Speakers

1 Tristan Bepler

2 Jonathan Bouvette

3 Muyuan Chen

4 Anchi Cheng

5 Daisuki Kihara

6 Dari Kimanius

7 Mikhail Kudryashev

8 Yilai Li

9 Duane Loh

10 Jola Mirecka

11 Dong Si

12 Amit Singer

13 Carlos Sorzano

14 Sameer Velankar

15 Shruthi Viswanath

16 Thorsten Wagner

17 Min Xu

Abstracts

Workshop on Machine Learning for cryoEM

(19–22 Sep 2022)

1 Tristan Bepler

Simons Machine Learning Center, USA
Frontiers of ML in cryoEM

Abstract

2 Jonathan Bouvette

National Institutes of Health, USA
Automated systematic evaluation of cryo-EM specimens with SmartScope

Abstract

Finding the conditions to stabilize a macromolecular target for imaging remains the most critical barrier to determining its structure by cryo-electron microscopy (cryo-EM). While automation has significantly increased the speed of data collection, specimens are still screened manually, a laborious and subjective task that often determines the success of a project. Here, we present SmartScope, the first framework to streamline, standardize, and automate specimen evaluation in cryo-EM. SmartScope employs deep-learning-based object detection to identify and classify features suitable for imaging, allowing it to perform thorough specimen screening in a fully automated manner. A web interface provides remote control over the automated operation of the microscope in real time and access to images and annotation tools. Manual annotations can be used to re-train the feature recognition models, leading to improvements in performance. Our automated tool for systematic evaluation of specimens streamlines structure determination and lowers the barrier of adoption for cryo-EM.

3 Muyuan Chen

Stanford University, USA
Deep learning based Gaussian mixture model for characterizing variability in CryoEM

Abstract

Structural flexibility and/or dynamic interactions with other molecules is a critical aspect of protein function. CryoEM provides direct visualization of individual macromolecules sampling different conformational and compositional states, but characterization of continuous conformational changes or large numbers of discrete states without human supervision remains challenging. Here, we present a machine learning algorithm to determine a conformational landscape for proteins or complexes using a 3-D Gaussian mixture model mapped onto 2-D particle images in known orientations. Using a deep neural network architecture, the method can automatically resolve the structural heterogeneity within the protein complex and map particles onto a small latent space describing conformational and compositional changes. The usage of Gaussian mixture model also makes it possible to incorporate prior information, including domain segmentation and molecular model of the target protein, to guide the heterogeneity analysis. The algorithm is applied to multiple biological systems, to explore the structural heterogeneity of protein complexes at a range of scales.

4 Anchi Cheng

New York Structural Biology Center, USA
Fully Automated Multi-Grid Cryo-EM Screening using Smart Leginon

Abstract

CryoEM screening entails the identification of optimal grids and optimal sample conditions for subsequent high-resolution data collection. In order to obtain optimized grids that may produce high-resolution data, researchers usually aim to optimize several factors. Here, we incorporate machine learning model from a new software package Ptolemy and new purpose-built image processing algorithms into our automated data collection software, Leginon, to provide an open-source solution for fully automated, high-throughput grid

screening. This new version, broadly called Smart Leginon, emulates the actions of an operator in identifying areas on the grid to explore as potentially useful for data collection. The autoscreen workflow using these new features can also sequentially load and examine grids from an automated specimen exchange system to provide completely unattended grid screening across a set of grids. I will present our evaluation of the system and share our experience in incoorporate machine learning into this practical use case.

5 Daisuki Kihara

Purdue University, USA

Building and Validating Protein Structure Models for cryo-EM Maps Using Deep Learning

Abstract

Cryo-electron microscopy (cryo-EM) has become one of the main experimental methods for determining protein structures. protein structure modeling from cryo-EM is in general more difficult than X-ray crystallography since the resolution of maps is often not high enough to specify atom positions. We have been developing a series of computational methods for modeling protein structures from cryo-EM maps. For maps at medium resolution, deep learning can provide useful structure information for modeling. Captured features from density maps can be use for modeling as well as validating existing structure models. We present tools for structure modeling, fitting, and validation for cryo-EM. All the tools we developed are available at https://kiharalab.org/emsuites/.

