

# Yang Yan

Ph.D. Candidate, Zhejiang University / Westlake University

AI for Science | Computer Vision | Representation Learning | Structural Biology

y.yan@zju.edu.cn yanyang92@westlake.edu.cn

<https://yanyang1998.github.io/> Google Scholar



## Research Profile

**AI-for-science Ph.D. candidate** developing foundation models and scalable ML systems for cryo-EM particle analysis and heterogeneous reconstruction. First author of a *Nature Methods* study introducing **Cryo-IEF**, the first foundation model for cryo-EM particle processing, pretrained on **65.3M particle images** and transferred to structural classification, pose-aware clustering, particle-quality ranking, and automated reconstruction. Current work develops CryoDECO, a foundation-prior framework for compositional and conformational heterogeneity in cryo-EM. Longer term, my research aims to build structural world models and agentic reconstruction systems that reason over molecular identity, dynamics, and native cellular context.

## Education

2023–Present	<b>Ph.D. Candidate in Computer Science</b> Joint Ph.D. program supervised by Prof. Fajie Yuan and Prof. Huaizong Shen. Research: foundation models for cryo-EM image processing, automated workflows, and heterogeneity analysis.	Zhejiang University / Westlake University
2020–2023	<b>M.S. in Translational Medicine (Engineering)</b> National Institute of Diagnostics and Vaccine Development in Infectious Diseases, supervised by Prof. Ningshao Xia. Research: machine learning for medical image analysis.	Xiamen University
2016–2020	<b>B.S. in Electrical Information Engineering</b> Research included machine-learning applications in localization.	Northeastern University (Qinhuangdao)

## Selected Publications

**A comprehensive foundation model for cryo-EM image processing** 2026

**Yang Yan**, Shiqi Fan, Fajie Yuan, Huaizong Shen

*Nature Methods*, 23(1), 88–95. DOI: [10.1038/s41592-025-02916-8](https://doi.org/10.1038/s41592-025-02916-8). Introduced Cryo-IEF as the first cryo-EM particle-processing foundation model and established a foundation-model paradigm for automated particle analysis.

**Artificial intelligence foundation model automates cryo-EM structure determination** 2026

**Yang Yan**, Huaizong Shen

*Nature Methods Research Briefing*, 23, 26–27, DOI: [10.1038/s41592-025-02917-7](https://doi.org/10.1038/s41592-025-02917-7).

**CryoDECO: Deconstructing Extreme Compositional and Conformational Heterogeneity in Cryo-EM via Foundation Model Priors** 2026

**Yang Yan**, Yanwanyu Xi, Shiqi Fan, Yifei Wang, Ziyun Tang, Fajie Yuan, Huaizong Shen

LangTaoSha Preprint Server. DOI: [10.65215/LTSpreprints.2025.12.30.000075](https://doi.org/10.65215/LTSpreprints.2025.12.30.000075). SOTA foundation-prior model for heterogeneous cryo-EM reconstruction, advancing high-throughput panoramic structural biology.

**Unsupervised cryo-EM images denoising and clustering based on deep convolutional autoencoder and k-means++** 2022

D. Zhang, **Yang Yan**, Y. Huang, B. Liu, Q. Zheng, J. Zhang, N. Xia

*IEEE Transactions on Medical Imaging*, 42(5), 1509–1521. Deep autoencoder and balanced K-means++ framework for cryo-EM denoising and clustering.

**Machine learning for structure determination in single-particle cryo-electron microscopy: A systematic review** 2021

J. G. Wu, **Yang Yan**, D. X. Zhang, B. W. Liu, Q. B. Zheng, X. L. Xie, S. Q. Liu, S. X. Ge, et al.

*IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 452–472. Review of ML methods for single-particle cryo-EM structure determination.

**Selected earlier medical-imaging publications** 2022–2023

S. Li, B. Liu, S. Li, X. Zhu, **Yang Yan**, D. Zhang; D. Zhang, B. Liu, Y. Huang, **Yang Yan**, et al.

*Complex & Intelligent Systems*, 2022; *Journal of Digital Imaging*, 2023. Deep-learning and ordinal-regression methods for bone-age assessment.

