AN JU, YANGKUN ZHANG, TIANXIAO SHEN

PROJECT 1: AUTOMATED PARTICLE PICKING IN CRYO-EM

1.1 ALGORITHM OVERVIEW

- Two-step procedure:
  - 1. scan the whole micrograph by a sliding window and judge whether it contains a particle or not
  - 2. for all the windows containing a particle, compute their center and merge the ones nearby

1.2 WINDOW CLASSIFICATION

- Image classification
- Convolutional Neural Network (CNN)
  - 5 layers: 3 convoluntional, 2 fully-connected
  - Rectified Linear Units (ReLUs)
  - Stochastic Gradient Descent (SGD) with Nesterov momentum

1.2 WINDOW CLASSIFICATION

- Preprocessing
  - rescale the values so that they range between 0 and 1
- Skewed classes?
  - use all golden particles as positive cases
  - randomly sample negative cases of the same amount

1.3 MERGE NEIGHBORING WINDOWS

- $P = \{p_1, p_2, \ldots p_m\}$ particles picked so far
- p center of an identified window
- $p_i \leftarrow \text{nearestNeighbor}(p, P)$

if $\text{dist}(p, p_i) < d$

$p_i \leftarrow centerOfMass(p, p_i)$

else

$P \leftarrow P \bigcup \{p\}$

1.3 MERGE NEIGHBORING WINDOWS

- Define the confidence of p as

$$c(p) = \sum_{p_i \in p} \Pr[w_i \text{ contains a particle}] - 0.5$$

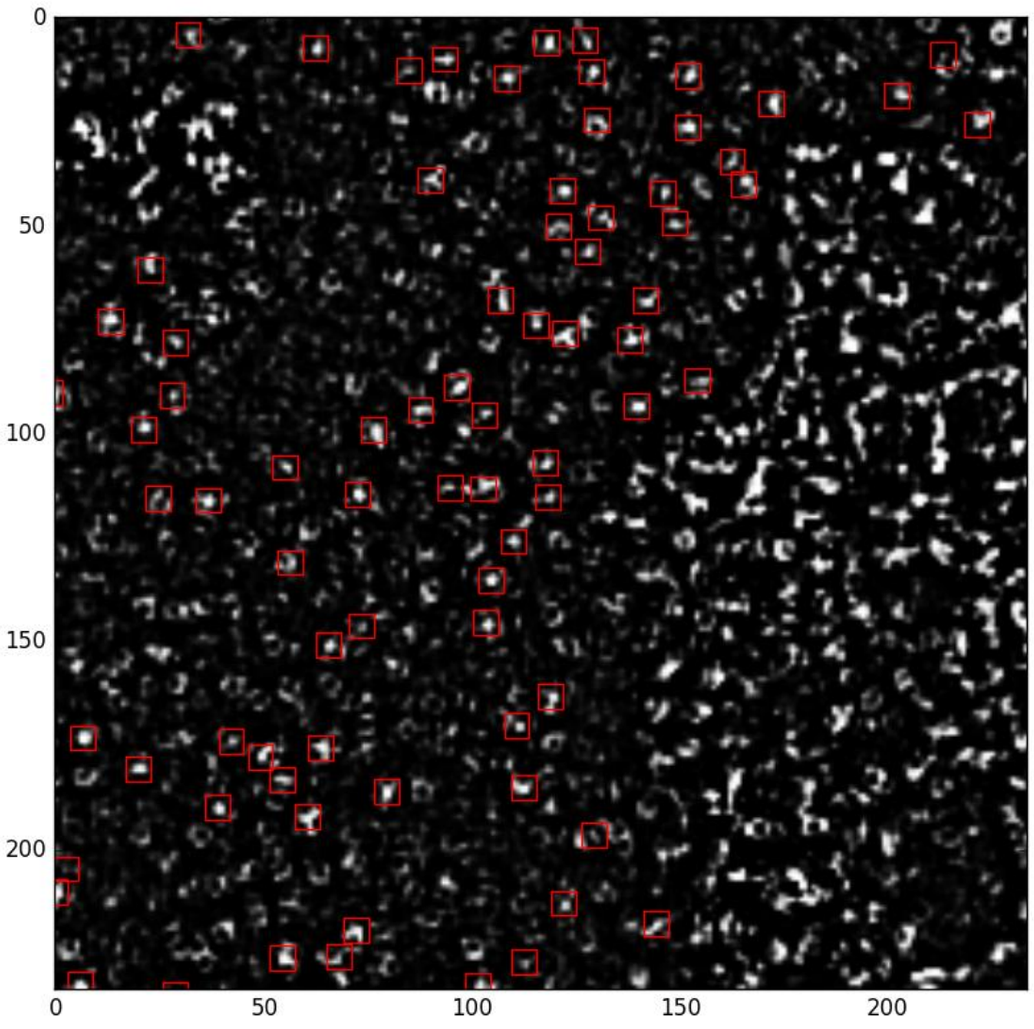
- If $c(p) > C$, report it as a final predicted particle

