DRACO: A Denoising-Reconstruction Autoencoder for Cryo-EM

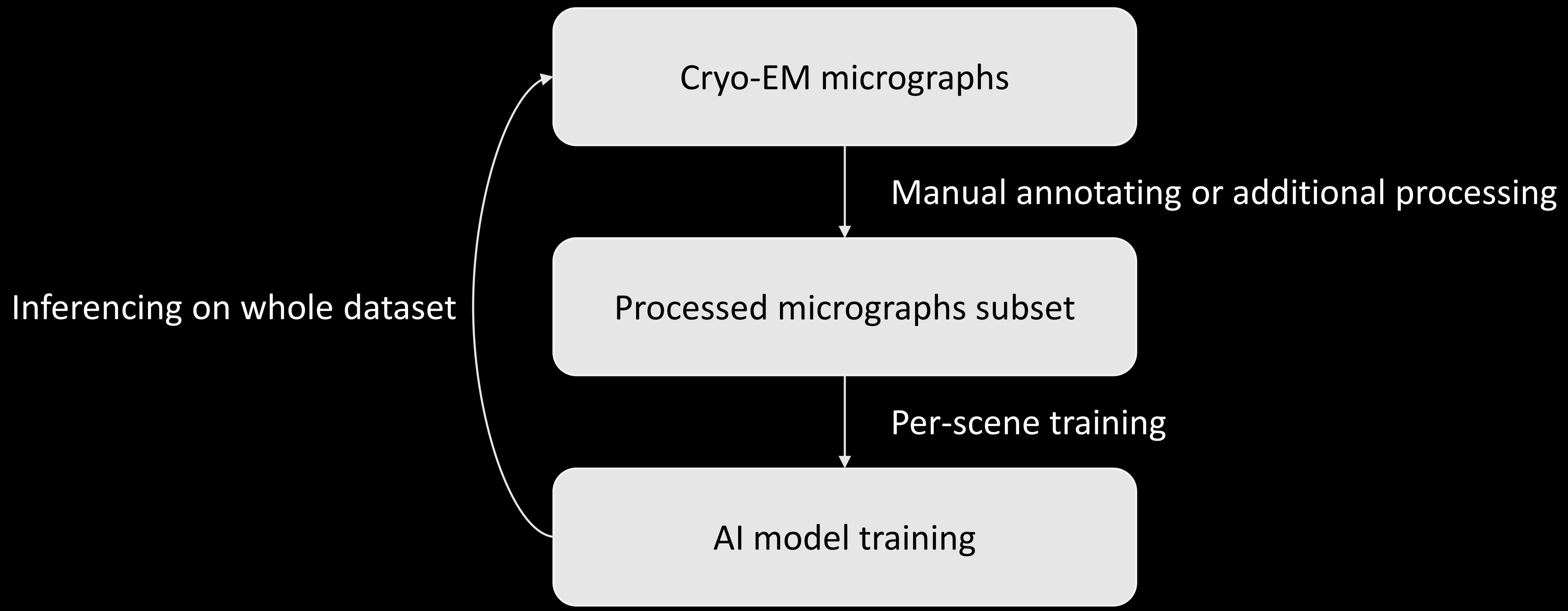
Yingjun Shen[*,1,2], **Haizhao Dai**[*,1,2], Qihe Chen[1,2], Yan Zeng[1,2], Jiakai Zhang[1,2], Yuan Pei[1,3], and Jingyi Yu[1,†]

[1]ShanghaiTech University [2]Cellverse [3]iHuman Institute

[*] Indicates Equal Contribution, [†] Indicates the corresponding author

Motivation: The time-consuming workflow in Cryo-EM

AI models have been widely used in the core tasks of cryo-EM pipeline. However, these methods often require per-scene training with addition process or manual annotations.



Can we skip these time-consuming processing or training steps?

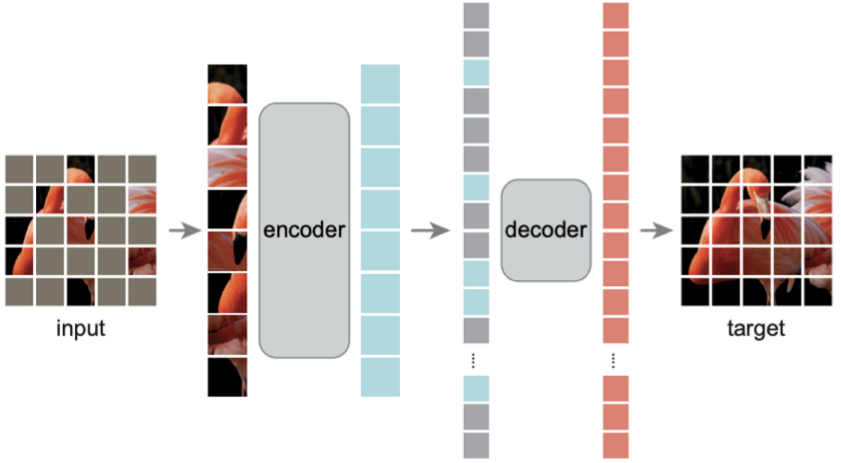
- A more general model is needed

Masked Autoencoders and Masked Feature Prediction

MAE: Using pixel values as targets also works great

- The encoder is applied to visible patches, mask tokens are introduced after this

MaskFeat: Other image features can be used as targets as well



method	pre-train data	AP^{box} ViT-B	AP^{box} ViT-L	AP^{mask} ViT-B	AP^{mask} ViT-L
supervised	IN1K w/ labels	47.9	49.3	42.9	43.9
MoCo v3	IN1K	47.9	49.3	42.7	44.0
BEiT	IN1K+DALLE	49.8	**53.3**	44.4	47.1
MAE	IN1K	**50.3**	**53.3**	**44.9**	**47.2**

