



上海科技大学
ShanghaiTech University



寰渺科技
Cellverse



CryoGEM: Physics-informed Generative Cryo-Electron Microscopy

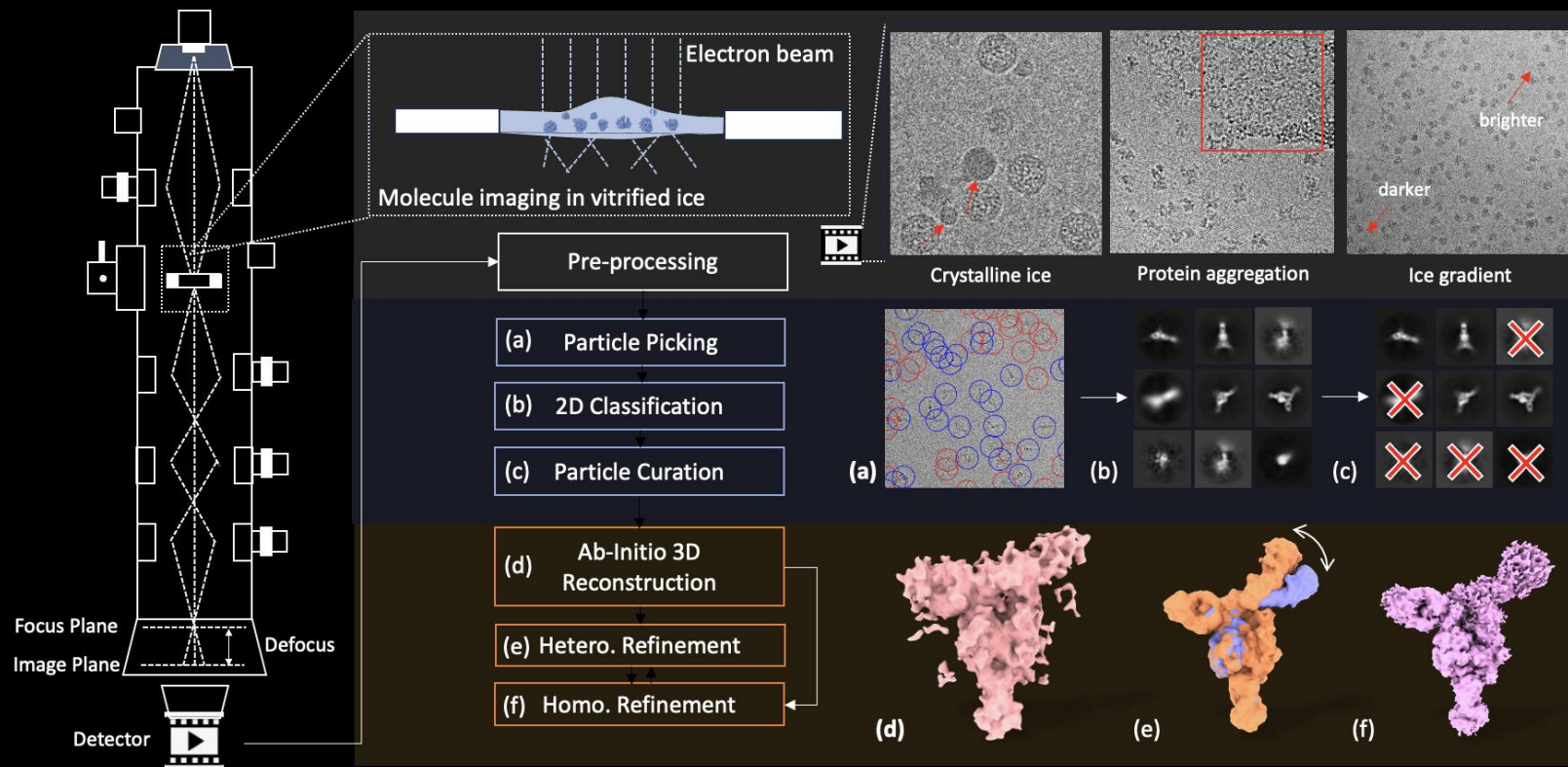
Jiakai Zhang^{*,1,2}, Qihe Chen^{*,1,2}, Yan Zeng^{1,2}, Wenyuan Gao¹,
Xuming He¹, Zhijie Liu^{1,3}, and Jingyi Yu^{1,†}

¹ShanghaiTech University ²Cellverse ³iHuman Institute

* Indicates Equal Contribution, † Indicates the corresponding author

Motivation: The Scarcity of Annotated Data in Cryo-EM

AI models have been widely used in the core tasks of cryo-EM pipeline. However, their performance are limited due to the lack of high-quality training datasets.

