

比較不同不確定性度量方法於基於資料池的主動學習框架下對低溫電子顯微鏡粒子 挑選的表現

Comparison of Different Uncertainty Measurement Methods for Pool-based Active Learning in Cryo-EM Particle Picking Performance

研究生： 李祐瑄 Yu-Hsuan Li
指導教授： 鍾思齊 博士 Dr. Szu-Chi Chung

National Sun Yat-sen University
Department of Applied Mathematics

January 20, 2025



Contents

① Introduction

② Literature Review

③ Method

④ Experiment

⑤ Discussion

Introduction

- Background
- Motivation
- Objective

Cryo-electron microscopy

Cryo-EM is an advanced microscopy technique that enables scientists **to study biomolecules at atomic resolution** by freezing specimens, revealing intricate molecular structures with unprecedented clarity [1].

- Key advantages:
 - Samples are studied in **native conditions** without dyes or chemical treatments.
 - Combines imaging with computational analysis for 3D structure determination.
- Impact on biomedical research:
 - **Faster and more accurate** acquisition of critical structural data.
 - For instance, studying viral capsid structures for vaccine design in virology.

Motivation: Importance of Particle picking

Goal of Cryo-EM Image Processing is reconstruct 3D density maps from particles embedded in 2D micrographs [2].

- Impact on 3D reconstruction
 - **Accurate particle picking** improves the quality of the final 3D molecular structure.
 - Selecting clear single particles with diverse viewing angles:
 - Enhances structural accuracy.
 - Prevents missing data or anomalies that could blur or distort the reconstruction.

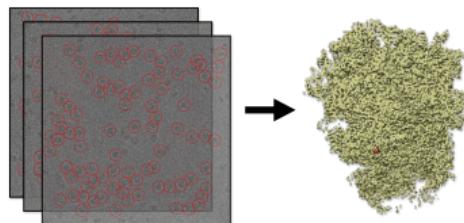


Figure 1-1: Schematic diagram

Motivation: Challenges and Developments

Cryo-EM images differ from general object detection tasks due to **extremely low-SNR** (signal-to-noise ratio), caused by background noise from vitreous ice.

- Manual particle picking: Traditionally time-consuming and error-prone.
- Particle shape variability: Traditional template-matching methods struggle with particles exhibiting diverse shapes or conformations, leading to false positives or missed detections [3].
- Advancements in Deep Learning powered by GPUs [4], enable:
 - Can **adapt to diverse particle shapes** and reduce errors.
 - **Enhance precision and efficiency** in automated particle picking.

Research Objectives

- **Active Learning** in automated particle picking
 - Goal: Reduce manual annotation efforts and enhance model training efficiency.
 - Approach: Target the **most informative** samples for labeling, reducing noise and unnecessary annotation.
- **Comparing uncertainty measurement methods**
 - Goal: **Compare and evaluate the effectiveness of different uncertainty measurements in Active Learning.**
 - Approach: **Develop a framework for a fair comparison of uncertainty methods to optimize the annotation process and improve training accuracy.**

Literature Review

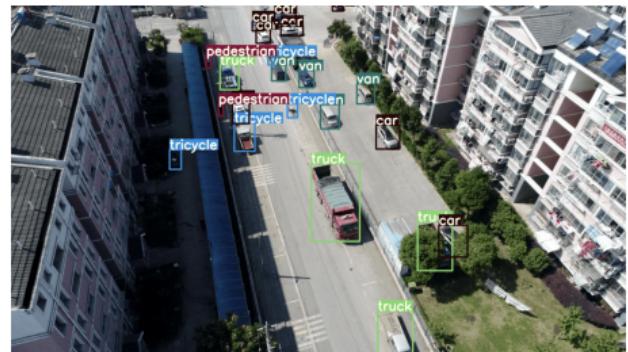
- Conception
- Relative work
- Implement

Object Detection

- Traditional object detection
 - Relied on handcrafted features (e.g., Histograms of oriented gradients, HOG) [5].
 - Required complex feature engineering; achieved limited accuracy.

Object Detection

- Traditional object detection
 - Relied on handcrafted features (e.g., Histograms of oriented gradients, HOG) [5].
 - Required complex feature engineering; achieved limited accuracy.
- Deep Learning-based detection
 - Revolutionized by automatic feature extraction from large datasets.
 - Two-stage methods: For instance "Faster Region-based Convolutional Neural Network (Faster R-CNN) [6]", which is more accurate, slower performance.
 - One-stage methods: For instance "You Only Look Once (YOLO)" [7], which is faster, slightly less accurate.



Particle picking in Cryo-EM

For automated particle picking in Cryo-EM, we can consider it a specialized task in object detection.

- Deep Learning-based approaches
 - Algorithms leverage high-capacity models to **improve detection accuracy in low-SNR conditions.**
 - Capable of handling **heterogeneous particle structures**, surpassing traditional template-matching methods [3], [4].
- Significance in Cryo-EM workflow [8]
 - Automation minimizes manual annotation efforts, improving efficiency and reproducibility.
 - Provides precise particle coordinates crucial for downstream structural studies.

Particle picking in Cryo-EM

For automated particle picking in Cryo-EM, we can consider it a specialized task in object detection.

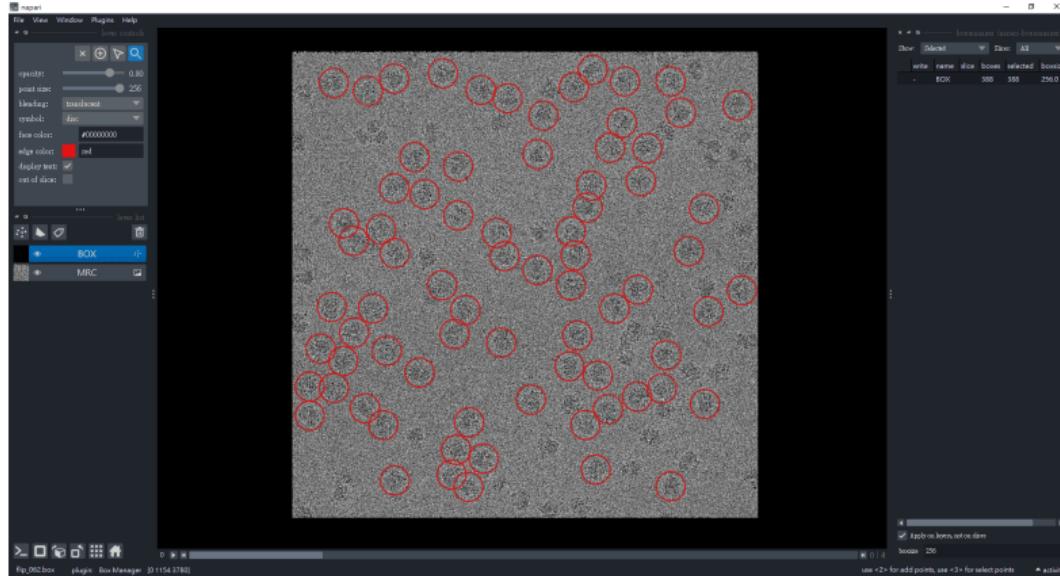


Figure 2-2: Napari GUI

Active Learning

According to Burr Settles' Artificial Intelligence and Machine Learning (2012) [9].

Three Main Scenarios:

- ① Query synthesis: The model generates new questions or samples for expert annotation.
- ② Stream-based selective: Decides in real time whether to select incoming data points for annotation.

Active Learning

According to Burr Settles' Artificial Intelligence and Machine Learning (2012) [9].

Three Main Scenarios:

- ① Query synthesis: The model generates new questions or samples for expert annotation.
- ② Stream-based selective: Decides in real time whether to select incoming data points for annotation.
- ③ Pool-based sampling
 - Evaluates a pre-collected pool of unlabeled data and **selects high-value samples** for annotation.
 - **Ideal for scenarios with a large volume of unlabeled data.**

Pool-based Active Learning Framework

- The labeled pool consists of a small initial set of annotated micrographs utilized for training the model [10], [11].
- The unlabeled pool comprises the remaining unannotated micrographs, from which uncertain regions are selected for expert annotation during the iterative process.

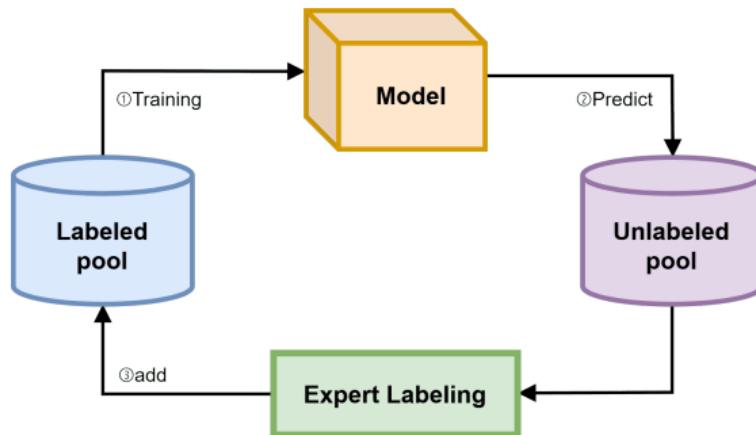


Figure 2-3: Schematic diagram of the abstract concept

Uncertainty Measurement [9]

① Least confident

- Selects samples with the lowest confidence in the most likely label.
- Ignores information beyond the most probable label.

② Margin

- Selects samples with the smallest difference between the top two label probabilities.
- Focuses on ambiguous samples for better discrimination.