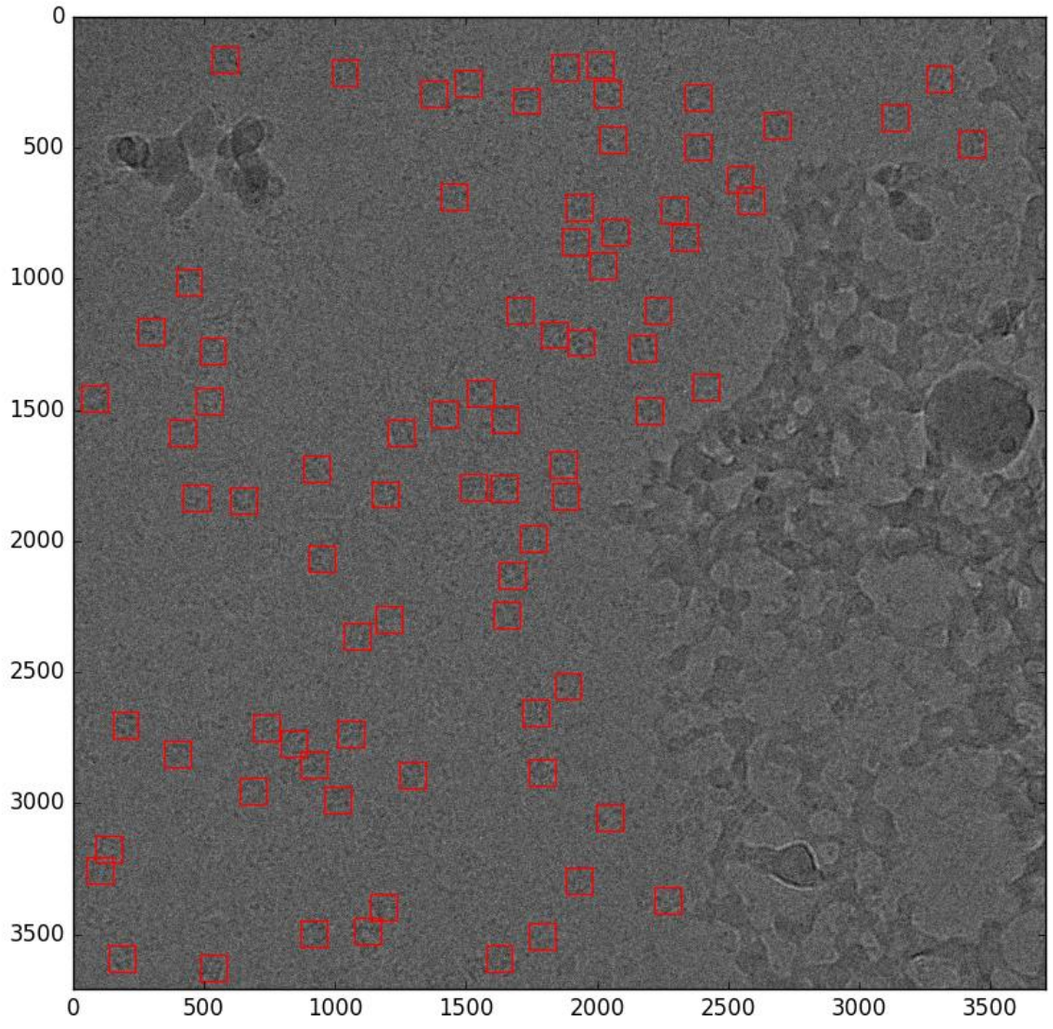
2.1 RESULTS

- 32 micrographs for training, 8 for validation
  - classification accuracy: 93.54%
- Average precision, recall, and F-measure on test data: 59.33%, 74.26%, 64.46%