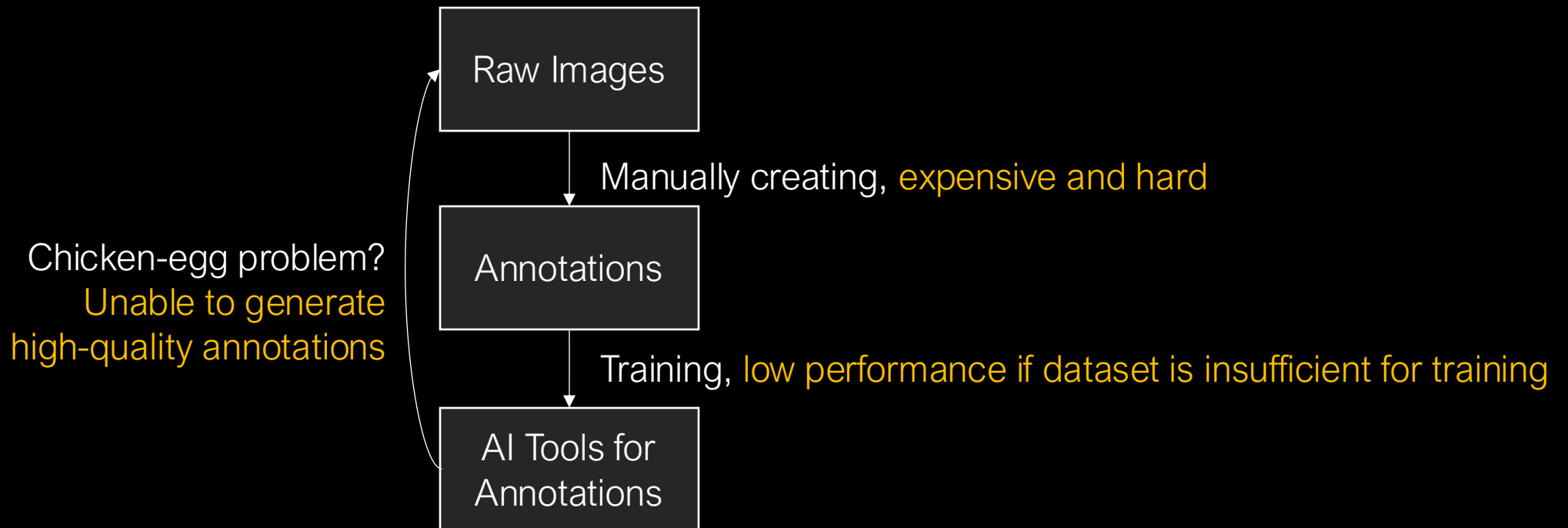


A comprehensive Cryo-EM pipeline for high-resolution reconstruction of proteins. (a) and (d) have the existing AI methods to tackle them.