③ Entropy

- Measures uncertainty by considering the entire label distribution.
- High entropy indicates higher uncertainty in predictions.

$$H(x) = - \sum_k p_k \log (p_k)$$

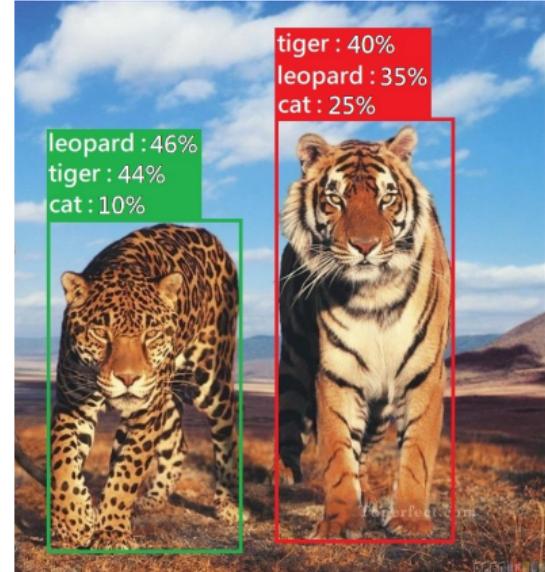


Figure 2-4: Example
Entropy
0.95
1.08

Entropy Score

In their 2023 study, Robert Kiewisz and Tristan Bepler introduced the "Entropy Score" as an uncertainty metric for Active Learning, specifically in single-particle picking tasks [12]. Utilizing a binary logistic regression (BLR) model with a decision boundary at 0.5,

Entropy Score

$$\text{Entropy Score} = -\sigma(\text{logits}) \times \log(\sigma(\text{logits})) - \sigma(-\text{logits}) \times \log(\sigma(-\text{logits})) \quad (1)$$

$$\sigma = \frac{e^x}{1 + e^x} \text{ (sigmoid function),}$$

where:

$$\text{logits} = \log\left(\frac{\text{confidence}}{1 - \text{confidence}}\right)$$

the Entropy Score reflects higher uncertainty when confidence values are near the threshold, signifying ambiguity in classification.

crYOLO

crYOLO is a Deep Learning software specifically designed for Cryo-EM [13]. crYOLO has several key features including:

- Automated particle picking: Enables fast, automated particle picking with human-level accuracy and minimal effort.
- Filament sample picking: Supports particle selection from filament-shaped samples.

crYOLO

crYOLO is a Deep Learning software specifically designed for Cryo-EM [13]. crYOLO has several key features including:

- Intuitive GUI: Features a user-friendly graphical user interface (GUI) that simplifies model training and parameter adjustment, making it accessible to researchers without extensive computational expertise.

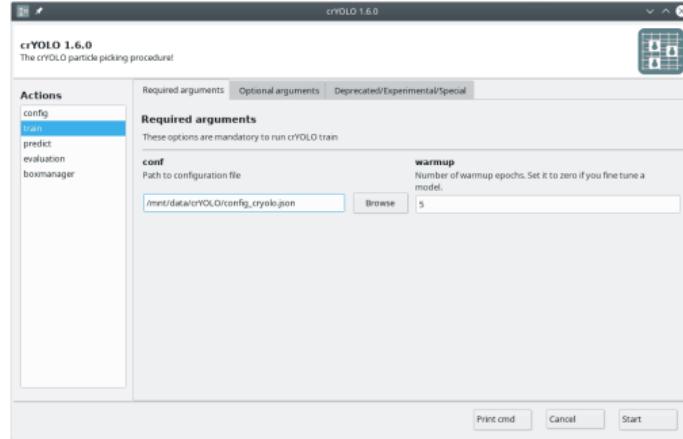


Figure 2-5: crYOLO GUI

crYOLO-PhosaurusNet

The architecture incorporates
Batch Normalization and
Leaky ReLU activation functions through-
out the convolutional layers.

Table 2-1: Architecture of PhosaurusNet
(given 1024×1024 as input)

Layer	FilterType	Filters	Kernel Size	Output Shape
Feature Extraction				
1	Convolutional Max-Pooling	32	3×3	$1024 \times 1024 \times 32$
2	Convolutional Max-Pooling	64	3×3	$512 \times 512 \times 64$
3	Convolutional	128	3×3	$256 \times 256 \times 128$
4	Convolutional	64	1×1	$256 \times 256 \times 64$
5	Convolutional Max-Pooling	128	3×3	$256 \times 256 \times 128$
6	Convolutional	256	3×3	$128 \times 128 \times 256$
7	Convolutional	128	1×1	$128 \times 128 \times 128$
8	Convolutional Max-Pooling	256	3×3	$128 \times 128 \times 256$
9	Convolutional	512	3×3	$64 \times 64 \times 512$
10	Convolutional	256	1×1	$64 \times 64 \times 256$
11	Convolutional	512	3×3	$64 \times 64 \times 512$
12	Convolutional	256	1×1	$64 \times 64 \times 256$
13	Convolutional Max-Pooling	512	3×3	$64 \times 64 \times 512$
14	Convolutional	1024	3×3	$32 \times 32 \times 1024$
15	Convolutional	512	1×1	$32 \times 32 \times 512$
16	Convolutional	1024	3×3	$32 \times 32 \times 1024$
17	Convolutional	512	1×1	$32 \times 32 \times 512$
18	Convolutional	1024	3×3	$32 \times 32 \times 1024$
19	Convolutional	1024	3×3	$32 \times 32 \times 1024$
20	Convolutional	1024	3×3	$32 \times 32 \times 1024$
20a	Upsampling			$64 \times 64 \times 1024$
21	Convolutional (using output of 13) Concatenate 20a & 21	256	1×1	$64 \times 64 \times 256$ $64 \times 64 \times 1280$
22	Convolutional	1024	3×3	$64 \times 64 \times 1024$
Detection				
1	Dropout (0.2) Convolutional	6	1×1	

CryoSPARC in Research

CryoSPARC is a software package designed for determining high-resolution 3D structures of biomolecules using Cryo-EM [14]. Its key functions include contrast transfer function (CTF) estimation, Ab-initio model generation, and **structure refinement**.

Feature comparison	Homogeneous Refinement	Non-Uniform Refinement
Assumption	Assumes structure is rigid and uniform	Allows for heterogeneity and flexibility
Suitable Samples	Stable and rigid molecules	Complex samples with flexibility or multiple conformations
Local Resolution	Lower in flexible or heterogeneous regions	Improved in flexible or heterogeneous regions
Computational Burden	Lower	Higher

Table 2-2: Comparison between Homogeneous Refinement and Non-Uniform Refinement

Method

- Framework
- Uncertainty
- Evaluation

Pool-based Active Learning

In the Active Learning framework, the process involves iterative predictions on the training set using the model [10], [11].

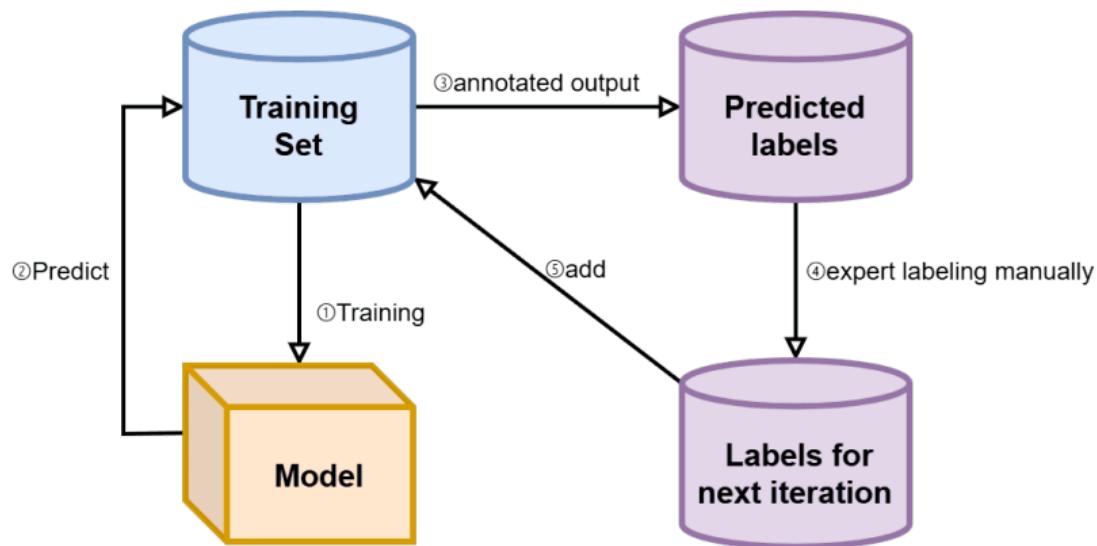


Figure 3-6: Abstract Framework

Particle Correction Process

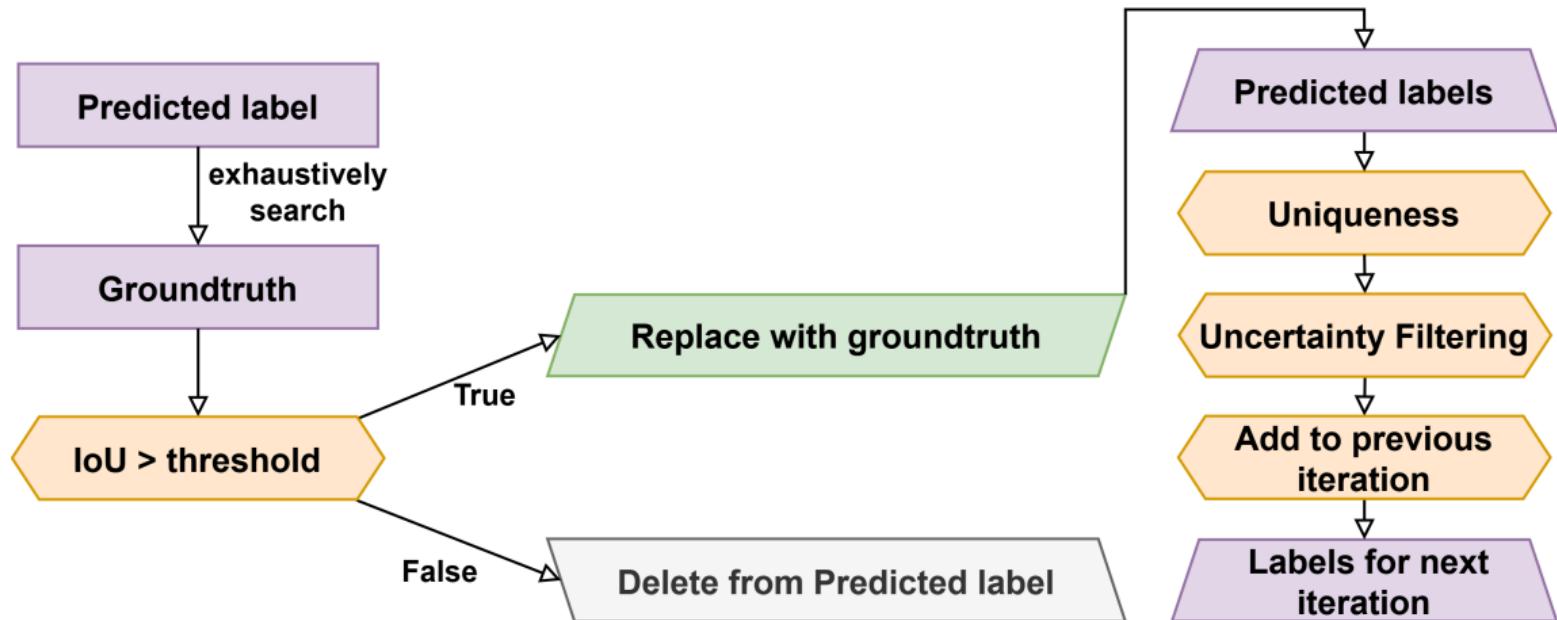


Figure 3-7: Particle correction process

IoU Calculation

Intersection over Union

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}, \quad \text{where:}$$

$$\text{Area of Intersection} = A \cap B$$

$$\text{Area of Union} = A + B - A \cap B = A \cup B$$

(2)

