red) particles are circled. Topaz avoids ice chunks, particles in proximity to the edge of the hole, and particles on carbon and correctly identifies many particles missing from the manually labeled/curated particles sets. We found this result to be consistent over >20 micrographs examined.



Sensitivity to π of GE-binomial.

Sensitivity of GE-binomial objective function to the setting of π . For EMPIAR-10096 and the EMPIAR-10234 datasets we report average-precision scores for classifiers trained with 100 and 1,000 labeled particles and values of π varying from 0.5x to 1.5x the values reported in table 1. We report the mean and standard deviation of 10 runs.