Motivation: Annotations are Expensive in Cryo-EM

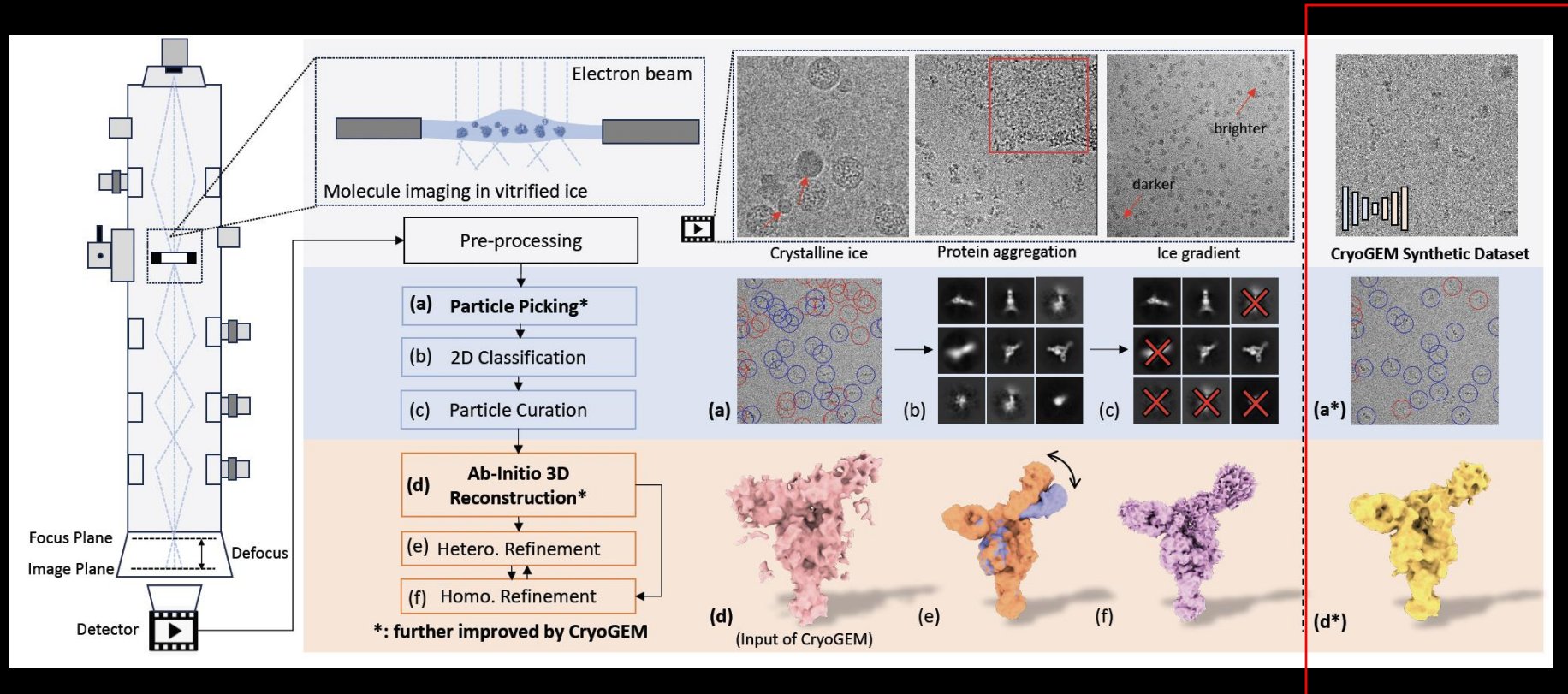
- Expensive to manually creating annotations: **picking** millions of particles, further **segmenting** them and even **determine** particles' orientation and conformation.
- Can we get annotated data effortlessly?
 - Physical simulation is possible, but it is computationally-intensive and unrealistic for cryo-EM.

Generative model is an alternative.