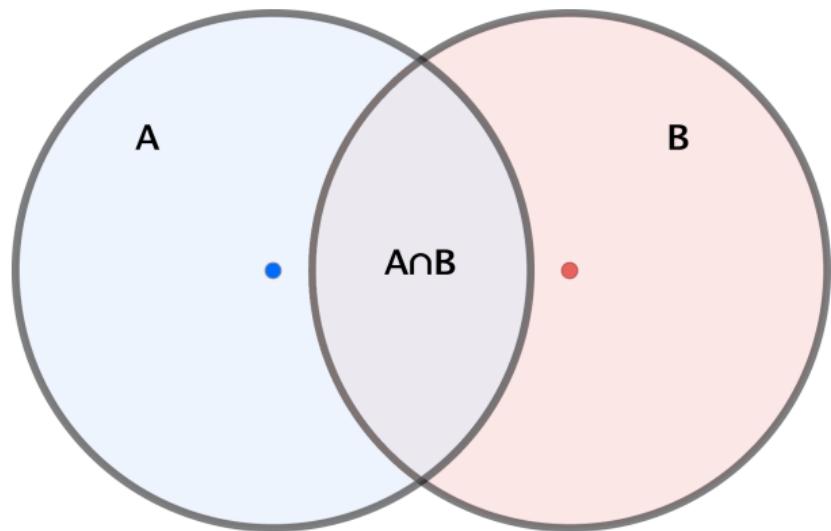


Figure 3-8: IoU

IoU Calculation

Intersection over Union

$$\frac{\text{Area of Intersection}}{\text{Area of Union}}, \quad \text{where:}$$

$$\text{Area of Intersection} = A \cap B$$

$$\text{Area of Union} = A + B - A \cap B = A \cup B$$

(2)

Assumption:

Circles A and B have equal radii.

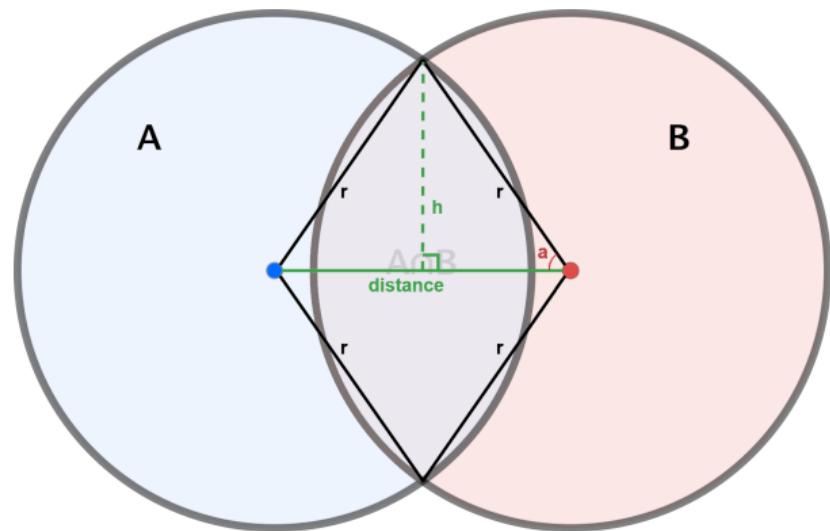


Figure 3-8: IoU Calculation

Area of Intersection

Area of intersection

$$2 \times \text{Sector} - \text{Diamond} \quad (3)$$

Area of Sector

$$\begin{aligned}\text{Sector} &= \frac{1}{2} \text{radius}^2 \times \text{radian} \\ \cos(\angle a) &= \frac{\frac{\text{distance}}{2}}{\text{radius}} = \frac{\text{distance}}{2 \times \text{radius}}, \quad \angle a = \arccos\left(\frac{\text{distance}}{2 \times \text{radius}}\right)\end{aligned} \quad (4)$$

∴ The radian of the sector = $2\angle a$

$$\begin{aligned}\text{Sector area} &= \frac{1}{2} \text{radius}^2 \times 2\angle a \\ &= \text{radius}^2 \times \angle a = \text{radius}^2 \times \arccos\left(\frac{\text{distance}}{2 \times \text{radius}}\right)\end{aligned}$$

IoU constraint

For explain,

The green arrow lines indicate
IoU>threshold.

The red line indicates that no
groundtruth label satisfies IoU >
threshold.

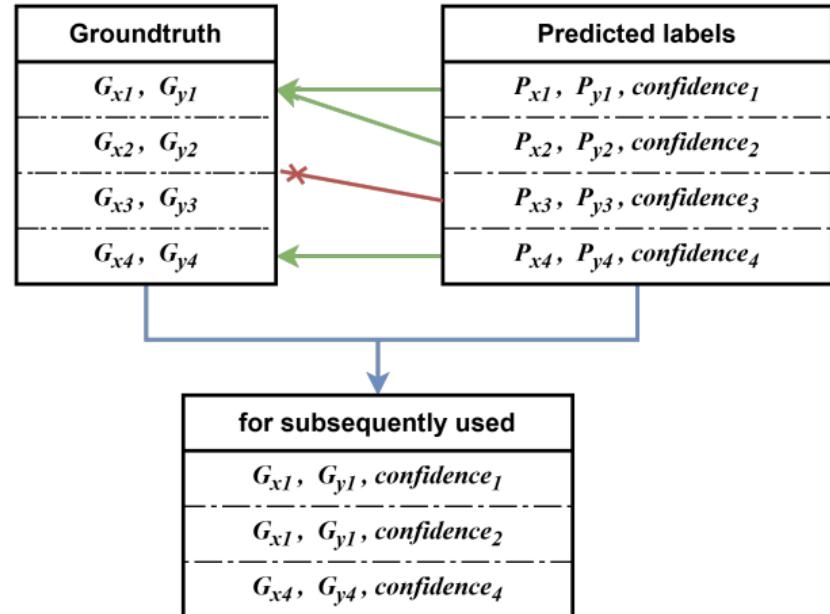


Figure 3-9: i.e., Particle correction process

Uncertainty Measurements

① Low Confidence

- After the model makes predictions, **non-maximum suppression (NMS)** is applied to eliminate overlapping predictions [15].
- NMS retains only the most confident predictions in overlapping areas, enabling focus on meaningful candidates.
- **Predictions with lower confidence are selected as particles with higher uncertainty [9].**
- This approach is similar to the least-confident selection method discussed in the literature review.

Uncertainty Measurements

② Entropy Score

This method, as applied by Robert Kiewisz and Tristan Bepler in the field of Cryo-EM, has already been discussed in the literature review [12].

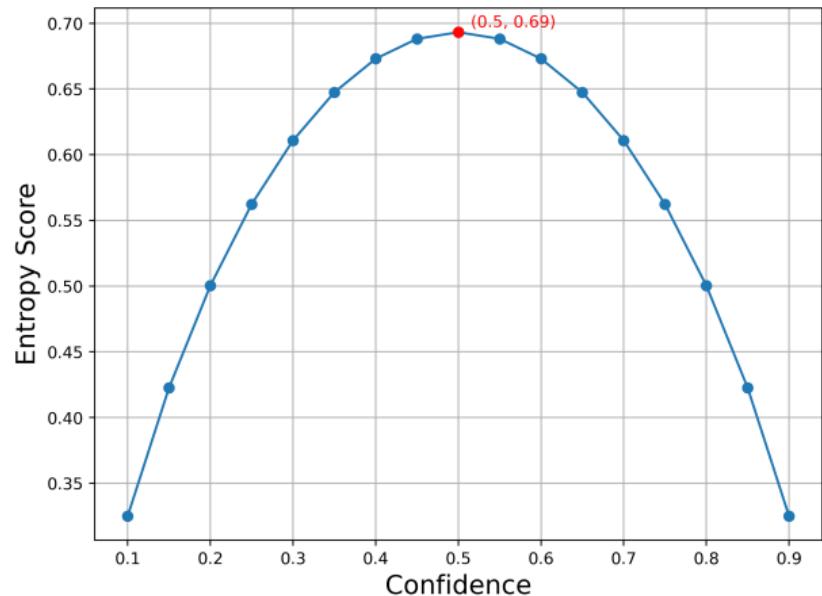


Figure 3-10: Entropy Score v.s. Confidence

Uncertainty Measurements

② Entropy Score

This method, as applied by Robert Kiewisz and Tristan Bepler in the field of Cryo-EM, has already been discussed in the literature review [12].

③ Entropy Score with normalized confidence :

- Confidence values are normalized to $[0.01, 0.99]$ via Min-Max scaling.
- This method will be referred to as "Norm-conf-es" in subsequent sections.

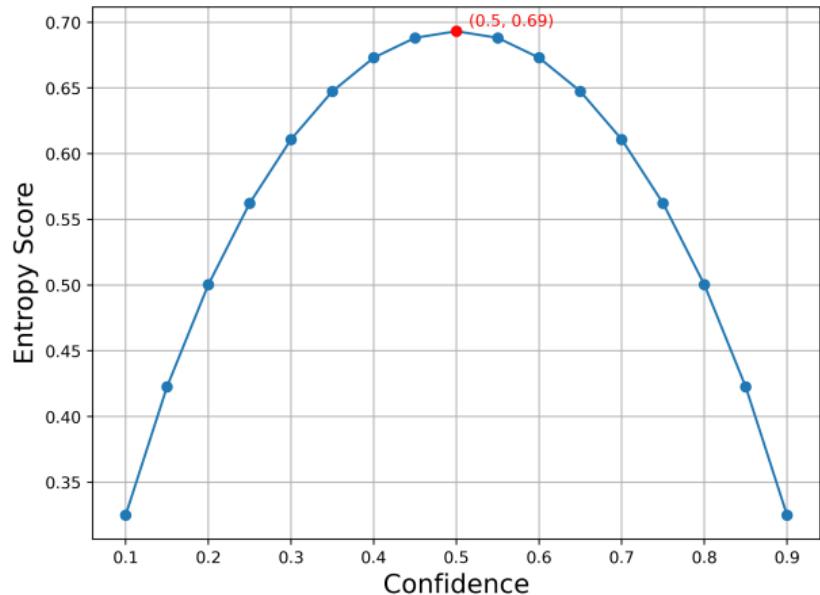


Figure 3-10: Entropy Score v.s. Confidence

Uncertainty Measurements

④ Boundary Distance

- crYOLO uses a grid search algorithm to find the optimal confidence threshold, referred to as ' T_{opt} ', which maximizes the F1-score and serves as the decision boundary for classification.
- When the confidence of a predicted label exceeds T_{opt} , the model classifies it as a particle, and when the confidence is below T_{opt} , it is classified as a non-particle.
- **The uncertainty is defined as the distance between the predicted confidence and T_{opt}**
- T_{opt} is obtained by evaluating the model on the validation set, since groundtruth is only available for the validation set in real-world applications.

