Deep learning of elastic 3D shapes for cryo electron microscopy analysis of continuous conformational changes of biomolecules

Tzu-Hsuan Chang

R e f e r e n c e

Deep learning of elastic 3D shapes for cryo electron microscopy analysis of continuous conformational changes of biomolecules

Ilyes Hamitouche and Slavica Jonić*
IMPMC - UMR 7590 CNRS, Sorbonne Université, MNHN, Paris, France
ilyes.hamitouche@upmc.fr, slavica.jonic@upmc.fr (corresponding*)

***Abstract*—Cryo electron microscopy (cryo-EM) allows high-resolution 3D reconstruction of biomolecular structures from highly noisy 2D parallel-beam projection images containing tens of thousands of copies of the same macromolecular complex but at different random orientations and positions. However, biomolecular complexes are not rigid but flexible entities that change their conformations gradually (continuous transition with many intermediate states) to accomplish biological functions (e.g., DNA replication, protein synthesis, etc.). The determination of the full distribution of conformations (conformational space or landscape) from cryo-EM images is challenging but could provide insights into working mechanisms of the complexes. In this paper, we present a method for conformational space determination, which uses deep learning in combination with cryo-EM image analysis and normal mode analysis (molecular mechanics simulation), where the amplitudes of normal modes are used as parameters of the elastic 3D shapes of complexes (the parameters determining the conformation). We show the performance of this new method using synthetic cryo-EM data.**

***Index Terms*—Deep Learning, Cryo-EM, Elastic 3D-to-2D alignment, Normal Modes, Molecular Dynamics**

analysis is suited to discrete conformational changes (e.g., two-state heterogeneity of binding and unbinding of a complex with another molecule) but, generally, it is suboptimal. Indeed, biomolecular complexes generally adopt gradual transitions with a large unknown number of intermediate conformational states (continuous conformational changes), which yields a particularly challenging type of hetereogeneity [3].

The majority of the existing methods are based on classification into a small number of discrete classes, which is often defined using a prior knowledge about the number of expected conformations [4], [5]. The other group of methods deal with continuous conformational heterogeneity by determining the full conformational distribution (also called conformational space, conformational landscape, or conformational manifold) [6], [7], [8], based on which groups of images with similar conformations are made and 3D reconstructions from these groups computed. The development of methods for continuous conformational hetereogeneity analysis is currently an active