Table 4. **COCO object detection and segmentation** using a ViT

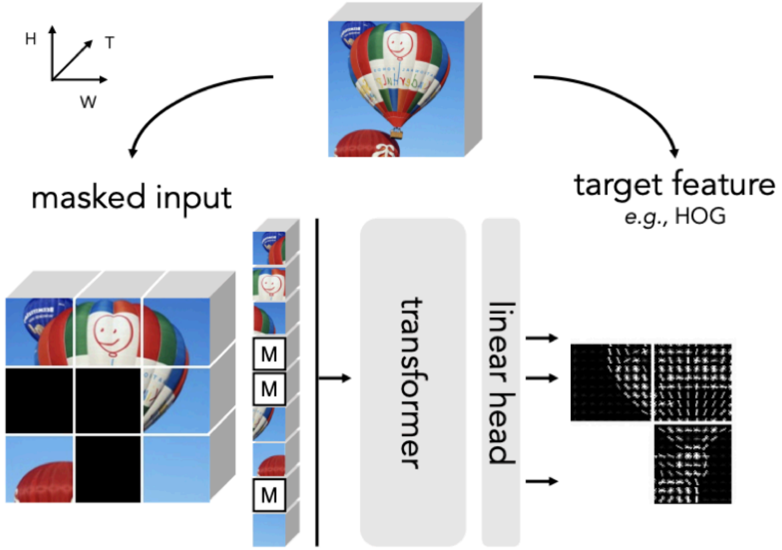


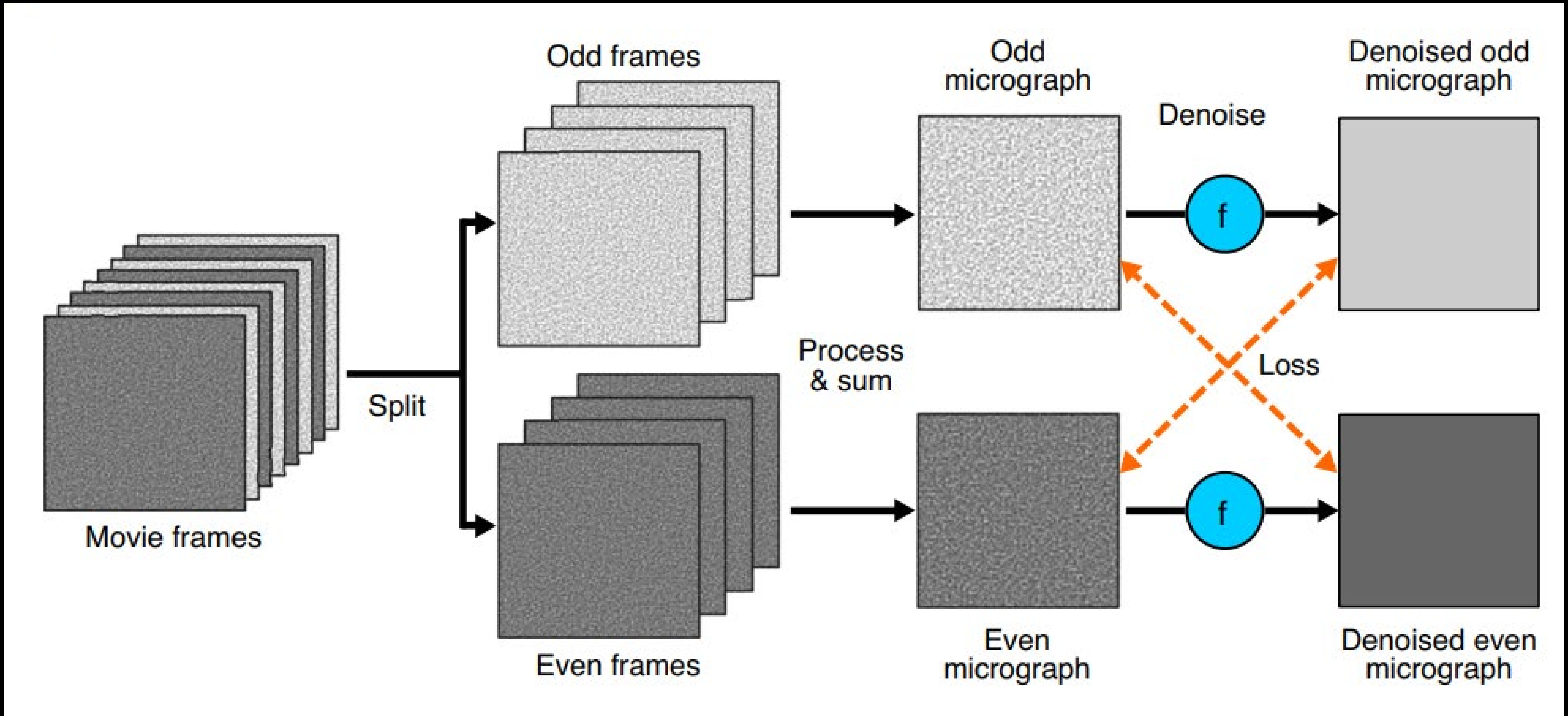
Figure 2. **MaskFeat pre-training.** We randomly replace the input space-time cubes of a video with a [MASK] token and directly regress features (*e.g.* HOG) of the masked regions. After pre-training, the Transformer is fine-tuned on end tasks.