Boundary Distance

$$Uncertainty = |Confidence - T_{opt}| \quad (5)$$

Evaluation Metrics

① Mean IoU

- The mean IoU is used as one of the model evaluation metrics in crYOLO, where predictions with confidence above T_{opt} are first selected and compared with the groundtruth labels.
- For each comparison, the bounding box with **the highest IoU is selected**, and if the IoU exceeds the threshold (**set to 0.5 in crYOLO**), it is considered a true positive (TP), with the IoU value added to the calculation of the mean IoU.
- The final mean IoU is the average of IoU values across all images.

Mean IoU

$$Mean\ IoU = \frac{1}{N} \times \sum_{i=1}^N IoU_i \quad (6)$$

Where N is the number of groundtruth labels, and IoU_i is the maximum IoU between the i -th predicted bounding box and the corresponding groundtruth, given that $IoU_i > 0.5$.

Evaluation Metrics

② Recall

A metric that evaluates a model's detection capability by calculating the TP and false negatives (FN).

Recall

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

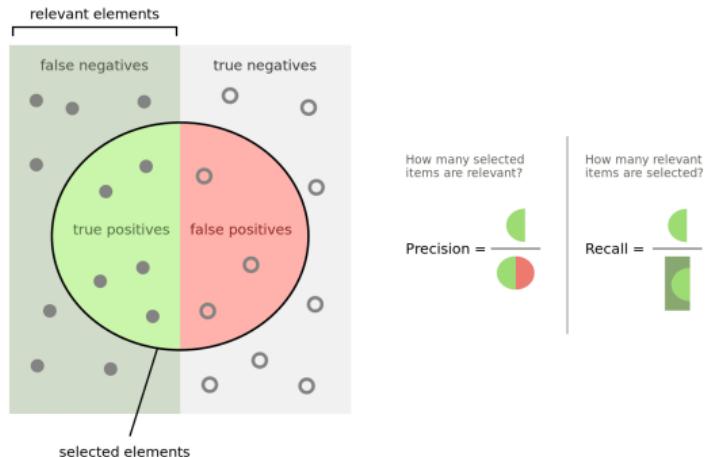


Figure 3-11: Outcome Description

Evaluation Metrics

③ AUC

The Area Under the Curve (AUC) measures a model's balance between precision and recall by calculating the area under the Precision-Recall curve. We know that TP and FP (False Positive) represent the number of particles correctly detected by the model and the number of particles incorrectly detected, respectively.

$$Precision = \frac{TP}{TP + FP}$$

AUC

$$AUC = \sum_{i=1}^n (Recall_i - Recall_{i-1}) \times Precision_i , \quad \text{where } n=1000 \quad (8)$$

Calculated with **a 0.001 step size**, iterating precision and recall for confidence values from 0 to 1 via **rectangular approximation**.

Evaluation Metrics

④ F1-Score

Since we have already described the calculation of recall and precision above, the F1-Score can be expressed as:

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Evaluation Metrics

⑤ GSFSC

The Gold Standard Fourier Shell Correlation (GSFSC) resolution serves as a vital approach for evaluating reconstruction quality [16].

FSC

$$FSC_{(1,2)}(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} F_1^2(r_i) \cdot \sum_{r_i \in r} F_2^2(r_i)}} \quad (10)$$

- $F_1(r)$ and $F_2(r)$: The Fourier coefficients of half-map 1 and half-map 2, respectively.
- $*$: Complex conjugate.
- r : Radius of the frequency shell.
- r_i : A point in Fourier space representing a specific spatial frequency.

Experiment

- Practical Workflow
- Dataset
- Experimental Setting
- Result

Practical Workflow

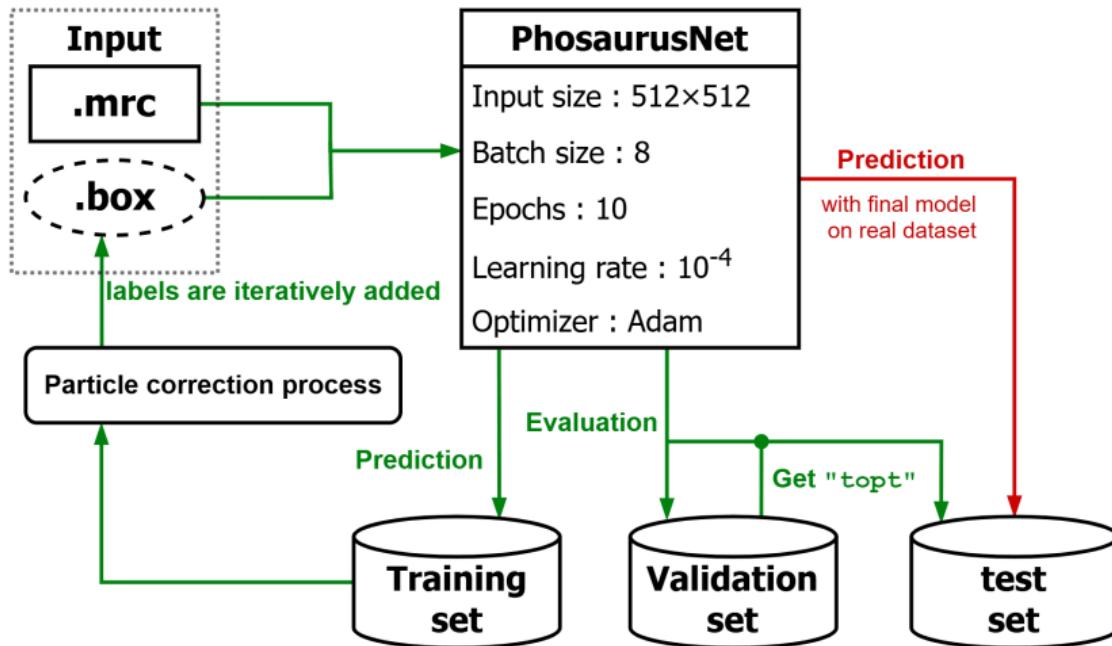


Figure 4-12: Practical Workflow

Dataset Source

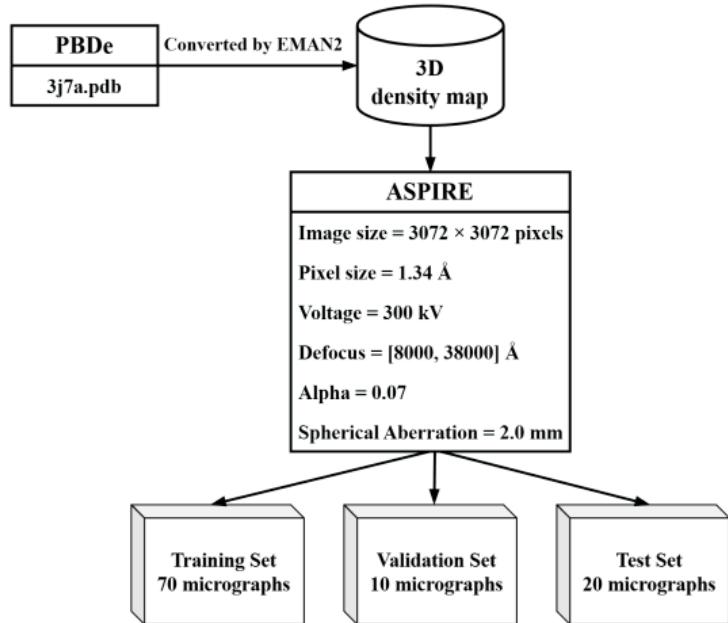


Figure 4-13: Synthetic dataset

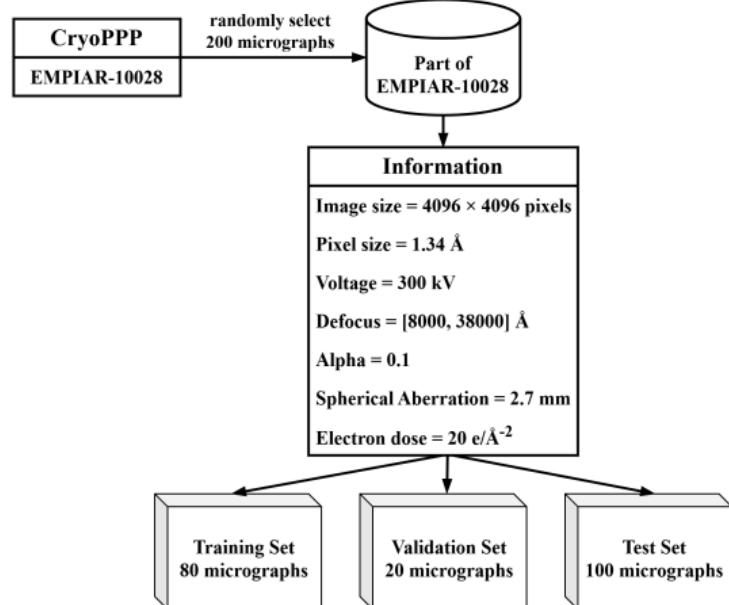
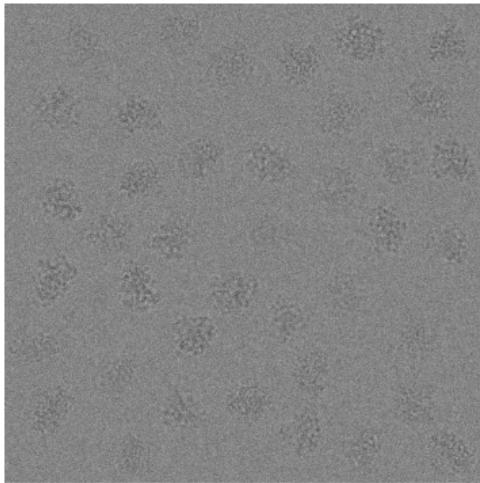


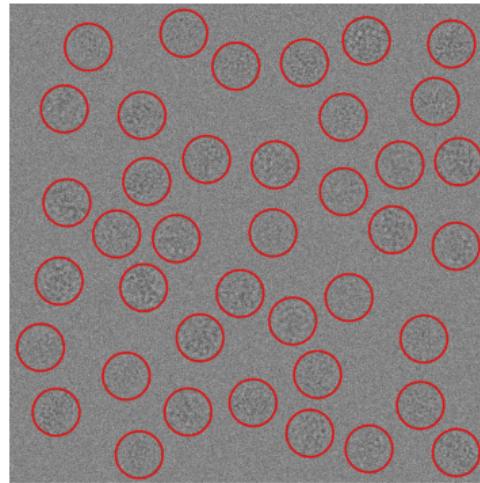
Figure 4-14: Real dataset

Synthetic Dataset

Each micrograph contained 40 particles with an average SNR of approximately 0.05.