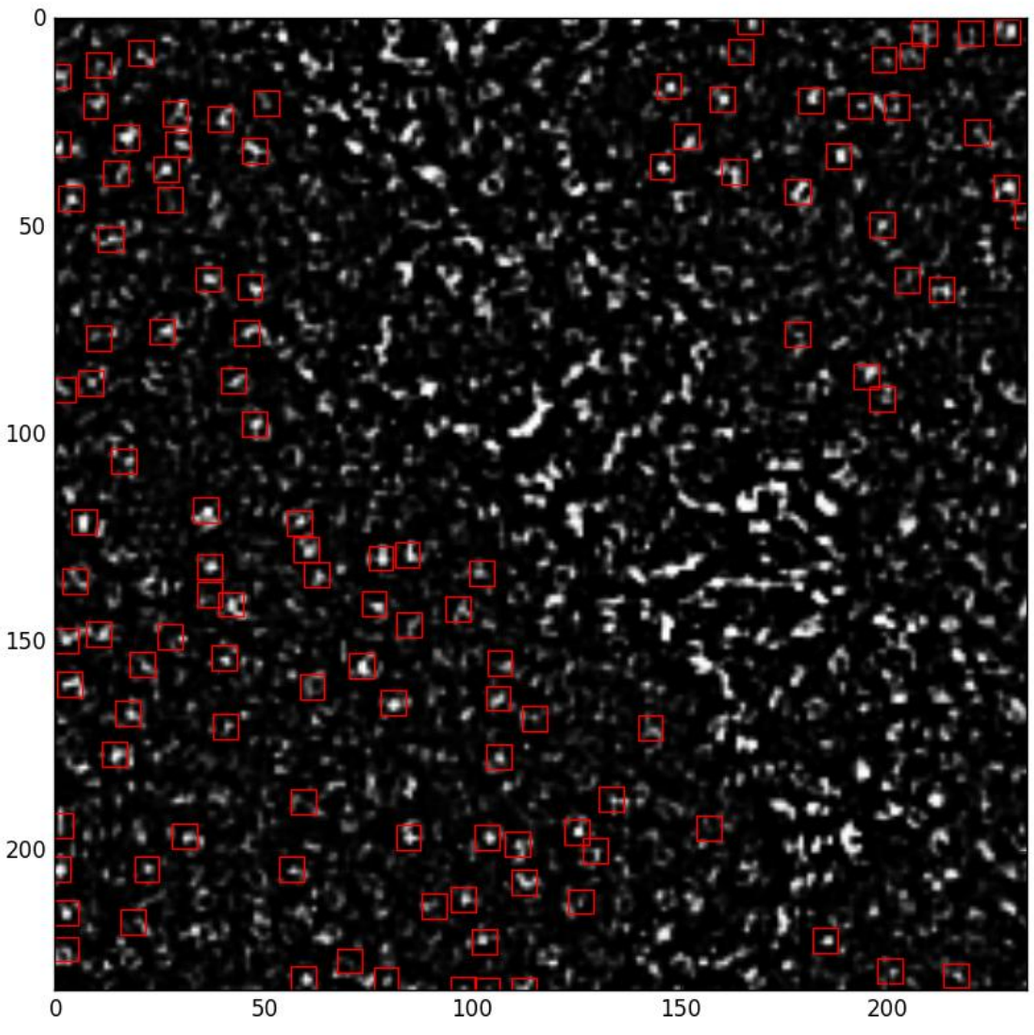
Index	1	2	3	4	5	6	7	8	9	10
Precision	64.96	57.19	67.77	53.33	49.47	70.00	72.00	66.94	60.53	23.25
Recall	83.33	96.62	86.42	71.15	86.96	56.44	64.29	42.94	85.21	70.89
F-measure	73.01	71.85	75.97	60.97	63.06	62.49	67.92	52.32	70.78	35.01
Index	11	12	13	14	15	16	17	18	19	20
Precision	58.30	69.48	58.42	57.84	59.80	74.74	71.74	23.59	58.21	69.09
Recall	87.65	87.76	84.66	73.76	85.55	69.89	54.34	38.60	84.38	74.47
F-measure	70.03	77.56	69.13	64.84	70.39	72.24	61.84	29.28	68.89	71.68