[1] Masked Autoencoders Are Scalable Vision Learners, CVPR 2022 [2] Masked feature prediction for self-supervised visual pre-training., CVPR 2022

Noise2Noise in Cryo-EM

Topaz denoise: use the odd-even paired micrographs

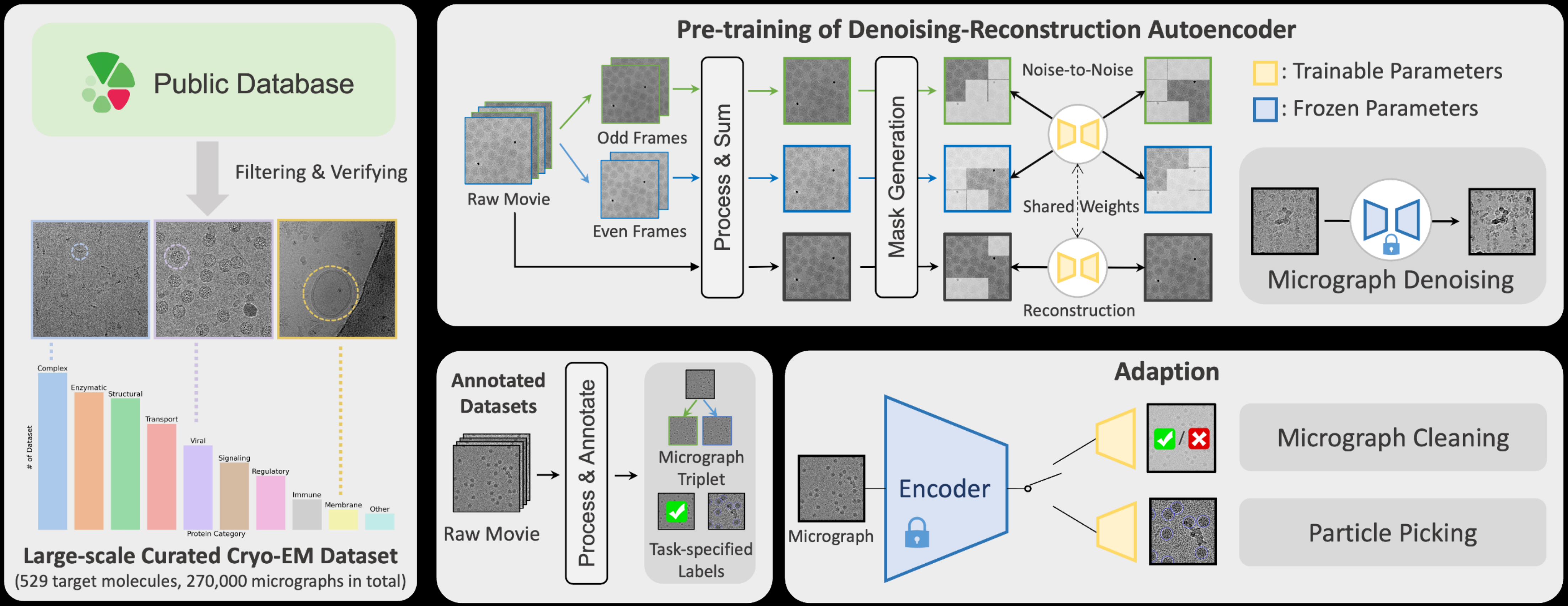
- With assumption of zero-mean noise, the network converges to predict expectation of micrographs, which is the signal value.
- Topaz denoise is only designed for denoising, lacking the ability of extracting general features.



[1] Topaz-denoise: general deep denoising models for cryoem and cryoet. Nature communications, 11(1):5208, 2020

DRACO: A Denoising-Reconstruction Autoencoder

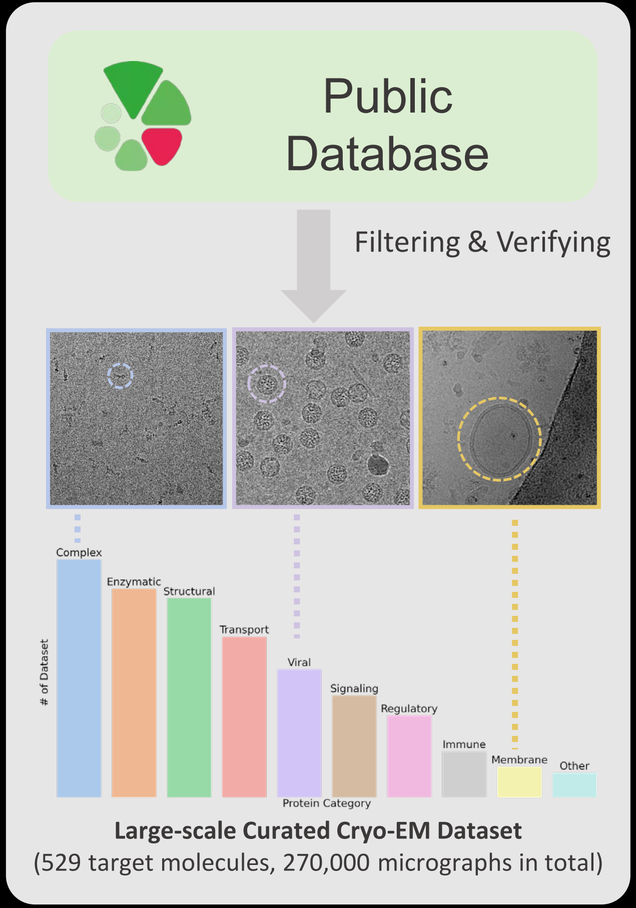
- The first foundation model for Cryo-EM trained on a large-scale curated dataset (~270,000 noisy pairs of cryo-EM images, ~100T raw data)
- Diverse downstream tasks: denoising, particle Picking, 3D Reconstruction…