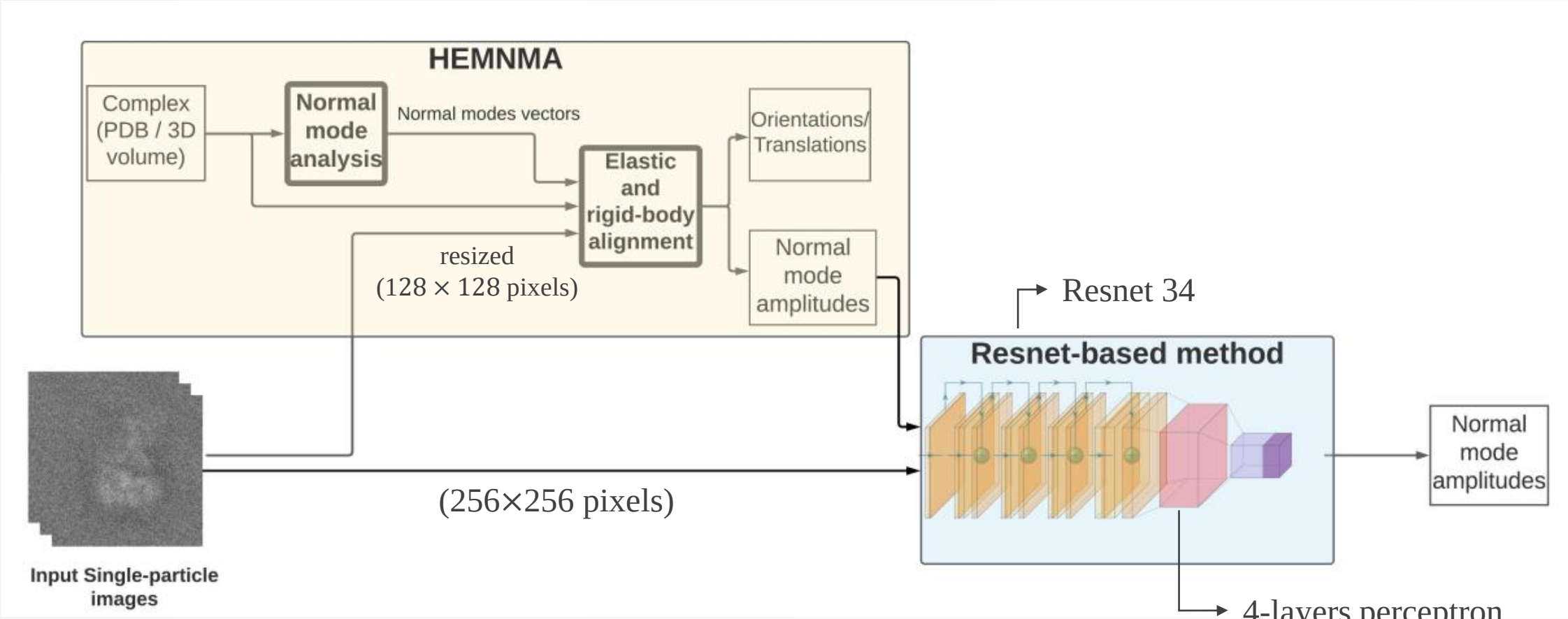
P u r p o s e

1. To know more about working mechanisms of the molecules.

2. Deal with the disadvantage that discard images while doing classification.

3. To speed up determination of conformation.

P r o c e d u r e



(256×256 pixels)

Normal Mode Analysis

→ A fast and simple method to calculate vibrational modes and protein flexibility.

Potential energy equation

$V(q) = V(q^0) + \left(\frac{\partial V}{\partial q_i}\right)^0 \eta_i + \frac{1}{2}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right) \eta_i \eta_j + \cdots$

Minimum energy conformation (coordinates)

$\eta_i = q_i - q_i^0$

$= \frac{1}{2}\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right) \eta_i \eta_j = \frac{1}{2} \eta_i V_{ij} \eta_j$