6 Dari Kimanius

MRC, UK

Representation learning with prior constraints

Abstract

I will present the ongoing work in algorithm development based on modern machine learning approaches in RELION and beyond. By carefully incorporating more prior information into the reconstruction process and better

accounting for structural heterogeneity, we aim to create robust algorithms with improved automation that can handle smaller datasets or datasets with lower signal-to-noise ratios. I will discuss how this relates to data representation learning and how to handle new challenges these new approaches pose for validation and certainty estimation.

7 Mikhail Kudryashev

Max Delbrück Center for Molecular Medicine, Germany
At-scale cryo-ET data processing for high-resolution structural determination of membrane proteins

Abstract

8 Yilai Li

University of Michigan, USA
Machine learning approaches to automate single particle cryo-EM data acquisition

Abstract

Single-particle cryo-electron microscopy (cryo-EM) continues to grow exponentially, serving as a go-to technique for users of all levels of expertise. Recent developments in data collection strategies alongside new sample preparation devices herald a future where users will collect multiple datasets per microscope session. However, human-guided cryo-EM data collection practices limit the impact of cryo-EM because cryo-EM datasets typically represent 2-5% of the total sample area. In this talk, I will present my work in the past few years mostly on the automation of cryo-EM data acquisition. By combining supervised regression and reinforcement learning, our algorithm, cryoRL, maximizes data quality in a limited time without human intervention, and performs in the top 10% of the human test subjects on the same dataset.

9 Duane Loh

National University of Singapore, USA
Cryo-EM inspired EM applications

Abstract

10 Jola Mirecka

CCP-EM, UK
Towards automatic shape-based clustering and recognition in cryo-ET

Abstract

11 Dong Si

University of Washington Bothell, USA
Fast, Accurate, and Fully Automated Macromolecular Complex Structure Prediction and Determination from 3D Cryo-EM

Abstract

Information about the macromolecular structure and related molecular mechanisms can assist in the understanding of its function in a living cell and the drug development processes. To obtain such structural information, we present DeepTracer, a fully automatic deep learning-based platform for fast de novo macromolecular complex structure determination from high-resolution cryo-electron microscopy (cryo-EM) density maps. It is a fully automated pipeline and users can perform cryo-EM data processing, target identification, and structure predictions through the intuitive graphical web server. The web service is globally accessible at `https://deeptracer.uw.edu`.

12 Amit Singer

Princeton University, USA
Heterogeneity analysis in cryo-EM by covariance estimation and manifold learning

Abstract

In cryo-EM, the 3-D molecular structure needs to be determined from many noisy 2-D tomographic projection images of randomly oriented and positioned molecules that are rapidly frozen in a thin layer of vitreous ice. A key assumption in classical reconstruction procedures for cryo-EM is that the sample consists of identical molecules. However, many molecules of interest exist in more than one conformational state. These structural variations are of great interest to biologists, as they provide insight into the functioning of the molecule. Determining the structural variability from a set of cryo-EM images is known as the heterogeneity problem, widely recognized as one of the most challenging and important computational problem in the field. Due to high level of noise in cryo-EM images, heterogeneity studies typically involve hundreds of thousands of images, sometimes even a few millions. Covariance estimation is one of the earliest methods proposed for heterogeneity analysis in cryo-EM. It relies on computing the covariance of the conformations directly from projection images and extracting the optimal linear subspace of conformations through an eigendecomposition. Unfortunately, the standard formulation is plagued by the exorbitant cost of computing the N^3 x N^3 covariance matrix. In the first part of the talk, we present a new low-rank estimation method that requires computing only a small subset of the columns of the covariance while still providing an approximation for the entire matrix. This scheme allows us to estimate tens of principal components of real datasets in a few minutes at medium resolutions and under 30 minutes at high resolutions. Furthermore, the method is automatically regularized and requires minimal user interaction. After the initial covariance computation, we can reconstruct conformational states from the latent space in real-time. In the second part of the talk, we discuss a manifold learning approach based on the graph Laplacian and the diffusion maps framework for learning the manifold of conformations. If time permits, we will also discuss the potential application of optimal transportation to heterogeneity analysis. Based on joint works with Joakim Andén, Marc Gilles, Amit Halevi, Joe Kileel, Amit Moscovich, and Nathan Zelesko.