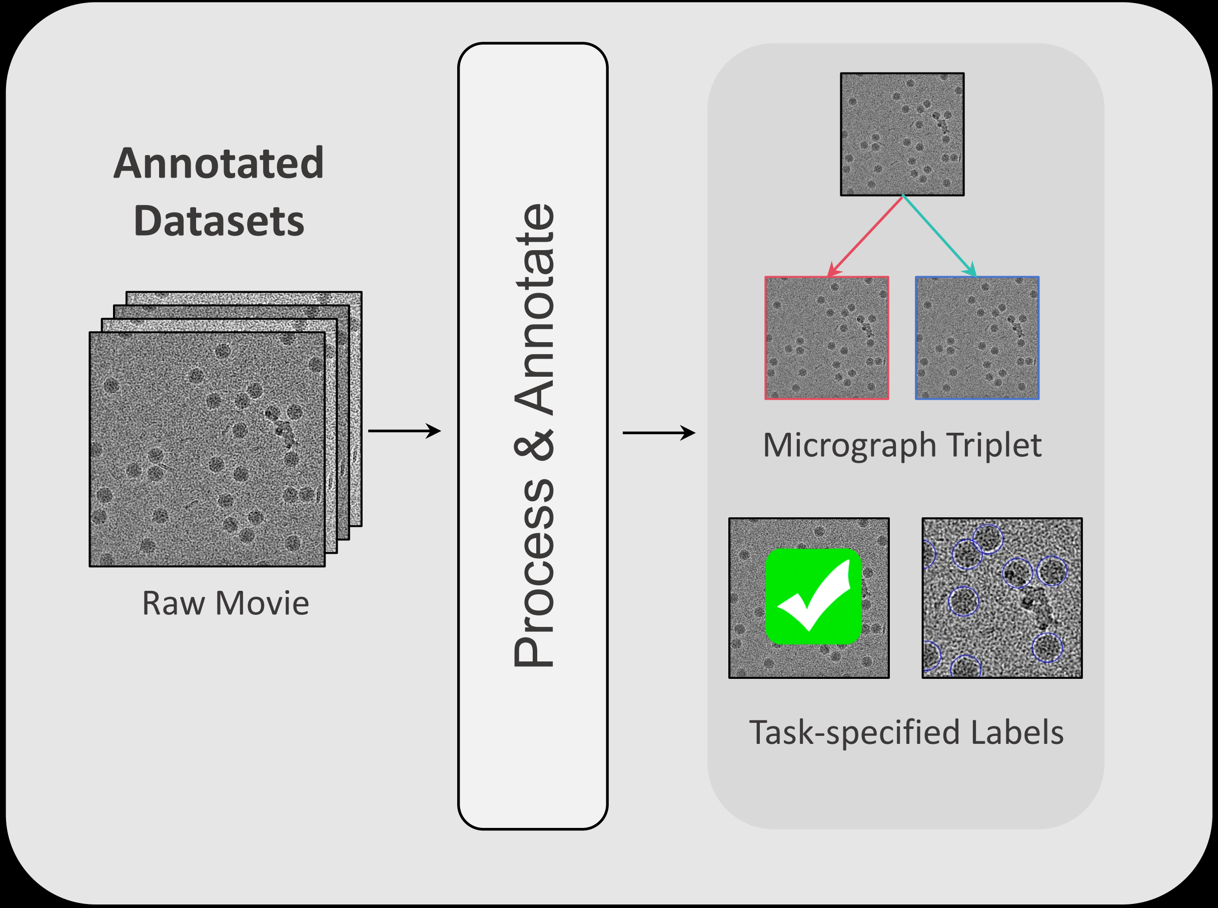


Large-scale dataset

We construct a large-scale, high-quality, and diverse single-particle cryo-EM dataset.

- Pretraining: ~270,000 odd-even-full micrograph triplets generated by Cryosparc.
- Particle picking: ~80000 micrographs with 8 million particles annotation generated by Cryosparc.
- Micrograph cleaning: 1194 micrographs with manually generated binary labels.