## Research Highlights

---

- **First cryo-EM foundation model:** developed **Cryo-IEF**, a foundation model developed for cryo-EM particle processing which is pretrained on **~65M particle images**. Cryo-IEF learns generalizable visual representations that can be transferred to classification, clustering, quality ranking, and 3D reconstruction, introducing the first foundation model paradigm for cryo-EM particle processing.
- **Automated reconstruction system:** developed **CryoWizard**, a fully automated computational pipeline for single-particle cryo-EM reconstruction that streamlines raw data processing, particle ranking, and high-resolution 3D structure determination without manual intervention.
- **Foundation-prior heterogeneous reconstruction:** developed **CryoDECO** to use pretrained particle representations as priors for ab initio heterogeneous reconstruction, targeting compositional and conformational mixtures.
- **Open-source Python package for cryo-EM data processing:** developed **cryodata**, a PyPI-installable data layer with **over 18.2k downloads** (according to pepy.tech) as of May 2026, converting CryoSPARC particle jobs into PyTorch-ready datasets with MRC/MRCS preprocessing, FFT/Hartley representations, LMDB storage, balanced sampling, and CryoSPARC-to-RELION metadata conversion.

## Research Software and Open Source

---

2023–2026	<b>Cryo-IEF: First Foundation Model for Cryo-EM Particle Processing</b> <a href="https://github.com/westlake-repl/Cryo-IEF">https://github.com/westlake-repl/Cryo-IEF</a> Designed and released the Cryo-IEF ecosystem, pretrained on ~65M cryo-EM particle images. Built PyTorch inference and downstream tooling for CryoRanker and CryoClustering, enabling practical structural classification, pose clustering, and particle-quality assessment.
2024–2026	<b>CryoDECO: Heterogeneous Reconstruction with Foundation Model Priors</b> <a href="https://github.com/yanyang1998/CryoDECO">https://github.com/yanyang1998/CryoDECO</a> Developed a prior-guided heterogeneous reconstruction framework that uses Cryo-IEF representations to reduce random ab initio initialization and disentangle compositional classification from 3D reconstruction.
2024–2025	<b>CryoWizard: Fully Automated Single-Particle Cryo-EM Reconstruction Pipeline</b> <a href="https://github.com/SMART-StructBio-AI/CryoWizard">https://github.com/SMART-StructBio-AI/CryoWizard</a> Built and extended an end-to-end computational pipeline integrating CryoRanker with CryoSPARC, streamlining raw movies, micrographs, or particles to high-resolution 3D volumes without manual intervention through command-line, web, and Chrome-extension interfaces.
2023–2026	<b>cryodata: Open-Source Python Package for Cryo-EM Data Processing</b> <a href="https://github.com/SMART-StructBio-AI/cryoief-data">https://github.com/SMART-StructBio-AI/cryoief-data</a> Created the <b>reusable data layer</b> used by Cryo-IEF, CryoDECO, and CryoWizard to convert raw CryoSPARC particle outputs into PyTorch-ready training and inference datasets. The package supports MRC/MRCS preprocessing, LMDB-backed dataset creation, Fourier/Hartley feature generation, balanced sampling, PyTorch data loading, and CryoSPARC-to-RELION metadata conversion. Packaged and documented the toolkit for <code>pip install cryodata</code> , reaching <b>over 18.2k downloads</b> (according to pepy.tech) as of May 2026 and turning lab-specific preprocessing into reproducible open-source scientific ML infrastructure.

## Technical Skills

---

<b>ML / Vision</b>	Self-supervised learning, representation learning, foundation models, clustering, autoencoders, generative/reconstruction models, image denoising.
<b>Structural Biology</b>	Single-particle cryo-EM, cryo-ET, CryoSPARC/RELION workflows, ChimeraX visualization, MRC/MRCS data formats.
<b>Systems</b>	Multiprocessing data pipelines, parallel GPU training/inference.

## Research Agenda

---

I aim to build foundation-model priors and agentic reconstruction systems for cryo-EM/cryo-ET that infer molecular identity, conformational dynamics, and native cellular context from large-scale imaging data. My near-term focus is representation-guided heterogeneous reconstruction; my long-term goal is AI-guided panoramic structural biology for complex molecular mixtures.