Qeios

Peer Review

Review of: "CryoSAMU: Enhancing 3D Cryo-EM Density Maps of Protein Structures at Intermediate Resolution with Structure-Aware Multimodal U-Nets"

Hung-Wen Chen[1]

1. National Tsing Hua University, Hsinchu, Taiwan

Summary

The manuscript by Zhang et al. introduces CryoSAMU, a multimodal deep-learning framework that enhances intermediate-resolution (4–8 Å) cryo-EM maps of protein structures. The method combines a 3D U-Net architecture with residue-level structural embeddings derived from the ESM-IF1 model, integrated via cross-attention. The authors demonstrate that CryoSAMU achieves significant improvements over state-of-the-art methods across real-space and Fourier-space metrics, as well as in downstream protein modeling accuracy. The model also offers notable efficiency gains, achieving a 13× speed-up over EMReady, suggesting strong potential for practical and large-scale applications.

Strengths

- Embedding residue-level structural features from ESM-IF1 into a 3D U-Net via cross-attention at the bottleneck is well motivated and clearly described (Section 3.3, Figure 1b).
- Real-space (CC_box, CC_volume, CC_peaks) and Fourier-space (FSC_0.5) metrics show consistent improvements over Autosharpen, DeepEMhancer, and EMReady (Section 4.2, Figure 3, Table 1).
- While limited in scope, the ablation in Section 4.5 (Figure 6) demonstrates a clear performance drop without structural embeddings.
- CryoSAMU achieves substantially faster processing speeds (32.49 s avg.), outperforming EMReady and others by up to 13× (Section 4.4, Figure 5, Table 3).

doi.org/10.32388/LLW6TV

- The release of source code and a well-documented high-level methodology—including model architecture, dataset preparation, and training procedures—provides a strong foundation for reproducibility.

Points for Revision

Q1: In the Abstract and Introduction, you position CryoSAMU as addressing a critical gap in intermediate-resolution cryo-EM map enhancement (4–8 Å). However, based on the resolution distribution of your examples in Section 4 and Supplementary Tables S1–S3, most test cases appear concentrated in the lower end of that range (i.e., 4.0–4.5 Å), with relatively few examples above 5 Å. Given that higher-end resolutions (5–8 Å) are often more challenging for model interpretation and map-based modeling, could you clarify how well CryoSAMU generalizes to this upper subrange?

Q2: Section 3.2 and Figure 1a describe a self-attention mechanism applied to residue embeddings before fusion, which is suggested to help preserve chain and residue relationships. However, no empirical evidence is provided to support this claim. Would it be possible to include: (a) attention heatmaps over residue indices to show that the model attends differently across chains or structural regions, and/or (b) quantitative analyses (e.g., correlation between attention scores and known residue-level importance)?

Q3: While Section 3.3 and Figure 1b provide a general description of the cross-attention design, several low-level implementation details remain unclear—such as the specific projection dimensions of Q/K/V, whether normalization layers (e.g., LayerNorm) or dropout are used within the attention block, and whether positional encoding is applied. Including these in a schematic or supplementary table would significantly improve reproducibility.

Q4: CryoSAMU introduces multiple novel components, including residue-level self-attention weighting, cross-attention fusion in the bottleneck, and the adoption of a smooth L1 loss function. However, the ablation study in Section 4.5 focuses solely on the inclusion of structural embeddings. It would improve clarity and rigor to include additional ablations isolating other key components—such as removing cross-attention or replacing the learned self-attention with uniform weights. In addition, although you mention that smooth L1 loss performs better than MSE (Section 3.4), no experiments are shown to support this choice. A brief comparison or citation to support this decision would strengthen your argument.

Declarations

Potential competing interests: No potential competing interests to declare.