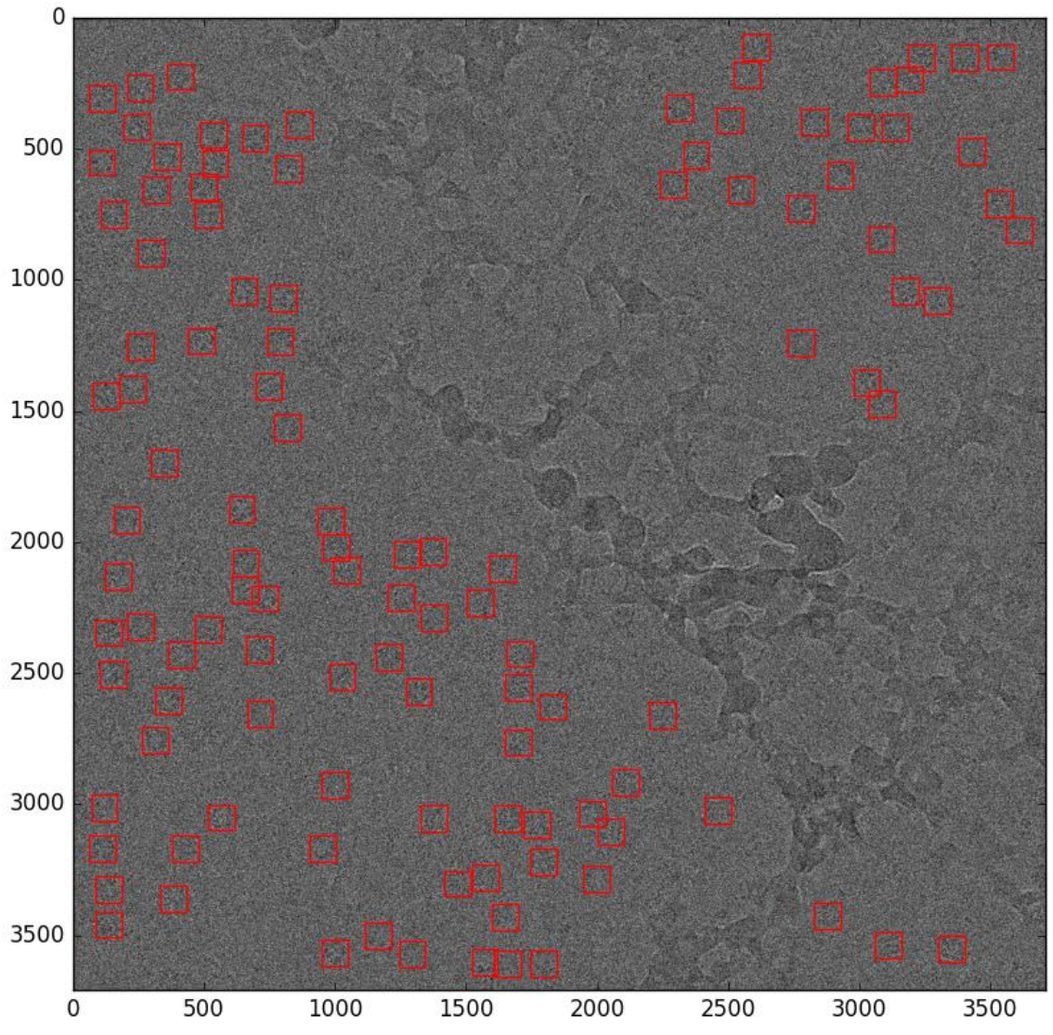
2.2 CONTAMINANTS





2.2 CONTAMINANTS



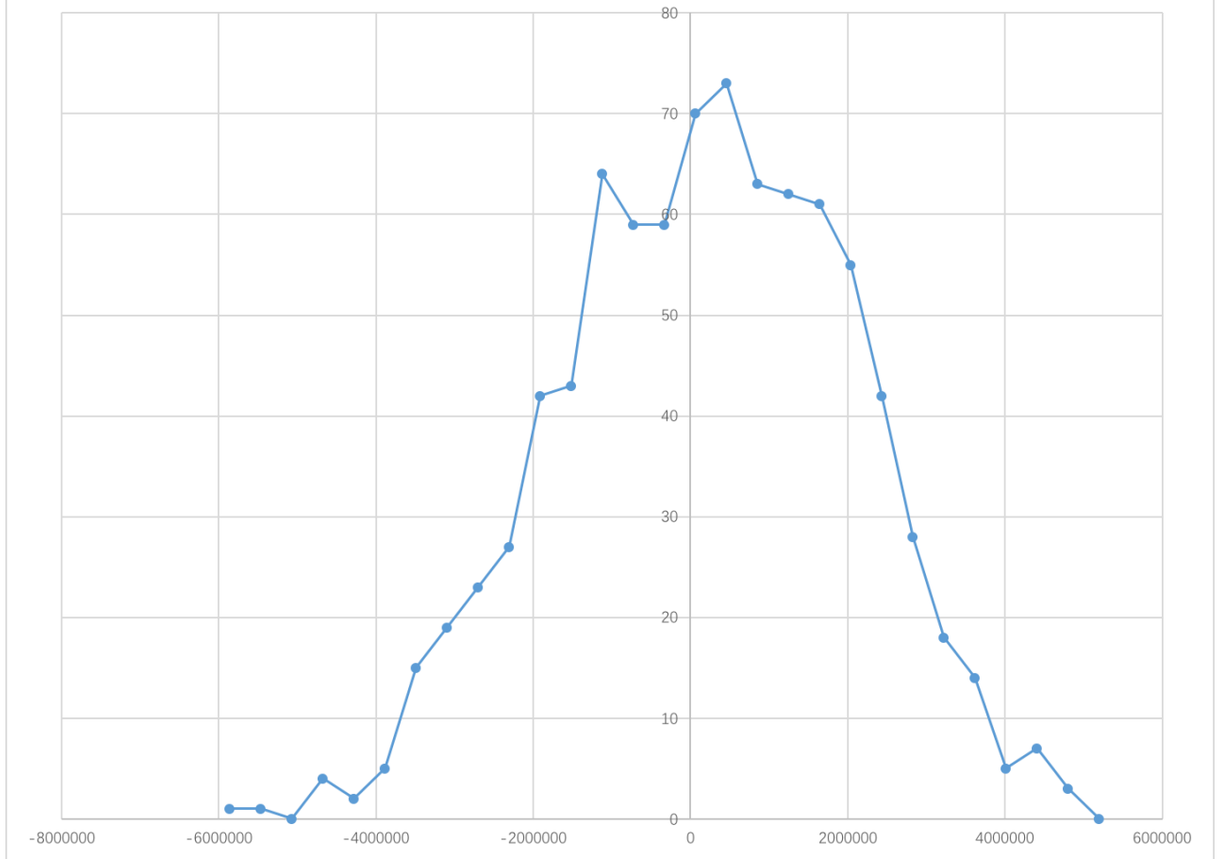


2.2 CONTAMINANTS

- Principal Component Analysis (PCA) over the power spectra of extracted windows, and eliminate outliers according to their distribution (Langlois et al., 2014)

2.2 CONTAMINANTS