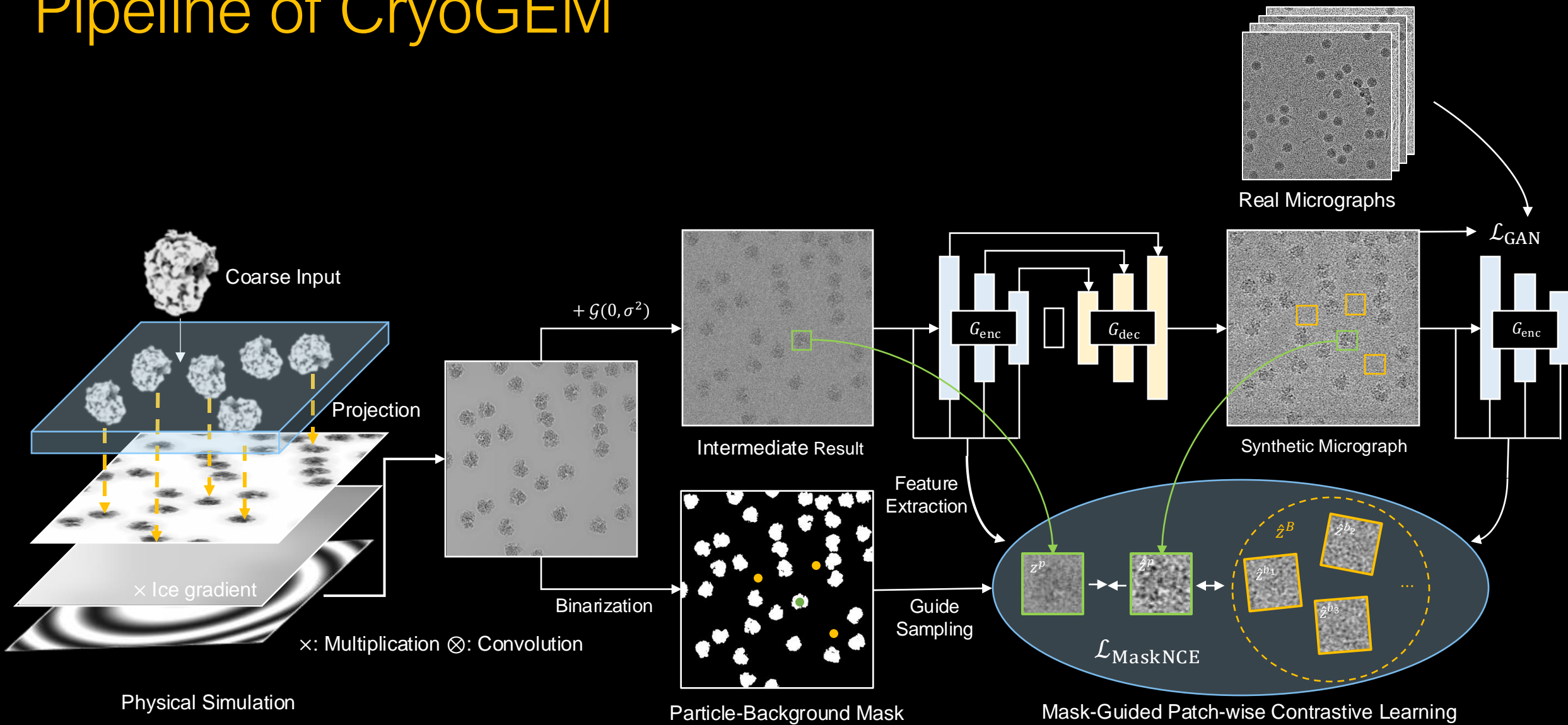
CryoGEM: Generative Cryo-Electron Microscopy

- The **first generative model** for micrograph and particle synthesis in cryo-EM
- CryoGEM can be **quickly trained** by only 100 real micrographs with an coarse 3D result in two hours
- The generated data are used to **pre-train** downstream AI models