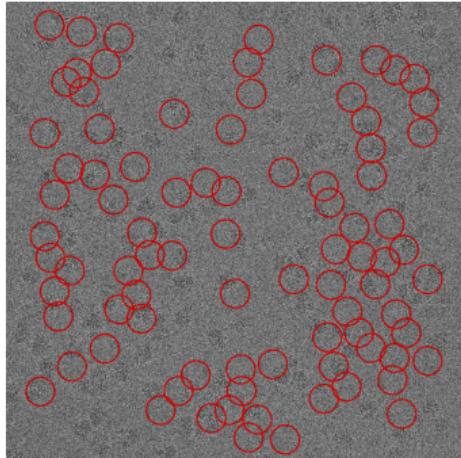


(a)

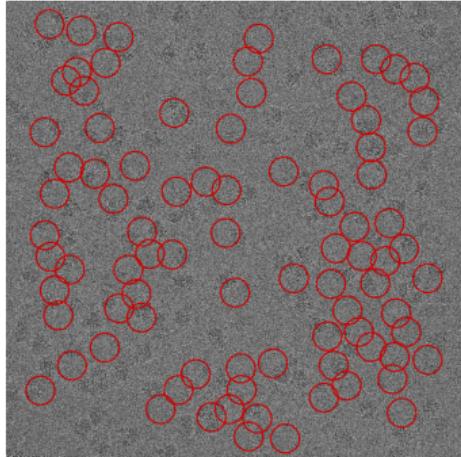


(b)

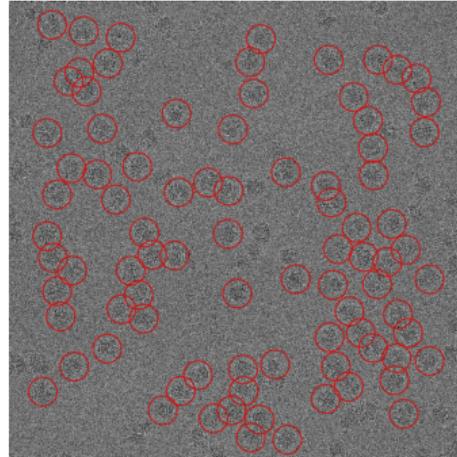
Figure 4-15: (a) One of the micrograph simulated by ASPIRE; (b) with the corresponding groundtruth represented by red bounding boxes.



(a)



(a)



(b)

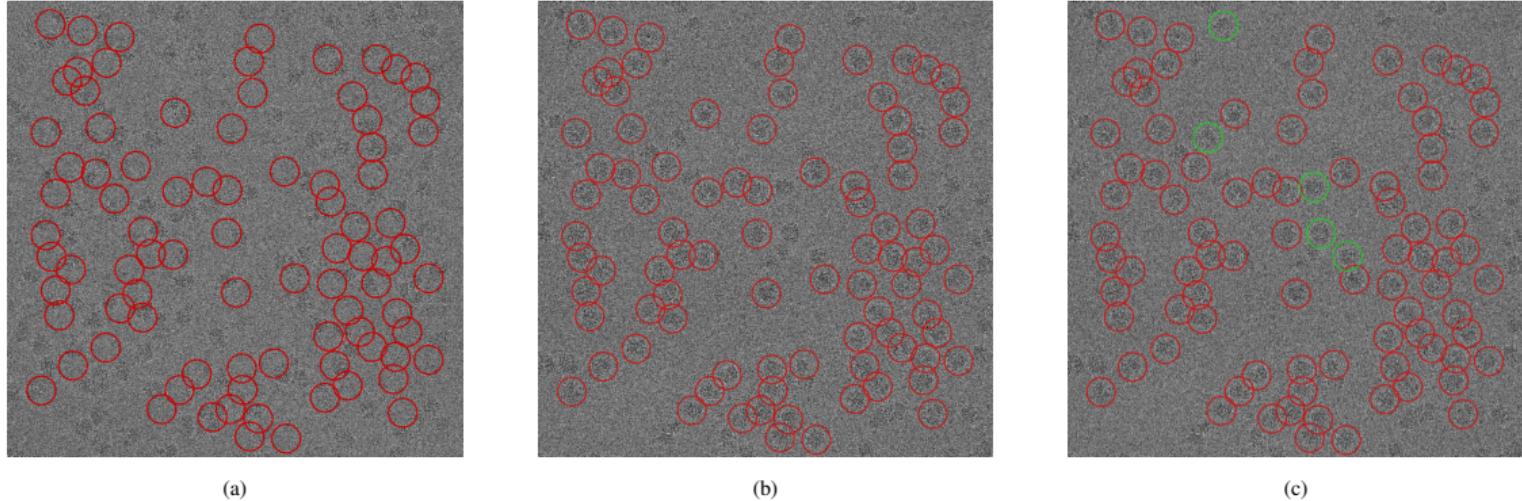


Figure 4-16: EMPIAR-10028 Dataset with annotations provided by CryoPPP:

- (a) Original annotation;
- (b) After flip micrographs;
- (c) The particles that CryoPPP did not pick are labeled with green bounding boxes.

Experimental Setting on Synthetic Data

	Synthetic Dataset
Origin training particles	2800 (100% Training set)
Initial particles used	70 (2.5% Training set)
Particles added per iteration	70 (2.5% Training set)
IoU constraint	> 0.7
Iterations of active learning	9 times
Epochs per iteration	10 epochs
Final particles used	630 (22.5% Training set)

Table 4-3: Hyperparameters used in synthetic data

Experimental Setting on Real Data

The sub-dataset of the EMPIAR-10028 dataset provided by CryoPPP			
Experiment	i5-a8.75	i15-a7.5	i25-a6.25
Origin training particles		7136 (100%)	
Initial particles used	357 ($\approx 5\%$)	1070 ($\approx 15\%$)	1784 ($\approx 25\%$)
Particles added per iteration	624 ($\approx 8.75\%$)	535 ($\approx 7.5\%$)	446 ($\approx 6.25\%$)
IoU constraint		> 0.3	
Iterations of active learning		9 times	
Epochs per iteration		10 epochs	
Final particles used	5349 ($\approx 75\%$)	5350 ($\approx 75\%$)	5352 ($\approx 75\%$)

Table 4-4: Hyperparameters used in subset of the EMPIAR-10028 dataset provided by CryoPPP, "i" indicates the proportion of initial particles used, while "a" represents the proportion of particles added per iteration. These proportions are calculated relative to the total count of original training particles, which is taken as 100%.

Distinction on Synthetic Dataset

While all methods used the same total number of particles, the table highlights the differences in the specific particles selected, reflecting the distinct selection behaviors of each method. The average number of distinct labels between pairs of methods is approximately $\frac{431}{560}$ ($\approx 77\%$).

	Random	Low confidence	Entropy Score	Norm-conf-es	Boundary Distance
Random	0	454	386	392	416
Low confidence	454	0	509	496	418
Entropy Score	386	509	0	313	470
Norm-conf-es	392	496	313	0	453
Boundary Distance	416	418	470	453	0

Table 4-5: Distinct particle counts on synthetic data.

i.e., observing the "Random" column and the "Entropy Score" row, the entry shows 386 unique particles between the two methods in the final training model.

Result on Synthetic Dataset

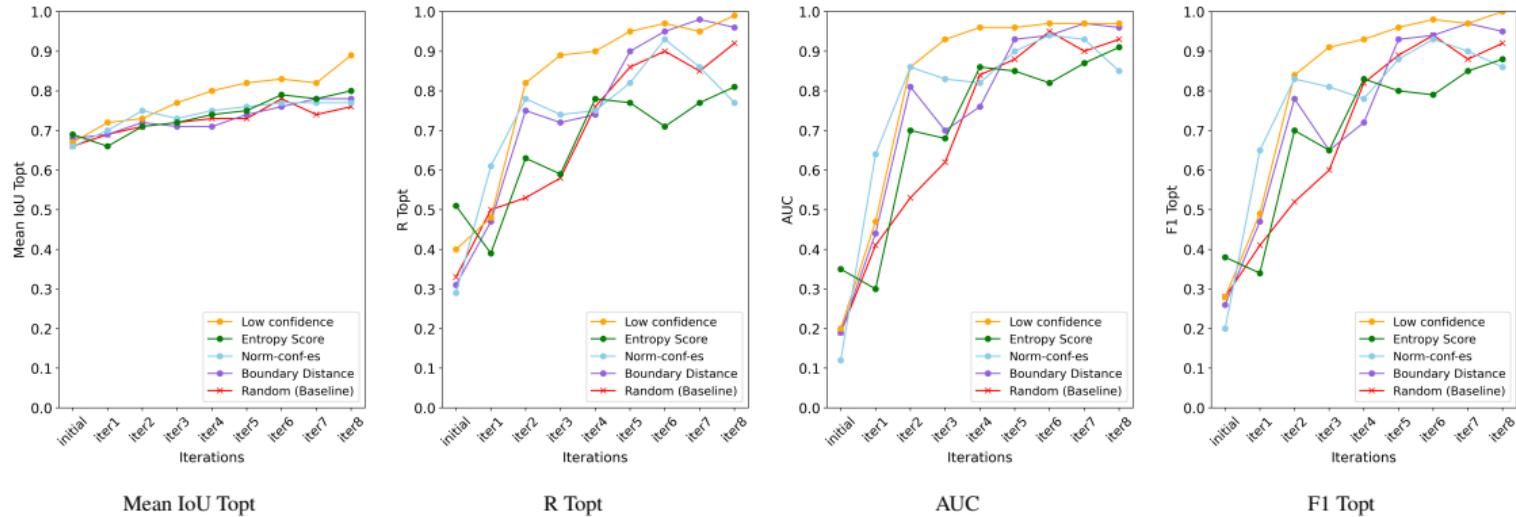


Figure 4-17: Performance metrics for synthetic datasets.