Hessian matrix

Kinetic energy

$T(q) = \frac{1}{2} M \frac{d^2 \eta_i}{dt^2}$

Matrix of mass of each particle

Entire equation of motion

$\frac{1}{2} M \frac{d^2 \eta i}{dt^2} + \frac{1}{2} \eta_i V_{ij} \eta_j = 0$

$\eta_i = a_{ik} \cos(\omega_k t + \delta_k)$

frequency

Phase factor

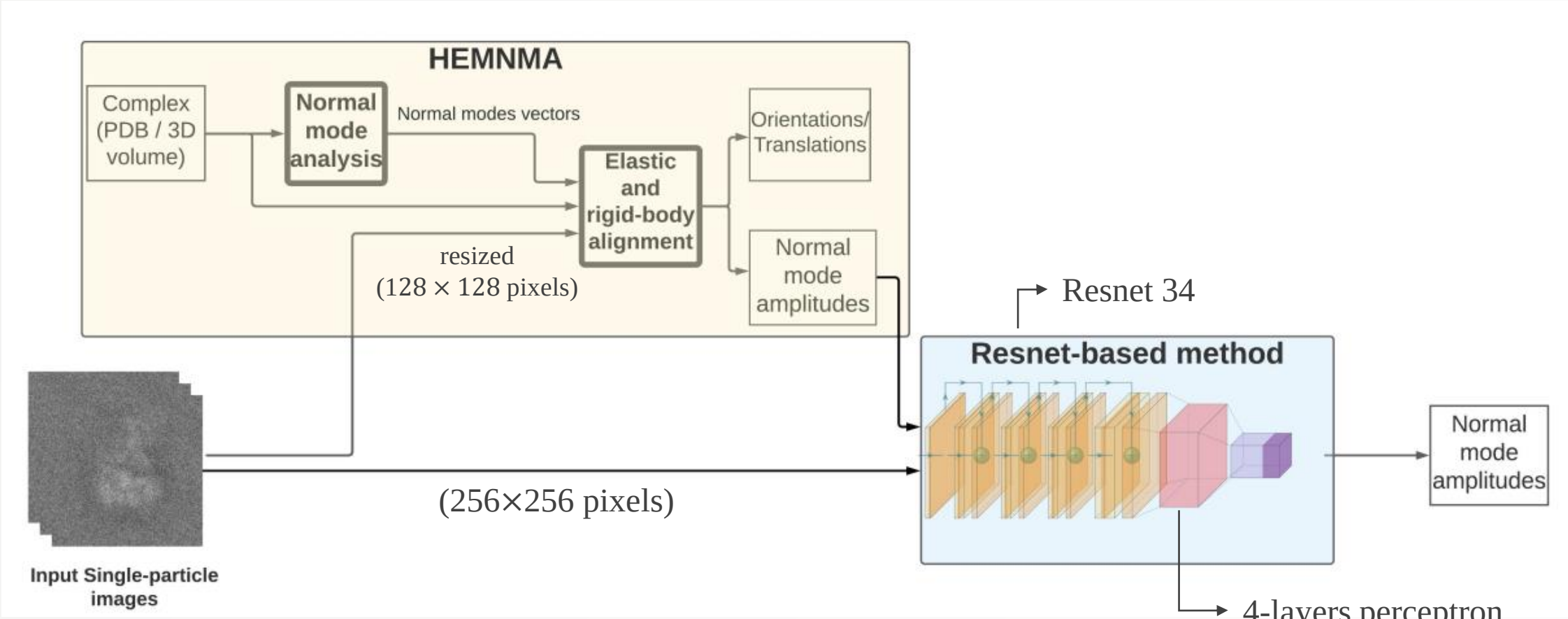
Amplitude of oscillation

$VA = \lambda A$

Eigenvalues

Eigenvectors of Hessian matrix V

P r o c e d u r e



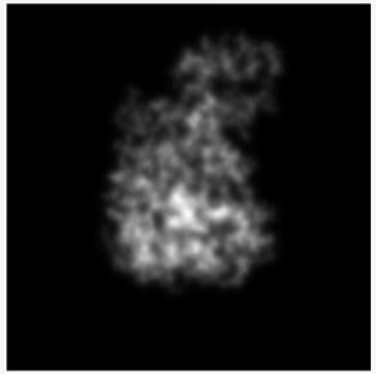
(256×256 pixels)