- Principal Component Analysis (PCA) over the power spectra of extracted windows, and eliminate outliers according to their distribution (Langlois et al., 2014)

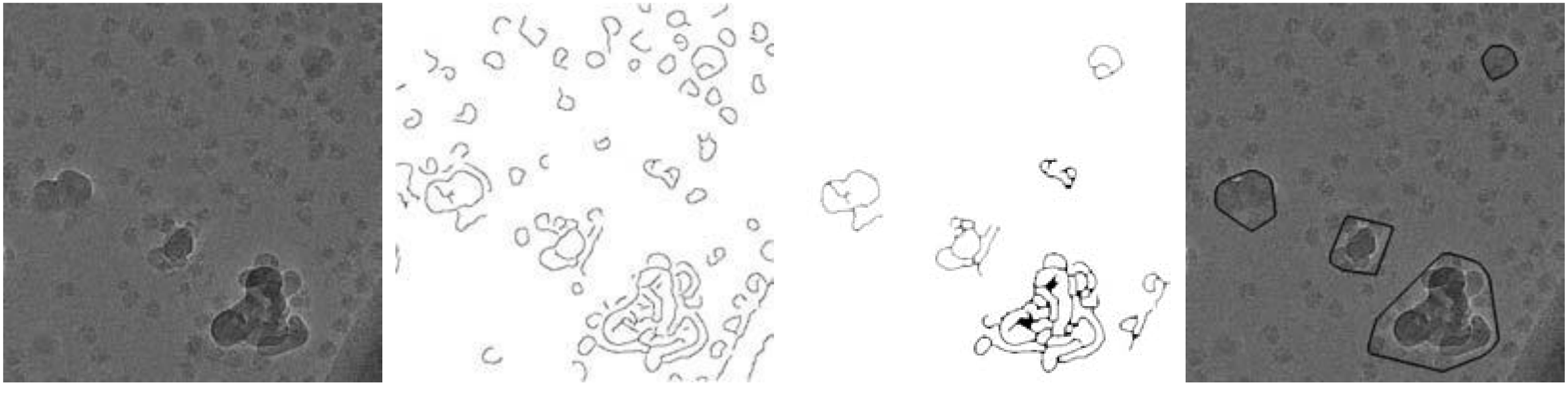


2.2 CONTAMINANTS

- Principal Component Analysis (PCA) over the power spectra of extracted windows, and eliminate outliers according to their distribution (Langlois et al., 2014)