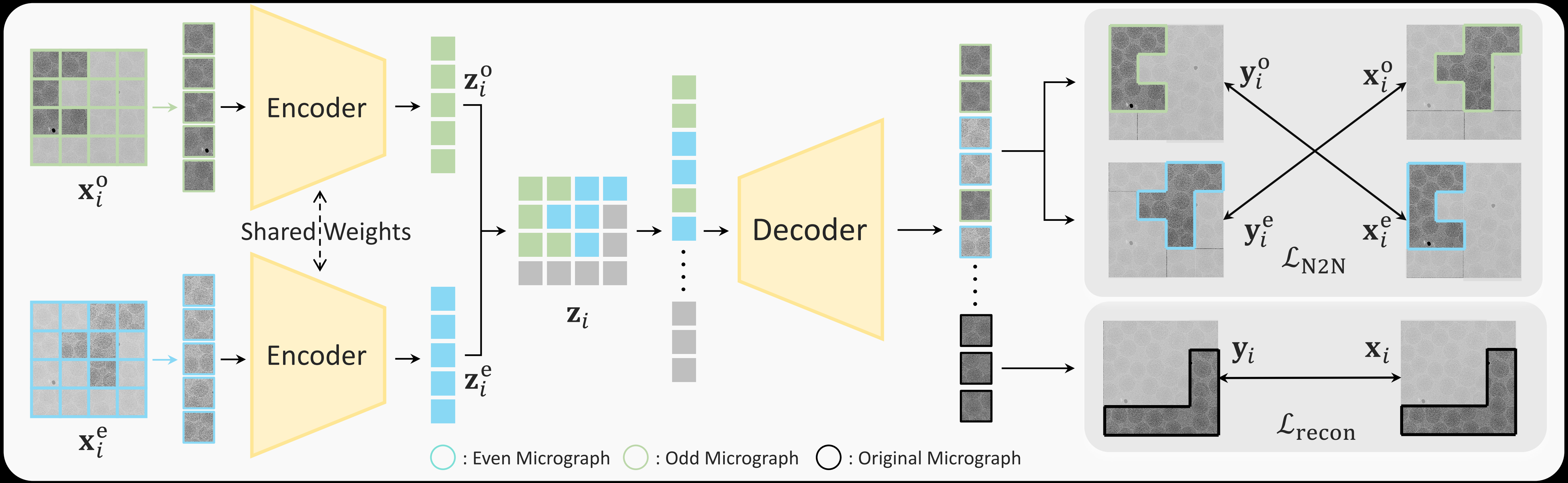




DRACO method

DRACO applies a denoising-reconstruction hybrid training scheme.

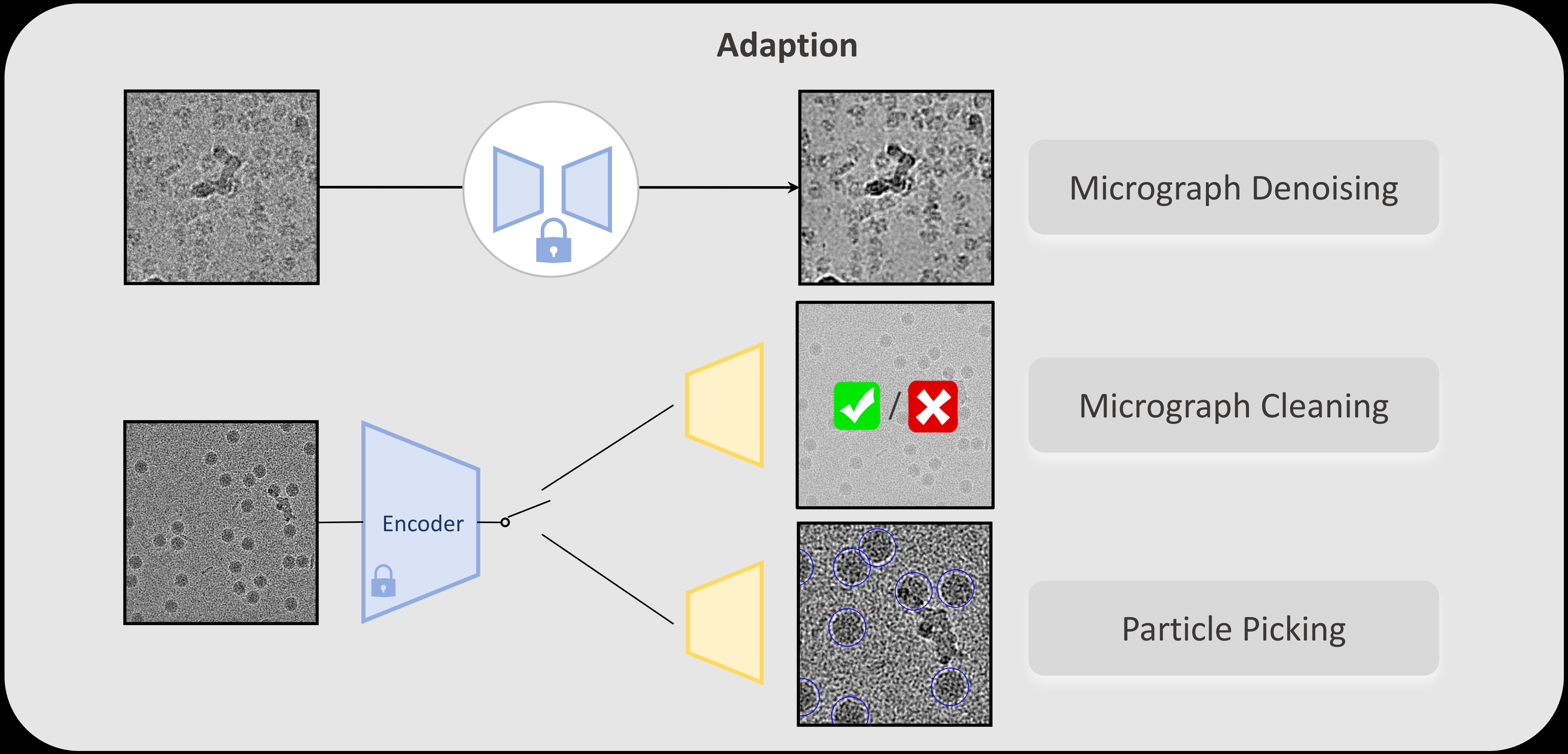
- The encoder takes odd-visible patches and even-visible patches as inputs.
- The N2N loss is applied on odd-even paired patches, which is inspired by Topaz-denoise.
- The reconstruction loss is applied to both invisible predicted patches, following the original MAE.