Experiment

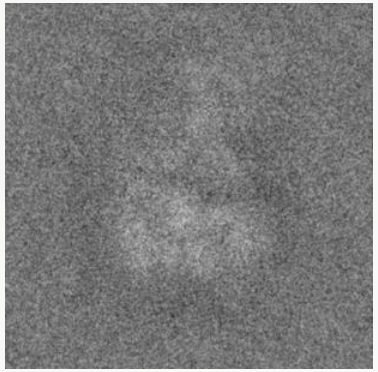
Atomic model

Adenylate Kinase chain A (AK)
PDB ： 4AKE

Synthetic image



Ideal image



SNR = 0.1
Defocus ： -0.5 μm

Calculate normal modes of structure

↓

Displace modes with random amplitudes
$(q_7 - q_9)$

↓

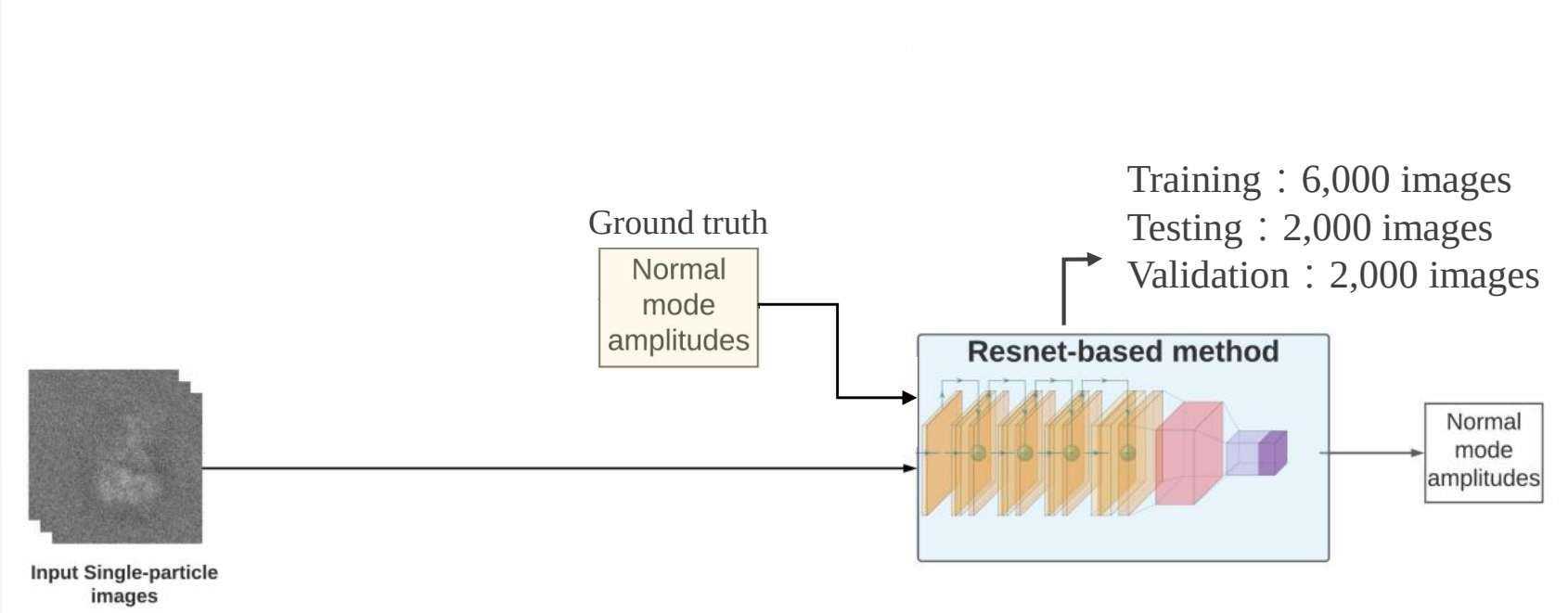
Convert into density map

↓

Project onto image plane at random orientation

$$q_7(r) = -200 \cdot \mathrm{r}$$
$$q_8(r) = 200 \cdot \sin(\pi \mathrm{r})$$
$$q_9(r) = 200 \cdot \cos(\pi \mathrm{r})$$
$$random\ variable\ r, r \in [0,1]$$