Result on "i5a-8.75"

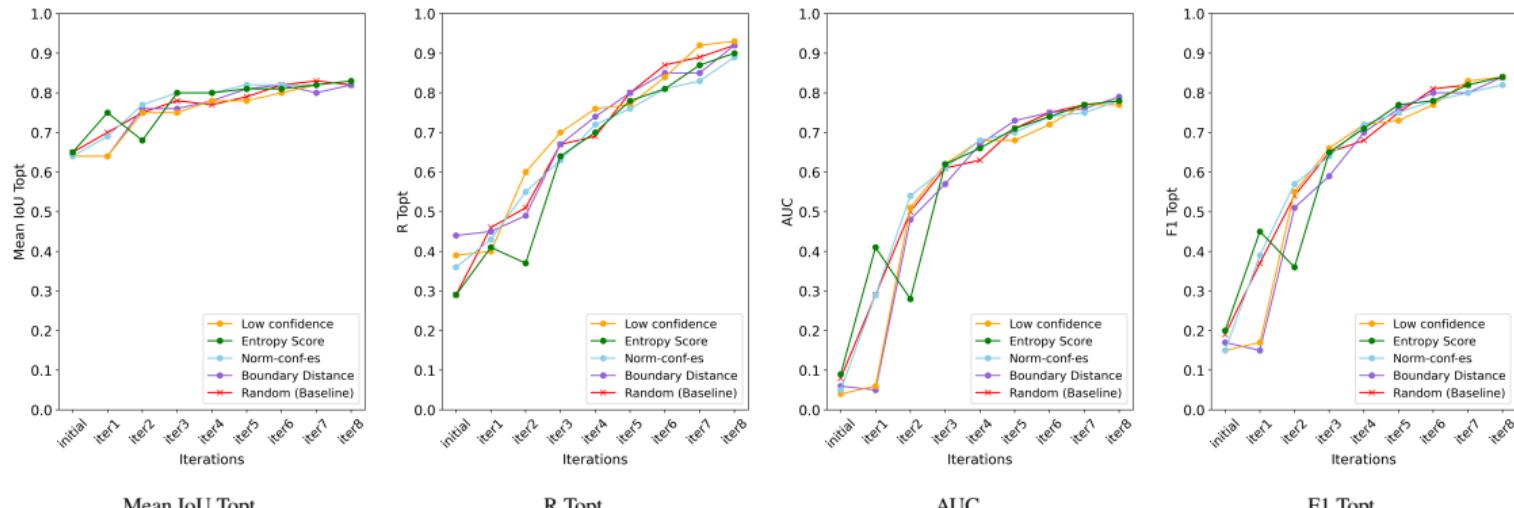


Figure 4-18: Performance metrics "i5a875".

Result on "i15-a7.5"

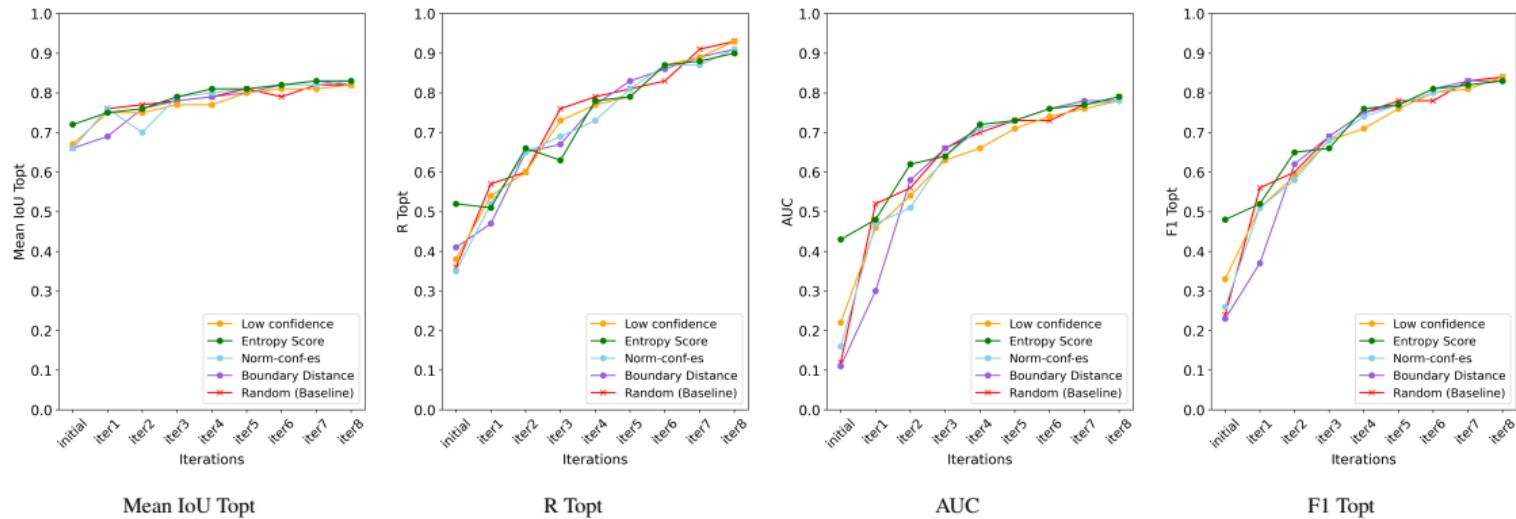


Figure 4-19: Performance metrics "i15a75".

Result on "i25-a6.25

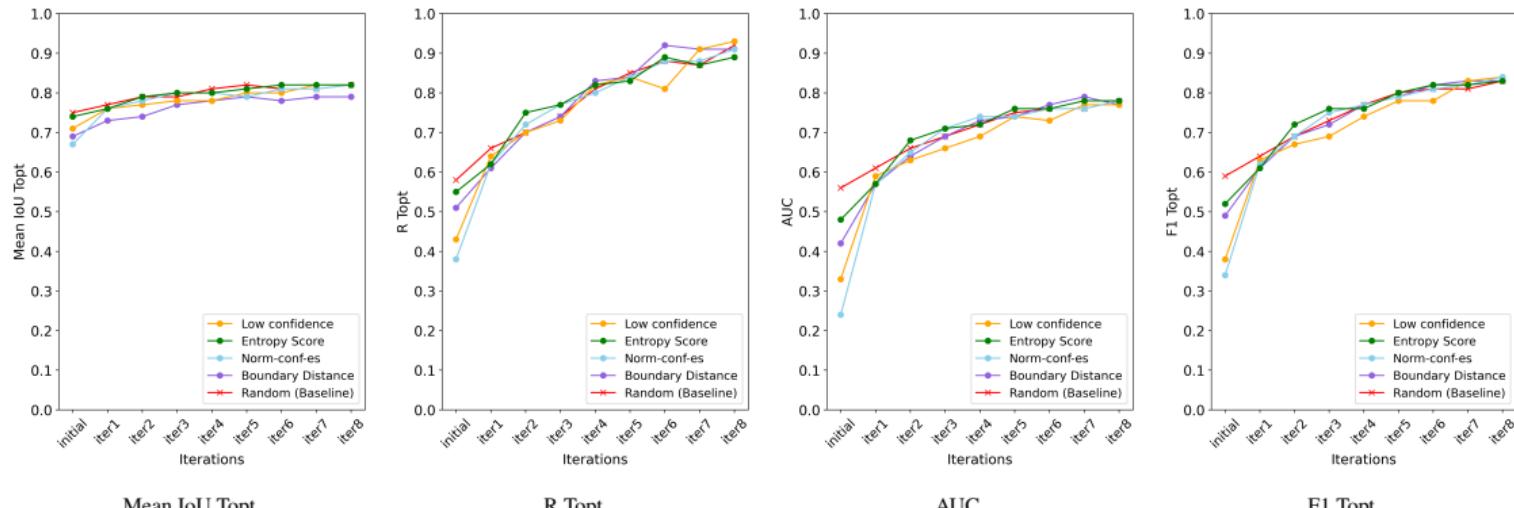


Figure 4-20: Performance metrics "i25a625".

Distinction on EMPIAR-10028

In the "i25-a6.25" experimental setting, we initially used $\frac{1,784}{7136}$ (25%) labels, which were ultimately increased to $\frac{5352}{7136}$ (75%) labels. As shown in the Table 4-6, the average number of distinct labels between pairs of methods is approximately $\frac{1170}{1784}$ ($\approx 65\%$).

	Random	Low confidence	Entropy Score	Norm-conf-es	Boundary Distance
Random	0	1182	1128	1139	1169
Low confidence	1182	0	1298	1280	1174
Entropy Score	1128	1298	0	983	1179
Norm-conf-es	1139	1280	983	0	1166
Boundary Distance	1169	1174	1179	1166	0

Table 4-6: With the experimental setting "i25-a6.25, the distinct particle counts between methods used for the final training model.

Reconstructure use with Non-Uniform Refinement

Experiment	Non-Uniform Refinement														
	i5-a8.75					i15-a7.5					i25-a6.25				
	1	2	3	Mean	SD	1	2	3	Mean	SD	1	2	3	Mean	SD
Random	6.22 Å	6.38 Å	6.42 Å	6.340 Å	0.086 Å	5.93 Å	6.12 Å	5.96 Å	6.003 Å	0.083 Å	6.35 Å	6.41 Å	6.38 Å	6.380 Å	0.025 Å
Low confidence	6.05 Å	5.70 Å	5.73 Å	5.840 Å	0.149 Å	6.10 Å	6.00 Å	5.83 Å	5.977 Å	0.112 Å	6.10 Å	6.01 Å	5.92 Å	6.010 Å	0.074 Å
Entropy Score	6.06 Å	6.11 Å	6.09 Å	6.087 Å	0.021 Å	6.20 Å	6.28 Å	6.13 Å	6.203 Å	0.061 Å	5.71 Å	5.81 Å	5.74 Å	5.753 Å	0.042 Å
Norm-conf-es	5.96 Å	5.95 Å	6.01 Å	5.973 Å	0.026 Å	6.06 Å	6.13 Å	6.21 Å	6.133 Å	0.061 Å	6.05 Å	6.02 Å	6.10 Å	6.057 Å	0.033 Å
Boundary Distance	5.80 Å	5.81 Å	5.80 Å	5.803 Å	0.005 Å	5.74 Å	5.74 Å	5.72 Å	5.733 Å	0.009 Å	5.29 Å	5.53 Å	5.32 Å	5.380 Å	0.107 Å

Table 4-7: Reconstructure use with Non-Uniform Refinement on the real dataset. ("SD" refers to the standard deviation.)

- Boundary Distance exhibiting a higher standard deviation (0.107) in the "i25-a6.25" setting, but its performance remains superior to random selection, even when accounting for variability within one standard deviation. **Indicating no significant disadvantage.**

Discussion

- Conclusion
- Challenge & Limitation
- Future work

Conclusion

- ① Active Learning Objective: Active Learning enhances model performance while reducing labeling costs by selectively annotating valuable samples, crucial for effective 2D particle picking in Cryo-EM to enable accurate 3D reconstruction.

Conclusion

- ① Active Learning Objective: Active Learning enhances model performance while **reducing labeling costs by selectively annotating valuable samples**, crucial for effective 2D particle picking in Cryo-EM to enable accurate 3D reconstruction.
- ② Synthetic Dataset Results: The "Low Confidence" and "Boundary Distance" uncertainty metrics significantly improved model iteration progress compared to other methods.

Conclusion

- ① Active Learning Objective: Active Learning enhances model performance while **reducing labeling costs by selectively annotating valuable samples**, crucial for effective 2D particle picking in Cryo-EM to enable accurate 3D reconstruction.
- ② Synthetic Dataset Results: The "Low Confidence" and "Boundary Distance" uncertainty metrics significantly improved model iteration progress compared to other methods.
- ③ Real Dataset Observations: 2D metrics showed no significant differences among methods, with initial particles intentionally set at low proportions to emphasize uncertainty-based selections.