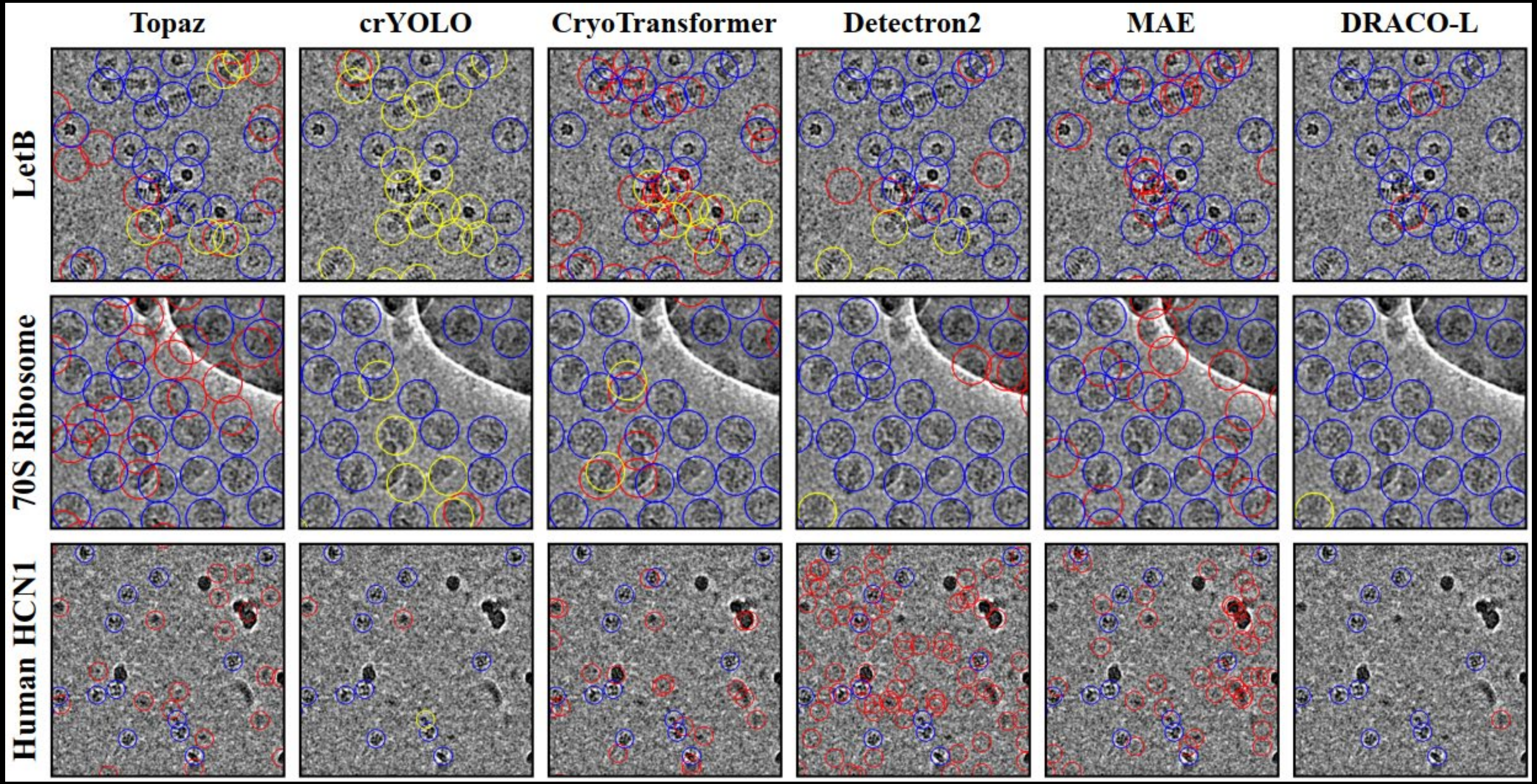
DRACO downstream task

Once pre-trained, our model can adapt to various downstream tasks.

- DRACO can naturally serve as a generalizable denoiser without any further fine-tuning.
- DRACO can also adapt to micrograph cleaning and particle picking with fine-tuning.