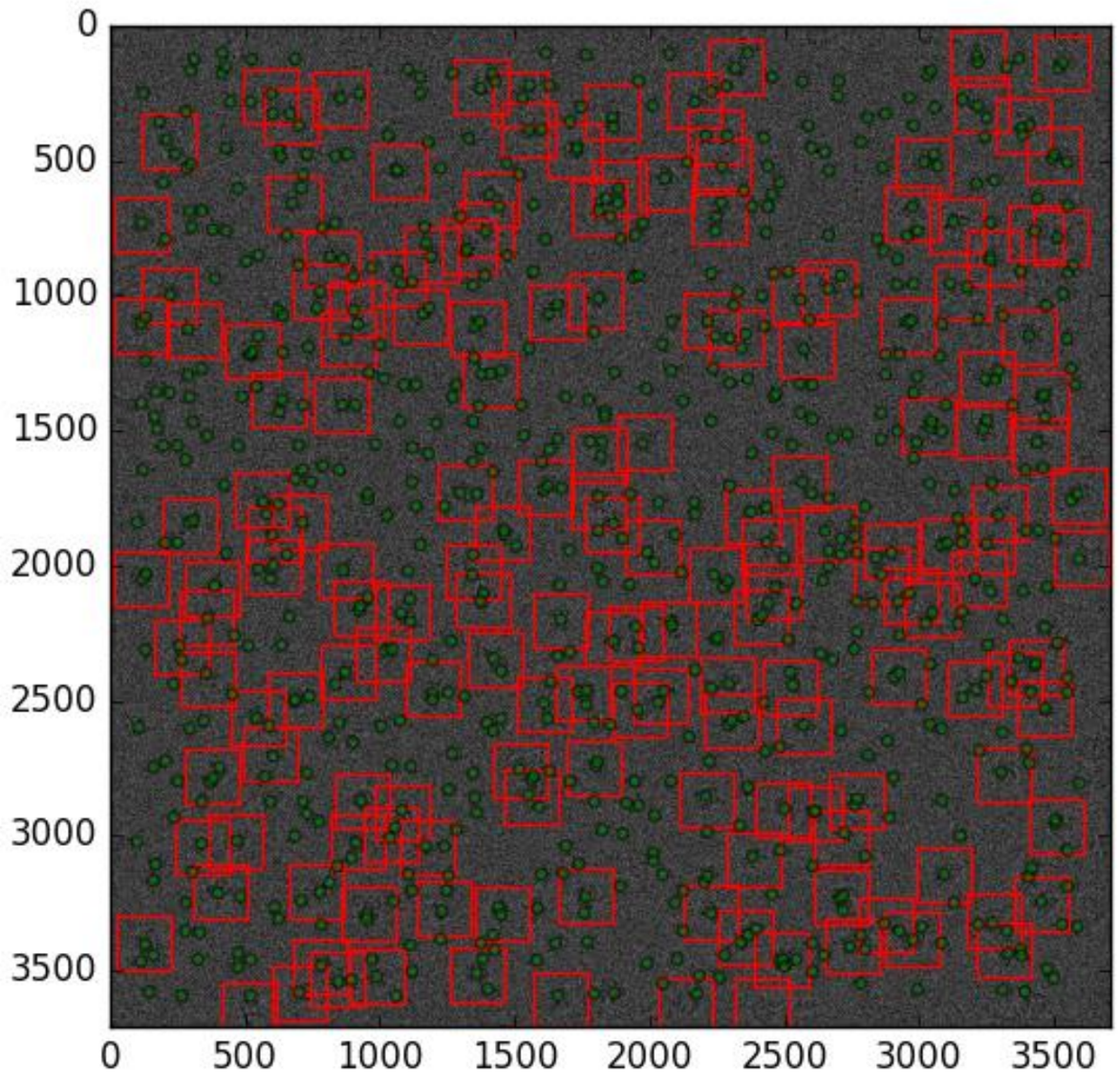
2.2 CONTAMINANTS

- Edge detection, connected components labeling and convex hull computation (Zhu et al., 2004)



