Qeios

Peer Review

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

Hany Akeel Al-hussaniy[1]

1. University of Baghdad, Iraq

The manuscript introduces **CryoSAMU**, a structure-aware multimodal framework for enhancing **intermediate-resolution (4–8 Å)** cryo-EM maps using a 3D U-Net augmented with **ESM-IF1 structural embeddings** via cross-attention. The work targets a very practical bottleneck in structural biology: maps in the 4–8 Å range are frequently challenging for interpretation and downstream model building, and existing DL enhancement methods are often trained/optimized outside this “intermediate” regime or rely purely on density features. Overall, the paper is **well organized**, the method is described with substantial detail, and the authors provide an open-source implementation. Importantly, the reported **speed advantage** (order-of-magnitude faster than some competitors) is a strong practical contribution for routine workflows and large-scale applications.

That said, several points need clarification to strengthen rigor, generalizability, and user confidence—particularly for the upper end of the target resolution range where hallucinated detail is a known risk in map enhancement.

Major/Key points to address

Generalization across resolution and SNR: The paper would benefit from stratifying evaluation by **resolution brackets** (e.g., 4–5, 5–6, 6–8 Å) and **SNR/quality tiers**. Intermediate maps are not uniform; a method that performs well at ~4.0–4.5 Å may behave differently at 6–8 Å, where over-sharpening and artificial features become more likely. Please include bracketed results (and ideally representative examples) to demonstrate robustness.

doi.org/10.32388/TIDPFA

“Structure-aware training but structure-free inference” design: A central design decision is that structural embeddings are used during training, but are not available at inference (cross-attention bypassed). This is interesting, but it also raises questions:

Why does training with a modality that is absent at inference not cause a train–test mismatch?

Can the authors quantify performance sensitivity to this bypass (e.g., using teacher forcing alternatives, or distillation-style training)?

A clearer explanation of why this improves inference would increase trust in the approach.

Risk of introducing non-physical detail: Since map enhancement can generate features not supported by the data, the manuscript should discuss safeguards and/or diagnostics. For example, do the authors observe cases where CryoSAMU produces overly detailed densities at 6–8 Å? Consider adding a brief “failure modes” subsection and guidance for end users on when to trust the output (e.g., cross-checks with local resolution, FSC behavior, or map-model validation metrics).

Declarations

Potential competing interests: No potential competing interests to declare.