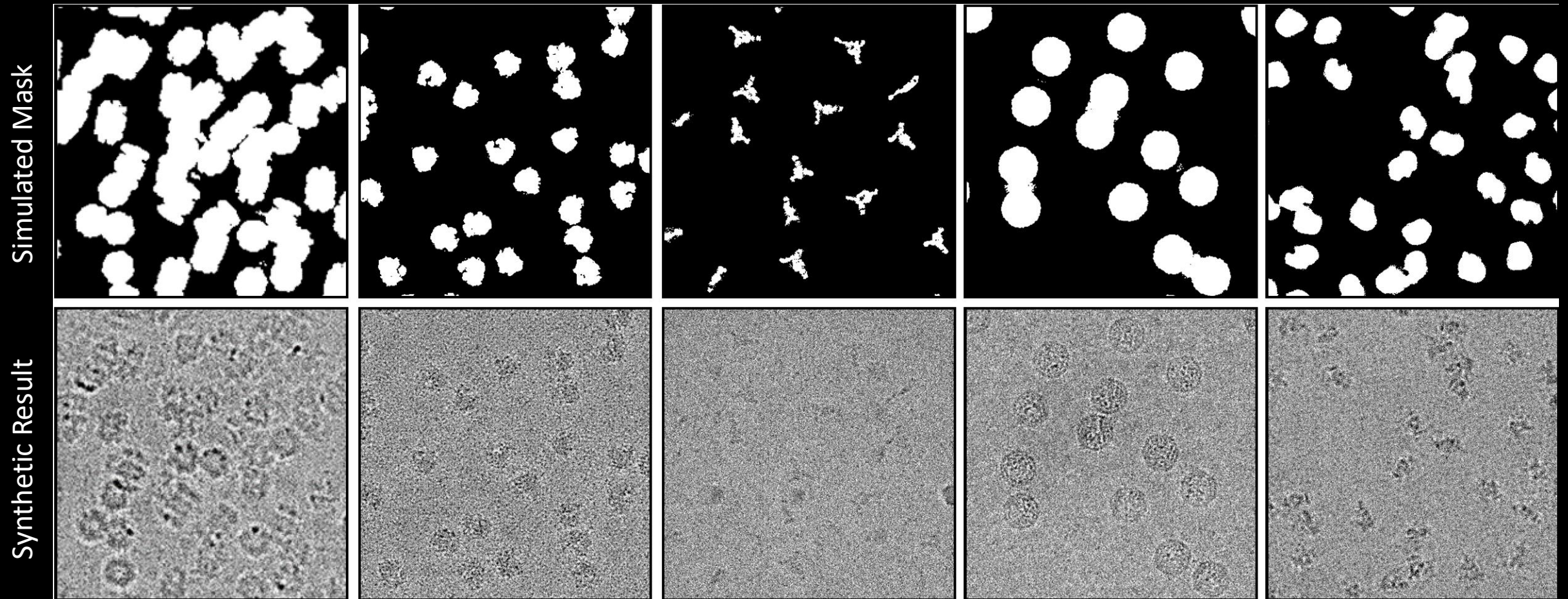


Improved by CryoGEM's annotations

Pipeline of CryoGEM

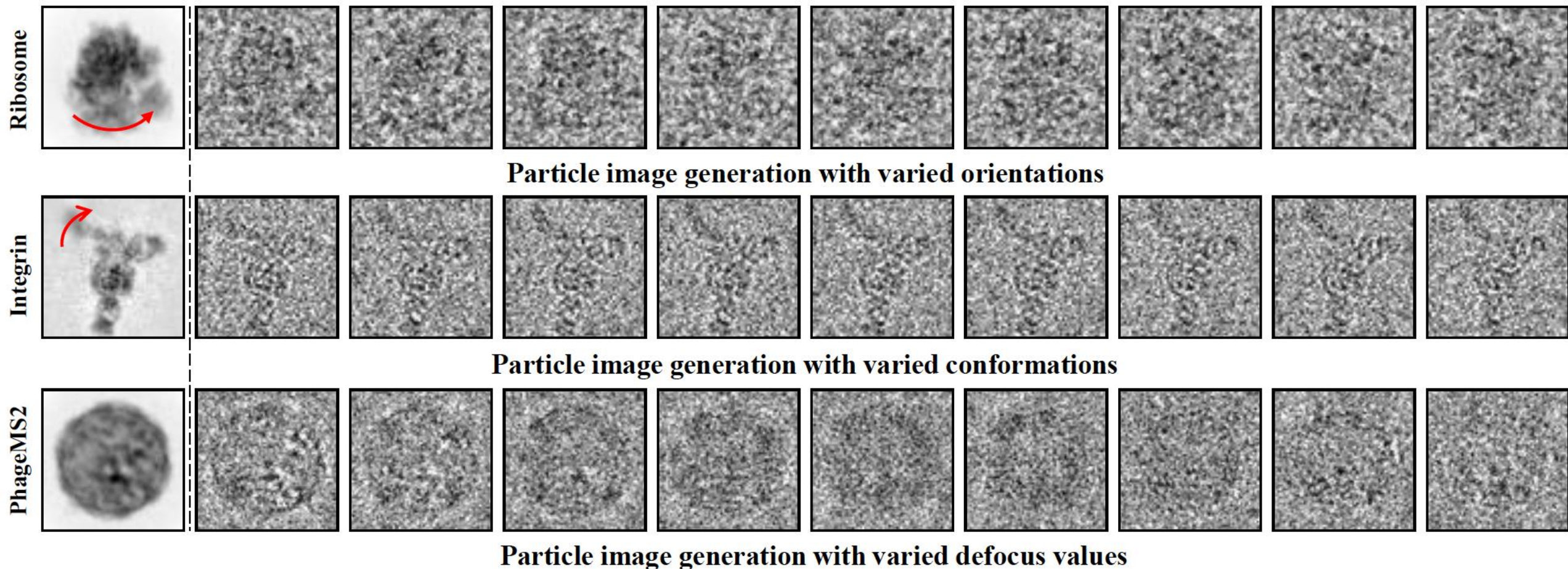


Micrograph Generation with Position Controls



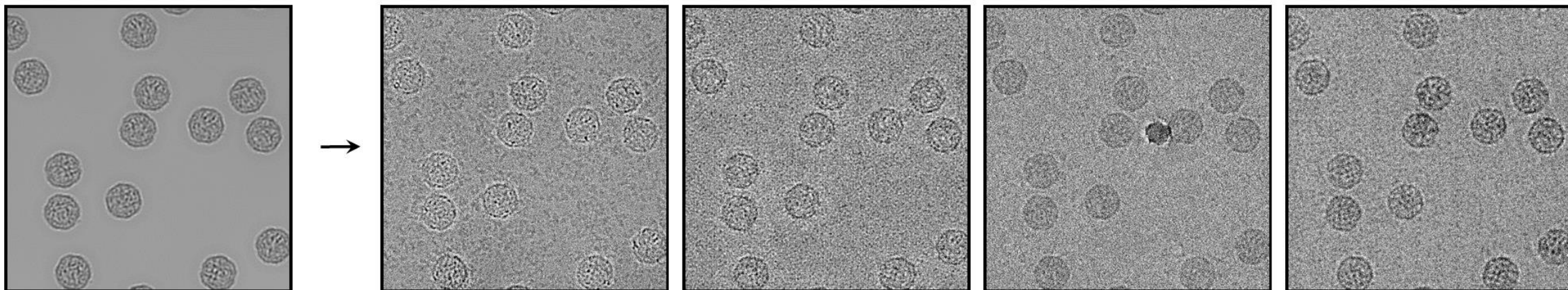
CryoGEM can generate the synthetic images that fits its particle masks

Particle Generation with Different Orientation, Conformation and Defocus

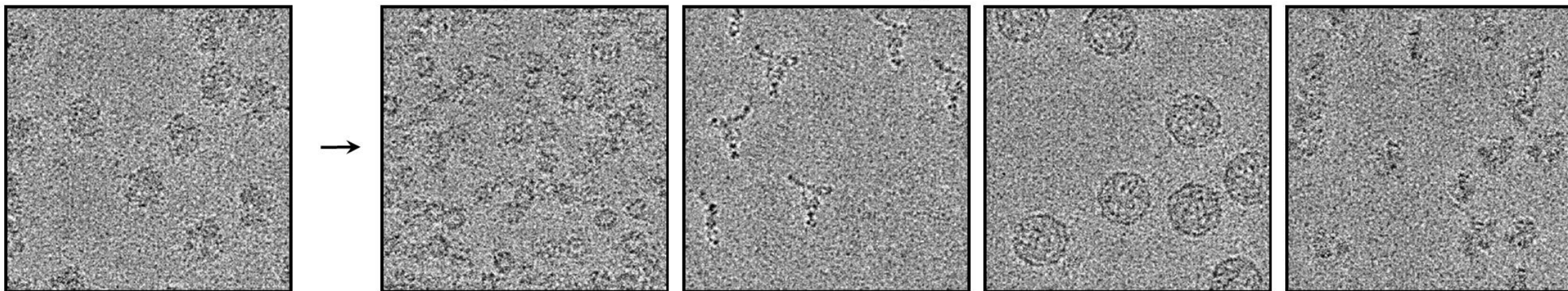


Zero-shot Particle/Noise Transfer to Different EM Setups

Zero-shot Results



(c) Zero-shot particle transfer from PhageMS2 to other datasets



(d) Zero-shot noise transfer from Ribosome to other datasets

Particle Picking

- Topaz: an AI model for particle picking, trained in several real datasets.
- Pre-train Topaz on synthetic data baselines and **tested on real datasets**
- Compared with both the **original Topaz**, CryoGEM significantly improves the picking accuracy and **the final reconstruction resolution**