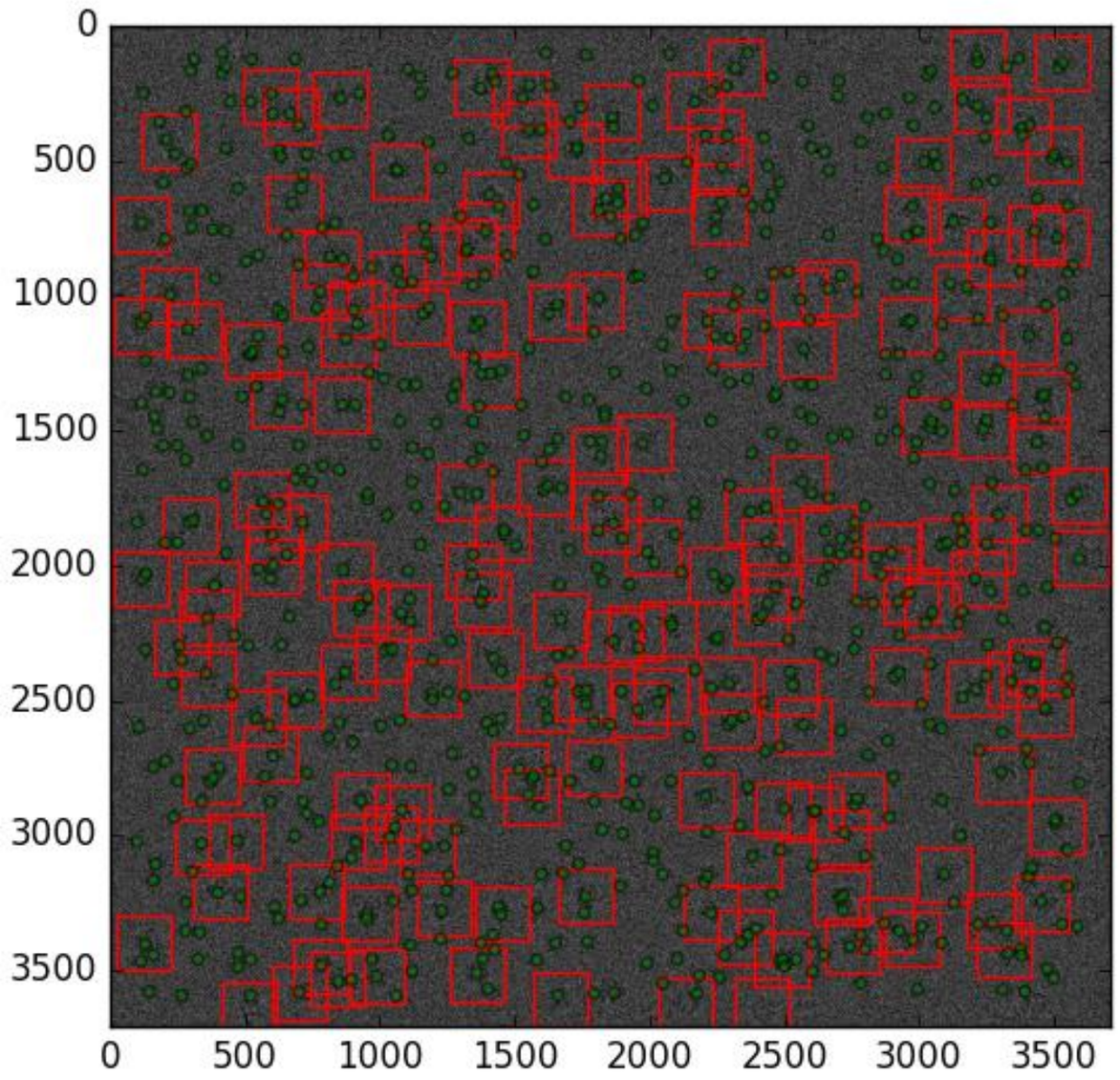
2.3 BOUNDARY

- particles on the boundary whose window lies out of scope



2.3 BOUNDARY

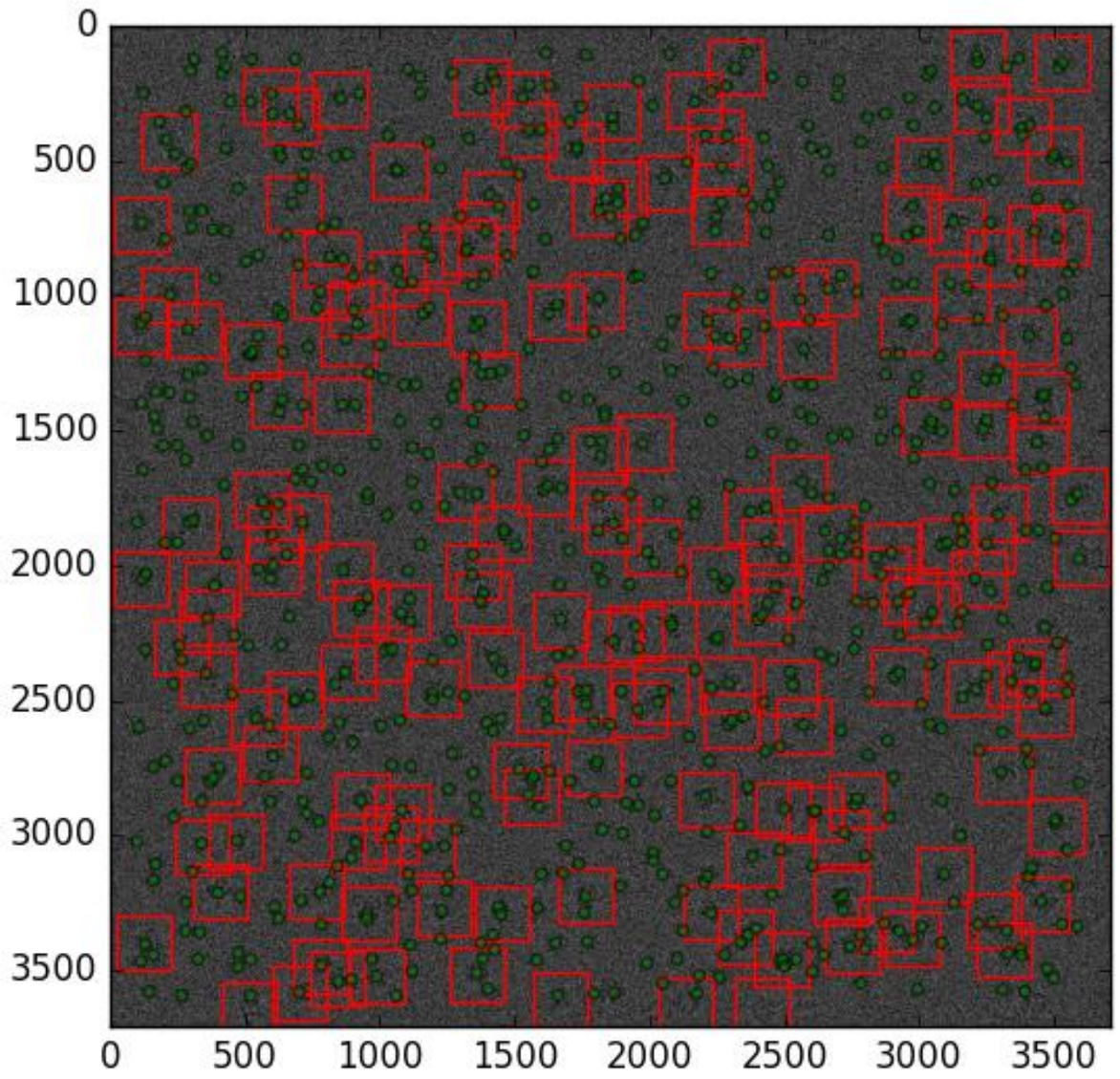
- particles on the boundary whose window lies out of scope
- padding
  - mean
  - Gaussian noise
  - folding



2.3 BOUNDARY

- particles on the boundary whose window lies out of scope
- padding
  - mean
  - Gaussian noise
  - folding

none of them works ☹



2.4 SPEED

- Efficient GPU implementation
- 20 min for training CNN
- 1 min for testing a micrograph

Acknowledgments

Zhipeng Jia

References

R. Langlois, J. Pallesen, J. T. Ash, D. N. Ho, J. L. Rubinstein, and J. Frank. Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. Journal of structural biology, 186(1):1-7, 2014.

Y. Zhu, B. Carragher, and C. S. Potter. Contaminant detection: improving template matching based particle selection for cryoelectron microscopy. In Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, pages 1071-1074. IEEE, 2004.