Conclusion

- ① Active Learning Objective: Active Learning enhances model performance while **reducing labeling costs by selectively annotating valuable samples**, crucial for effective 2D particle picking in Cryo-EM to enable accurate 3D reconstruction.
- ② Synthetic Dataset Results: The "Low Confidence" and "Boundary Distance" uncertainty metrics significantly improved model iteration progress compared to other methods.
- ③ Real Dataset Observations: 2D metrics showed no significant differences among methods, with initial particles intentionally set at low proportions to emphasize uncertainty-based selections.
- ④ 3D Reconstruction Outcome: The Boundary Distance method, in the "i25-a6.25" configuration with 100 micrographs, achieved the best 3D reconstruction with an average resolution of 5.38 Å. Although the publicly available resolution is 3.2 Å, **our result was obtained with only 100 micrographs**.

Conclusion

- ① Active Learning Objective: Active Learning enhances model performance while **reducing labeling costs by selectively annotating valuable samples**, crucial for effective 2D particle picking in Cryo-EM to enable accurate 3D reconstruction.
- ② Synthetic Dataset Results: The "Low Confidence" and "Boundary Distance" uncertainty metrics significantly improved model iteration progress compared to other methods.
- ③ Real Dataset Observations: 2D metrics showed no significant differences among methods, with initial particles intentionally set at low proportions to emphasize uncertainty-based selections.
- ④ 3D Reconstruction Outcome: The Boundary Distance method, in the "i25-a6.25" configuration with 100 micrographs, achieved the best 3D reconstruction with an average resolution of 5.38 Å. Although the publicly available resolution is 3.2 Å, **our result was obtained with only 100 micrographs**.

Challenge & Limitation

- Subjectivity in Manual Annotations: The "Particle Correction Process" depends on the quality and consistency of manual annotations, which can introduce subjectivity and impact model training.
- Uncertainty-Based Selection: While uncertainty-based particle selection is effective, the optimal ratio between initial and incremental particles in each iteration remains unclear.
- Small Validation Dataset: The relatively small validation dataset may result in biased evaluation outcomes if certain classes are over- or underrepresented, limiting the model's ability to generalize effectively.

Future work

① Particle Correction Process Limitation:

- The current method updates predicted particle coordinates upon identifying any groundtruth label meeting the IoU threshold, **potentially overlooking a label with a higher IoU later in the process.**

② Positional Information Integration [17], [18]:

- Traditional uncertainty-based methods may overlook critical regions or select similar samples, leading to suboptimal performance.
- Integrating positional context with uncertainty metrics could improve particle picking by ensuring **more diverse and representative sample** selection.

Future work

③ Hardware Limitations:

- The validation dataset was small due to hardware constraints, which may affect the model's ability to learn effectively across various classes.
- Exploring **lightweight neural network architectures** could reduce computational demands while maintaining performance.

④ Improve the determination of T_{opt} :

- Further studies could explore additional uncertainty measurement methods for more reliable decision boundary estimation, such as developing better techniques to determine the optimal confidence threshold (T_{opt}).
- One potential approach could involve using the F2-Score offered by crYOLO, which may provide a more stable and reliable decision boundary, helping to refine the model's performance.

References I

- [1] Jacques Dubochet, Joachim Frank, and Richard Henderson. *Popular science background: They captured life in atomic detail.* 2017.
- [2] Chin-Yu Chen, Bo-Lin Lin Yuan-Chih Chang, Chun-Hsiang Huang Kuan-Fu Lin, et al. “Use of Cryo-EM To Uncover Structural Bases of pH Effect and Cofactor Bispecificity of Ketol-Acid Reductoisomerase”. In: *Journal of the American Chemical Society* 141.15 (2019), pp. 6136–6140. doi: 10.1021/jacs.9b01354.
- [3] Ruijie Yao, Jiaqiang Qian, and Qiang Huang. “Deep-learning with synthetic data enables automated picking of cryo-EM particle images of biological macromolecules”. In: *Bioinformatics* 36.4 (Oct. 2019), pp. 1252–1259. ISSN: 1367-4803. doi: 10.1093/bioinformatics/btz728.
- [4] Adil Al-Azzawi, Anes Ouadou, Highsmith Max, et al. “DeepCryoPicker: fully automated deep neural network for single protein particle picking in cryo-EM”. In: *BMC Bioinformatics* 21.1 (2020), p. 509. ISSN: 1471-2105. doi: 10.1186/s12859-020-03809-7.
- [5] Zhengxia Zou, Keyan Chen, Zhenwei Shi, et al. “Object Detection in 20 Years: A Survey”. In: *Proceedings of the IEEE* 111.3 (2023), pp. 257–276. doi: 10.1109/JPROC.2023.3238524.
- [6] Shaoqing Ren, Kaiming He, Ross B. Girshick, et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015), pp. 1137–1149.

References II

- [7] Joseph Redmon, Santosh Divvala, Ross Girshick, et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [8] Tristan Bepler, Andrew Morin, Micah Rapp, et al. “Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs”. In: *Nature Methods* 16.11 (2019), pp. 1153–1160. issn: 1548-7105. doi: [10.1038/s41592-019-0575-8](https://doi.org/10.1038/s41592-019-0575-8).
- [9] Burr Settles. *Active Learning*. Ed. by Ronald J. Brachman, William W. Cohen, and Thomas Dietterich. Vol. 1. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer, 2012. isbn: 9783031004322.
- [10] Hiranmayi Ranganathan, Hemanth Venkateswara, Shayok Chakraborty, et al. “Deep active learning for image classification”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. 2017, pp. 3934–3938. doi: [10.1109/ICIP.2017.8297020](https://doi.org/10.1109/ICIP.2017.8297020).
- [11] Mingfei Wu, Chen Li, and Zehuan Yao. “Deep Active Learning for Computer Vision Tasks: Methodologies, Applications, and Challenges”. In: *Applied Sciences* 12.16 (2022). issn: 2076-3417. doi: [10.3390/app12168103](https://doi.org/10.3390/app12168103).
- [12] Robert Kiewisz and Tristan Bepler. “An Active Learning Framework for ML-Assisted Labeling of Cryo-EM Micrographs”. In: *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*. Conference presentation, no DOI available. 2023.

References III

- [13] Thorsten Wagner and Stefan Raunser. “The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows”. In: *Communications Biology* 3.1 (2020), p. 61. ISSN: 2399-3642. doi: [10.1038/s42003-020-0790-y](https://doi.org/10.1038/s42003-020-0790-y).
- [14] Ali Punjani, John L. Rubinstein, David J. Fleet, et al. “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination”. In: *Nature Methods* 14.3 (2017), pp. 290–296. ISSN: 1548-7105. doi: [10.1038/nmeth.4169](https://doi.org/10.1038/nmeth.4169).
- [15] A. Neubeck and L. Van Gool. “Efficient Non-Maximum Suppression”. In: *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 3. 2006, pp. 850–855. doi: [10.1109/ICPR.2006.479](https://doi.org/10.1109/ICPR.2006.479).
- [16] Marin van Heel and Michael Schatz. “Fourier shell correlation threshold criteria”. In: *Journal of Structural Biology* 151.3 (2005), pp. 250–262. ISSN: 1047-8477. doi: [10.1016/j.jsb.2005.05.009](https://doi.org/10.1016/j.jsb.2005.05.009).
- [17] Sebastian Schmidt, Qing Rao, Julian Tatsch, et al. “Advanced Active Learning Strategies for Object Detection”. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020, pp. 871–876. doi: [10.1109/IV47402.2020.9304565](https://doi.org/10.1109/IV47402.2020.9304565).
- [18] Siteng Ma, Haochang Wu, Aonghus Lawlor, et al. “Breaking the Barrier: Selective Uncertainty-Based Active Learning for Medical Image Segmentation”. In: *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2024, pp. 1531–1535. doi: [10.1109/ICASSP48485.2024.10446026](https://doi.org/10.1109/ICASSP48485.2024.10446026).

This is the end of my presentation.

I sincerely appreciate your attention today.

Presenter: Yu-Hsuan, Li

Appendix

Potential limitation

- Potential Error in Particle Correction Process:
 - When two groundtruth labels, G_1 and G_2 , are very close or slightly overlap, an issue may arise if the model predicts particle P with high IoU values for both labels.
 - The process might assign the coordinates of P to G_1 , even though P may have a higher IoU with G_2 .
 - This could result in inaccurate corrections, where the model preferentially detects G_2 but assigns P to G_1 .

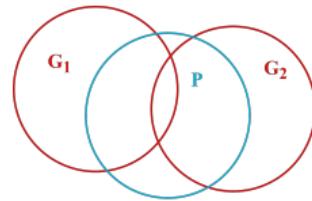


Figure 9-21: Schematic diagram of the issue

Previous weight exploration

- ① In our study, the model from the previous iteration was not used as pretrained weights due to its uncertain effectiveness.
- ② In 2024, Yamani et al. discussed in their paper "Active Learning for Single-Stage Object Detection in UAV Images" that training from scratch outperformed continuing from previous weights, especially in later stages of active learning.
- ③ They attributed this to overfitting in early iterations, which training from scratch avoids, enabling better improvement later.

Observation of Evaluation Metrics

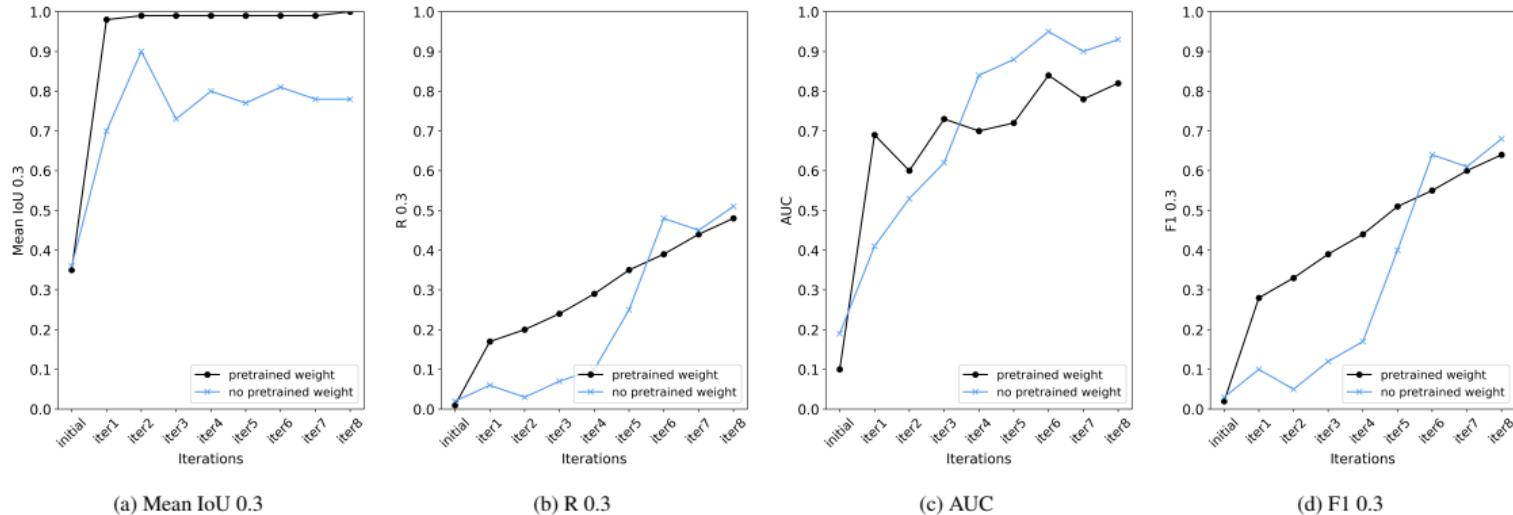


Figure 9-22: In the titles of (a), (b), and (d), "R" and "F1" denote Recall and F1-Score, respectively. The value of 0.3 that follows the title signifies the confidence threshold employed for the calculation of the corresponding metrics.