13 Carlos Sorzano

Spanish National Research Council (CSIC), Spain
Machine learning needs and trends for Single Particle Analysis

Abstract

Single Particle Analysis by Cryo-electron Microscopy has established as a key player in Structural Biology as a way to determine the three-dimensional structure of biological macromolecules. Several advances have contributed to this success at different levels: sample preparation, microscope and image acquisition, and image analysis. In this latter step, the number of already existing methods and those appearing is huge. Among them, classical image processing as well as machine and deep learning algorithms have a central role. In this talk we will present an overview of the most active topics and their needs. In particular, there are several pressing needs: robustness to noise, accuracy, speed, automation, and validation. Especially, the one of speed is in contradiction with the other three and automated decisions are not easy to take in absolutely all situations. This is a complicated balance although the field is quickly advancing through it. One of the main tools to have reliable results is the comparison of the estimations of several methods on the same set of images. This is seldom seen in the field at the moment, but these comparisons should be encouraged. Finally, it is important as a field to move into a FAIR (Findable, Accessible, Interoperable and Reusable) data regime (for instance, only 2% of the structures deposited in EMDB for SARS-CoV2 have their corresponding raw data in EMPIAR). In this way scientific transparency will be promoted, and Single Particle Analysis will be based on more solid grounds.

14 Sameer Velankar

EMBL-EBI, Hinxton, UK

A new era in (structural) biology - Impact of structure prediction using AI methods

Abstract

The AlphaFold Protein Structure Database (AlphaFold DB) provides high-accuracy 3D structure predictions for more than 200 million proteins and was developed in a collaboration between PDBe and DeepMind. The abundance of experimental and predicted structure models is revolutionising structural biology and bioinformatics and opens up many new opportunities. These new developments also require an overhaul of the current tools and infrastructure

for efficient access to such models. I will describe the status and plans for AlphaFold DB and its early impact on life-science research, and also present our work on developing a comprehensive infrastructure for accessing structure data.

15 Shruthi Viswanath

Tata Institute of Fundamental Research, India
Modeling large macromolecular assemblies: can we combine integrative approaches with deep learning?

Abstract

16 Thorsten Wagner

MPI of Molecular Physiology, Germany
TomoTwin: Generalized particle picking for cryo-ET with metric learning

Abstract

Cryoelectron tomography enables the visualization of cellular environments in extreme detail through the lens of a benign observer; what remains lacking however are tools to analyze the full amount of information contained within these densely packed volumes. Detailed analysis of macromolecules through subtomogram averaging requires particles to first be localized within the tomogram volume, a task complicated by a low signal to noise ratio and the crowding of the cellular space. To assist in this crucial particle picking step we developed TomoTwin: A general picking model for cryo-electron tomograms based on deep metric learning that does not require manual annotation of training data. By embedding tomograms in an information-rich, high-dimensional space which separates macromolecules according to their 3-dimensional structure, TomoTwin allows users to identify proteins in tomograms de novo without manually creating training data or retraining the network each time a new protein is to be located.

17 Min Xu

Carnegie Melon University, USA

Reducing the training data annotation cost for deep learning based cryo-electron tomography analysis

Abstract

Cryo-election tomography (cryo-ET) has become an important imaging technique for visualizing cellular structure and organization at sub-molecular resolution and in a near-native state. However, the automated analysis of cryo-ET images is very difficult due to the high level of image content complexity, noise, distortions, and a large amount of data. Deep learning based methods have become promising for such automated analysis. However, a major problem is that supervised deep learning often requires a large amount of annotated training data. On the other hand, high-quality annotations of cryo-ET images are often not readily available, and require a laborious process to obtain. In this talk, we present a number of our recent exploratory studies on how to reduce the amount of annotation needed for deep learning based cryo-ET analysis, particularly on classifying and recovering macromolecule structures from subtomograms.