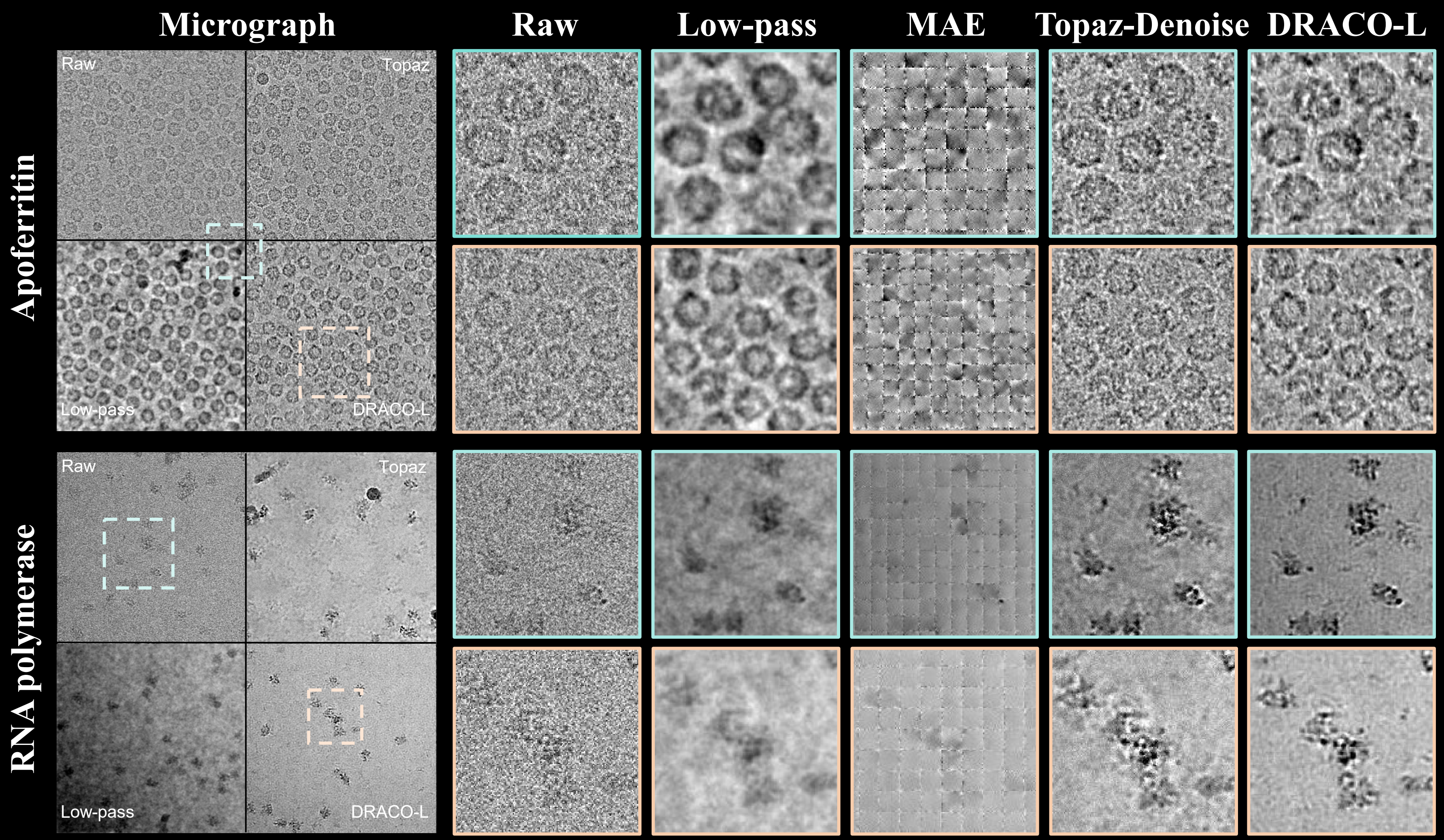
Generalized Cryo-EM Particle Picking



(Top) Blue indicates correctly picked particles, red and yellow indicate false positives and false negatives.
(Bottom) Particle picking metrics: Our method outperforms existing methods on the test dataset.

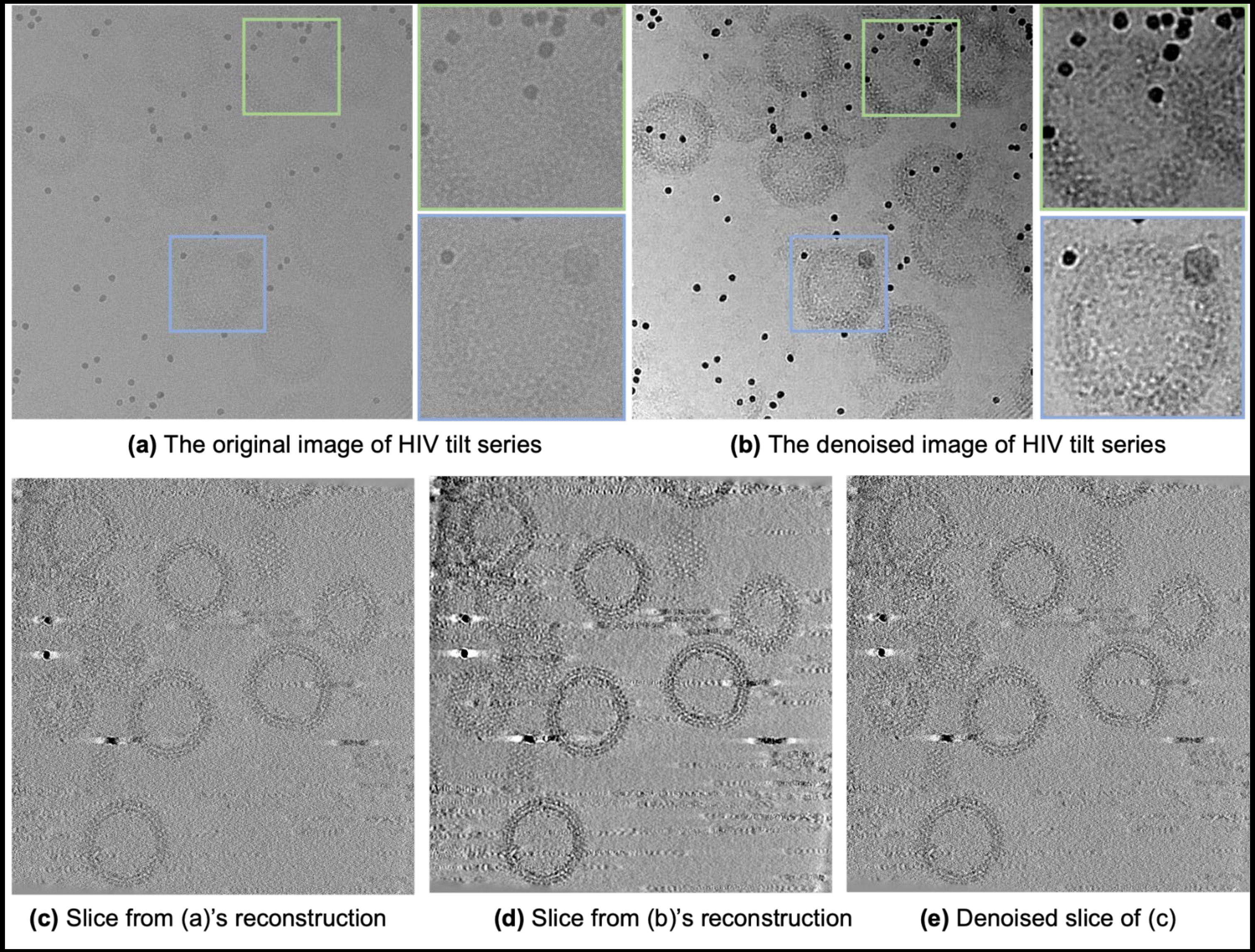
	Human HCN1				70S ribosome				LetB			
Method	Precision (↑)	Recall (↑)	F1 score (↑)	Res. (↓)	Precision	Recall	F1 score	Res.	Precision	Recall	F1 score	Res.
Topaz	0.462	**0.956**	0.623	4.20	0.362	**0.943**	0.523	2.80	0.518	0.761	0.617	3.67
crYOLO	**0.818**	0.748	0.782	4.15	0.602	0.869	0.711	2.78	0.632	0.163	0.224	4.62
CryoTransformer	0.475	0.910	0.624	4.13	0.517	0.887	0.654	2.79	0.429	0.706	0.534	3.67
Detectron	0.392	0.834	0.533	4.50	0.668	0.901	0.767	2.85	0.589	0.804	0.680	3.86
MAE	0.703	0.649	0.675	4.32	0.712	0.876	0.786	2.84	0.591	**0.805**	0.682	4.03
DRACO-B	0.768	0.799	0.793	4.03	0.732	0.905	0.810	2.61	0.637	0.779	0.701	3.55
DRACO-L	0.830	0.802	**0.816**	**3.90**	**0.803**	0.846	**0.824**	**2.51**	**0.678**	0.780	**0.725**	**3.53**