	Topaz	Ours	Poi-Gau	CycleGAN	CUT	CycleDiffusion
Ribosome						
Integrin						

Metric	AUPRC \uparrow						Res(\AA) \downarrow					
	Proteasome	Ribosome	Integrin	PhageMS2	HumanBAF	Avg.	Proteasome	Ribosome	Integrin	PhageMS2	HumanBAF	Avg.
Gaussian	0.471	0.485	0.259	0.886	0.470	0.514	2.76	3.84	7.62	7.44	10.29	6.39
Poisson	0.469	0.375	0.213	0.490	0.483	0.406	2.77	4.16	7.52	10.50	10.67	7.12
Poi-Gau	0.455	0.618	0.210	0.706	0.381	0.474	2.80	3.87	8.13	10.90	13.18	7.77
CycleGAN	0.200	0.308	0.414	0.895	0.228	0.409	5.10	3.93	8.03	7.51	11.83	7.28
CUT	0.442	0.224	0.335	0.592	0.513	0.421	2.77	4.78	7.16	8.91	12.92	7.30
CycleDiffusion	0.346	0.205	0.233	0.469	0.392	0.329	2.92	4.56	6.35	9.51	6.68	6.66
Topaz	0.302	0.679	0.526	0.329	0.493	0.466	3.13	3.84	6.34	10.67	9.59	6.71
Ours	0.490	0.797	0.606	0.915	0.562	0.674	2.68	3.25	5.54	7.16	7.74	5.27

Pose Estimation

- **CryoFIRE**: an AI model for ab-initio reconstruction in a self-supervised manner
- We pre-train CryoFIRE's pose estimation module on synthetic data generated from different baselines and **tested on real datasets**
- Compared with both the **original CryoFIRE and other baselines**, we achieve the best reconstruction resolutions and the minimal rotation errors.

Metric	Res(px)↓						Rot.(rad)↓					
Dataset	Proteasome	Ribosome	Integrin	PhageMS2	HumanBAF	Avg.	Proteasome	Ribosome	Integrin	PhageMS2	HumanBAF	Avg.
Gaussian	2.97	4.59	7.01	5.85	6.82	5.44	0.48	0.50	1.20	0.64	1.49	0.88
Poisson	3.02	4.912	7.01	5.98	8.03	5.79	1.20	0.90	1.19	0.69	1.47	1.10
Poi-Gau	3.02	4.39	9.06	5.90	8.12	6.09	1.05	0.39	1.40	0.61	1.43	0.98
CycleGAN	3.02	4.74	6.24	5.71	6.91	5.32	0.46	0.61	1.55	0.74	1.48	0.97
CUT	2.66	5.40	6.13	6.03	9.58	5.96	0.44	1.15	1.53	0.66	1.45	1.05
CycleDiffusion	3.61	5.79	8.5	5.91	9.33	6.62	0.44	1.42	1.55	0.58	1.53	1.10
CryoFIRE	5.94	16.92	13.87	17.23	6.98	12.18	1.55	0.64	0.93	0.75	1.53	1.08
Ours	2.59	4.27	4.88	5.54	6.56	4.29	0.41	0.32	0.88	0.43	1.42	0.69

Discussion & Future Work

- We have presented the first **generative Cryo-EM, CryoGEM**, which can be trained by only 100 raw micrographs and an initial reconstruction result. CryoGEM can synthesize **high-quality annotated dataset** that can be used by downstream AI models in cryo-EM.
- Limitations:
 - Depend on a coarse reconstructed result as an input, can be replaced by the structure prediction models.
- Future directions:
 - A **generalized version** of cryoGEM, more control parameters, better performance.
 - Explore more downstream tasks, such as **denoising and pose refinement**.
- See you at the poster session!