Observation of Evaluation Metrics

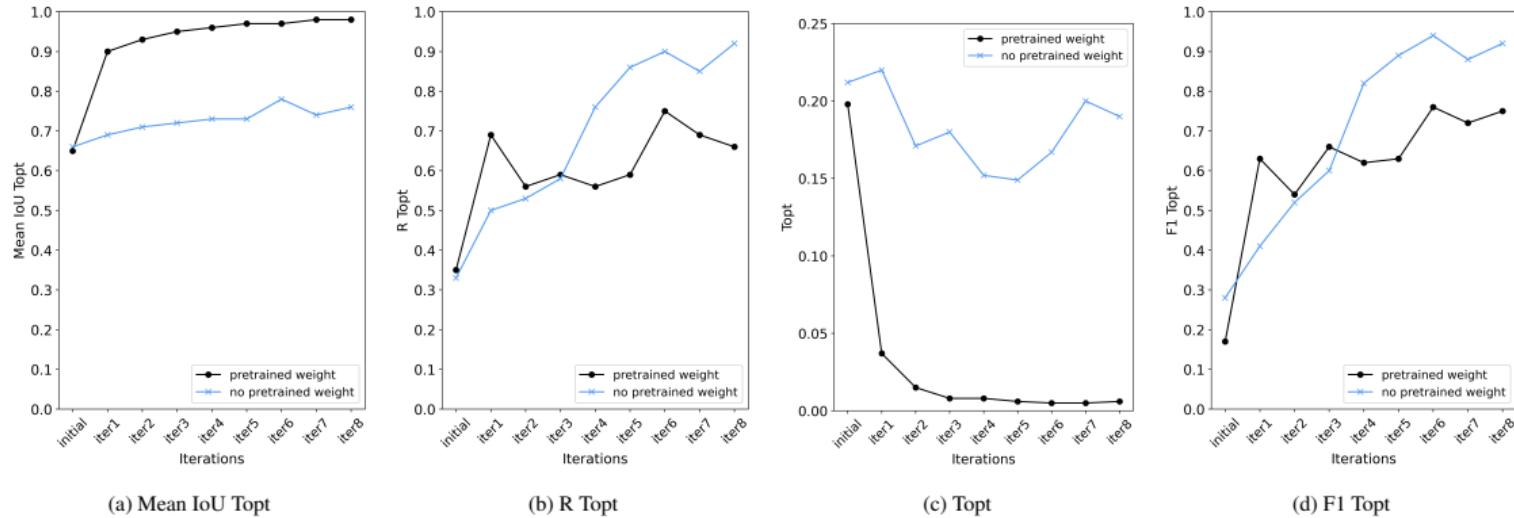


Figure 9-23: In the titles of (a), (b), and (d), the value of Topt that appears subsequent to the title signifies the confidence threshold employed for the calculation of the corresponding metrics. (c) Illustrates the Topt value used in each iteration.

i5-a8.75 distinction analysis

	Random	Low confidence	Entropy Score	Norm-conf-es	Boundary Distance
Random	0	1296	1287	1259	1273
Low confidence	1296	0	1398	1394	1340
Entropy Score	1287	1398	0	1054	1190
Norm-conf-es	1259	1394	1054	0	1227
Boundary Distance	1273	1340	1190	1227	0

Table 9-8: With the experimental setting "i5-a8.75, the distinct particle counts between methods used for the final training model.

i15-a7.5 distinction analysis

	Random	Low confidence	Entropy Score	Norm-conf-es	Boundary Distance
Random	0	1246	1205	1213	1217
Low confidence	1246	0	1325	1314	1338
Entropy Score	1205	1325	0	1105	1108
Norm-conf-es	1213	1314	1105	0	1080
Boundary Distance	1217	1338	1108	1080	0

Table 9-9: With the experimental setting "i15-a7.5, the distinct particle counts between methods used for the final training model.

Topt

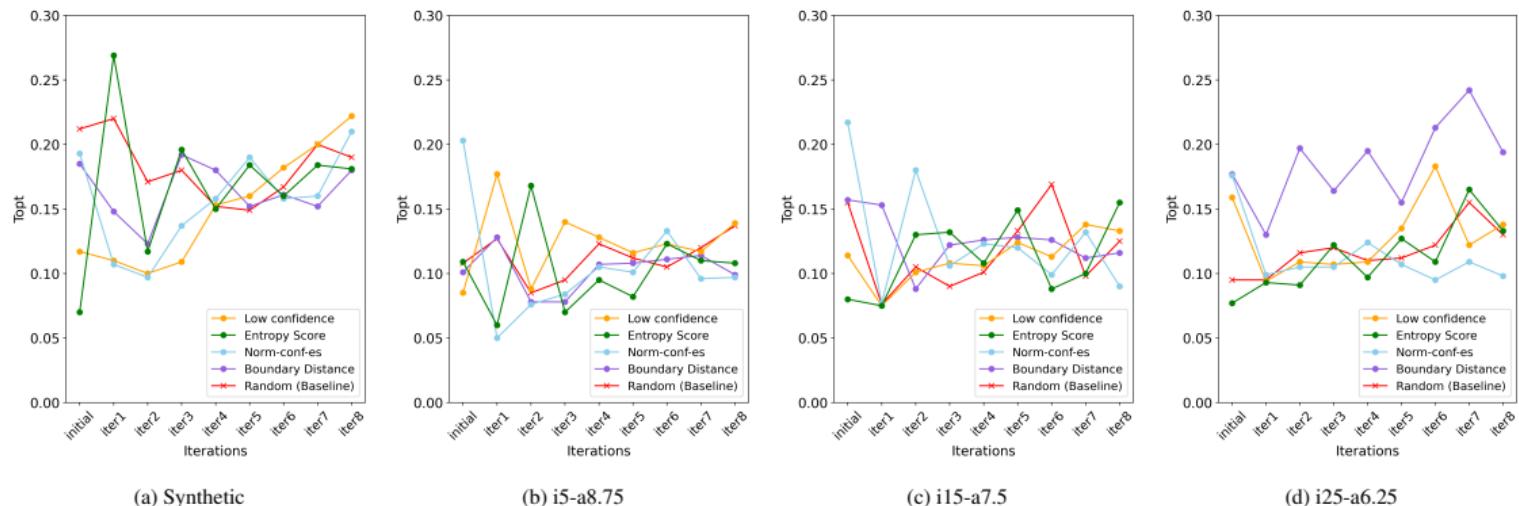


Figure 9-24: $Topt$ value for each iteration across all experimental setting.

i5-a8.75 Homogeneous Refinement

i5-a8.75							
	1	2	3	4	5	Mean	SD
Random	6.54 Å	6.55 Å	6.55 Å	6.63 Å	6.55 Å	6.564 Å	0.037 Å
Low confidence	6.44 Å	6.46 Å	6.4 Å	6.49 Å	6.43 Å	6.444 Å	0.034 Å
Entropy Score	6.49 Å	6.36 Å	6.49 Å	6.56 Å	6.58 Å	6.496 Å	0.086 Å
Norm-conf-es	6.59 Å	6.55 Å	6.53 Å	6.58 Å	6.51 Å	6.552 Å	0.033 Å
Boundary Distance	6.10 Å	6.29 Å	6.18 Å	6.12 Å	6.13 Å	6.164 Å	0.076 Å

Table 9-10: Reconstruct use with Homogeneous Refinement on the real dataset. The red values indicate the baseline, while the green values represent the best performance under the corresponding experimental settings, where "SD" denoting the standard deviation.

i15-a7.5 Homogeneous Refinement

i15-a7.5							
	1	2	3	4	5	Mean	SD
Random	6.48 Å	6.55 Å	6.52 Å	6.4 Å	6.41 Å	6.472 Å	0.066 Å
Low confidence	6.49 Å	6.16 Å	6.51 Å	6.54 Å	6.43 Å	6.426 Å	0.154 Å
Entropy Score	6.59 Å	6.53 Å	6.52 Å	6.58 Å	6.54 Å	6.552 Å	0.031 Å
Norm-conf-es	6.58 Å	6.75 Å	6.75 Å	6.77 Å	6.75 Å	6.720 Å	0.079 Å
Boundary Distance	6.08 Å	5.96 Å	6.09 Å	6.07 Å	6.06 Å	6.052 Å	0.053 Å

Table 9-11: Reconstruct use with Homogeneous Refinement on the real dataset. The red values indicate the baseline, while the green values represent the best performance under the corresponding experimental settings, where "SD" denoting the standard deviation.

i25-a6.25 Homogeneous Refinement

i25-a6.25							
	1	2	3	4	5	Mean	SD
Random	6.84 Å	6.63 Å	6.51 Å	6.78 Å	6.49 Å	6.650 Å	0.157 Å
Low confidence	6.38 Å	6.40 Å	6.07 Å	6.36 Å	6.34 Å	6.310 Å	0.136 Å
Entropy Score	6.39 Å	6.31 Å	5.95 Å	6.43 Å	6.11 Å	6.238 Å	0.203 Å
Norm-conf-es	6.38 Å	6.42 Å	6.48 Å	6.49 Å	6.48 Å	6.450 Å	0.048 Å
Boundary Distance	5.55 Å	5.55 Å	5.62 Å	5.61 Å	5.62 Å	5.590 Å	0.037 Å

Table 9-12: Reconstruct use with Homogeneous Refinement on the real dataset. The red values indicate the baseline, while the green values represent the best performance under the corresponding experimental settings, where "SD" denoting the standard deviation.