Generalized Cryo-EM denoising

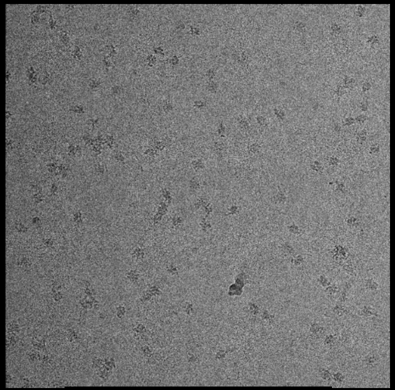


Generalized Cryo-ET denoising

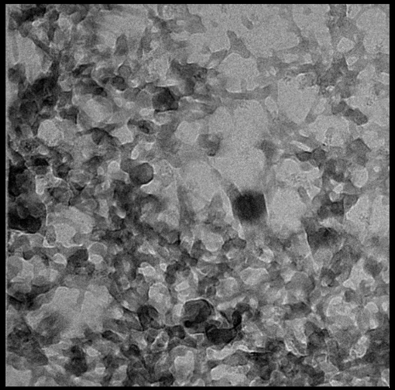
Although not pre-trained on Cryo-ET datasets, our model can be directly applied to Cryo-ET tilt series.



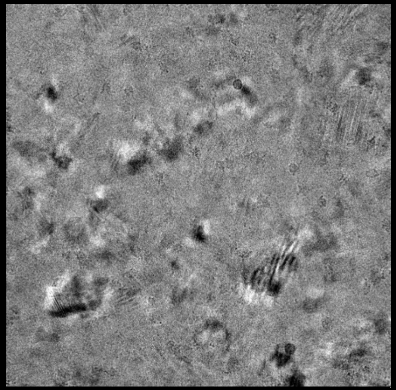
Generalized Cryo-EM micrograph curation



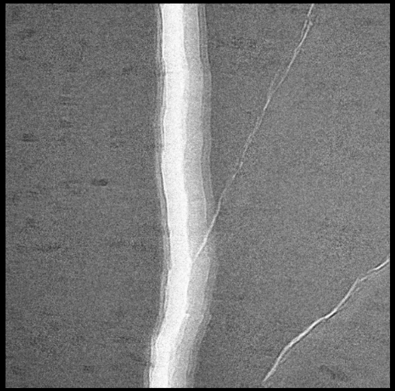
normal sample



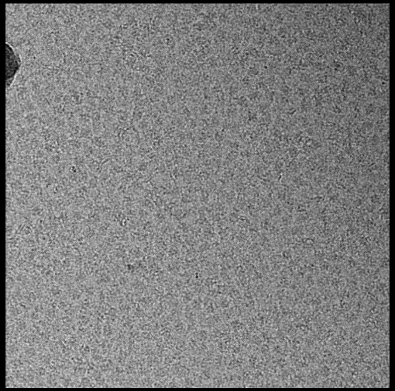
ice contamination



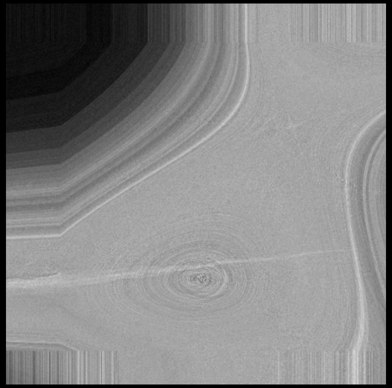
ice crystalline



crack sample



bad film coverage



multiple issues

Low-quality micrographs may arise from various artifacts and should be filtered.

DRACO can easily adapt to this 2-class classification task by linear probing.

Method	Accuracy	Precision	Recall	F1 score
Miffi	0.836	0.899	0.845	0.871
ResNet18	0.938	0.923	0.960	0.940
MAE	0.904	0.927	0.892	0.909
DRACO-B	0.963	**0.976**	0.953	0.964
DRACO-L	**0.983**	**0.976**	**0.992**	**0.984**

Future work

- Future direction:
  - Design a more comprehensive denoising tasks for movies
  - A more diverse datasets, including Cryo-ET datasets.
  - Extend DRACO to particle level, supporting particle tasks such as pose estimation.
- See you in poster session