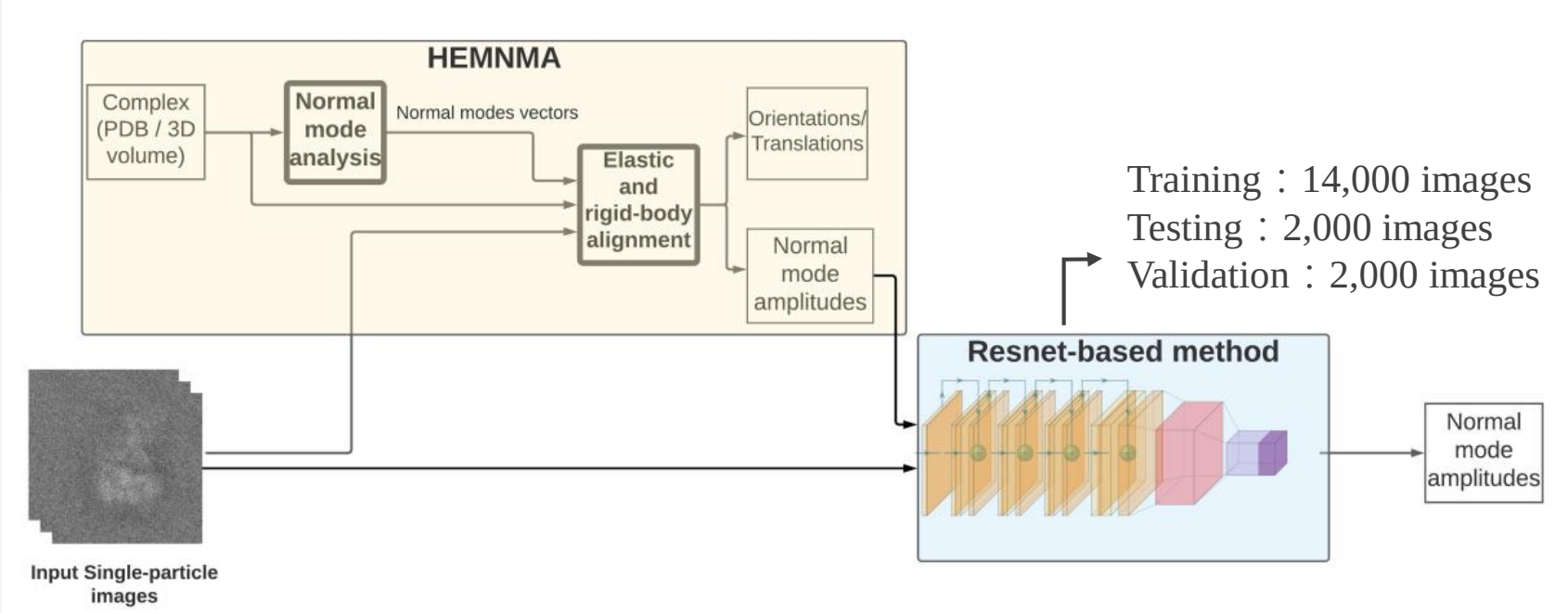
Experiment I



Result I

Cases	Noise	Defocus	Global error	q_7 error	q_8 error	q_9 error
No in-plane rotation, no shift	No	No	2.32	1.49 ± 2.06	3.10 ± 4.49	2.35 ± 2.90
No in-plane rotation, no shift	SNR 0.1	-0.5 µm	5.79	3.62 ± 4.91	7.27 ± 10.69	6.49 ± 9.09
No in-plane rotation, random shift	No	No	4.83	3.02 ± 4.55	6.14 ± 10.11	5.32 ± 8.61
No in-plane rotation, random shift	SNR 0.1	-0.5 µm	7.95	4.90 ± 6.72	9.85 ± 14.72	9.10 ± 13.32
Random in-plane rotation, no shift	No	No	16.86	10.32 ± 15.38	19.53 ± 29.60	20.74 ± 33.43
Random in-plane rotation, no shift	SNR 0.1	-0.5 µm	19.62	12.00 ± 17.64	22.32 ± 31.95	24.53 ± 39.03
Random in-plane rotation and shift	No	No	23.51	14.67 ± 21.14	24.44 ± 34.00	31.42 ± 49.10
Random in-plane rotation and shift	SNR 0.1	-0.5 µm	27.60	17.14 ± 23.65	29.12 ± 39.42	36.55 ± 54.46

Experiment II

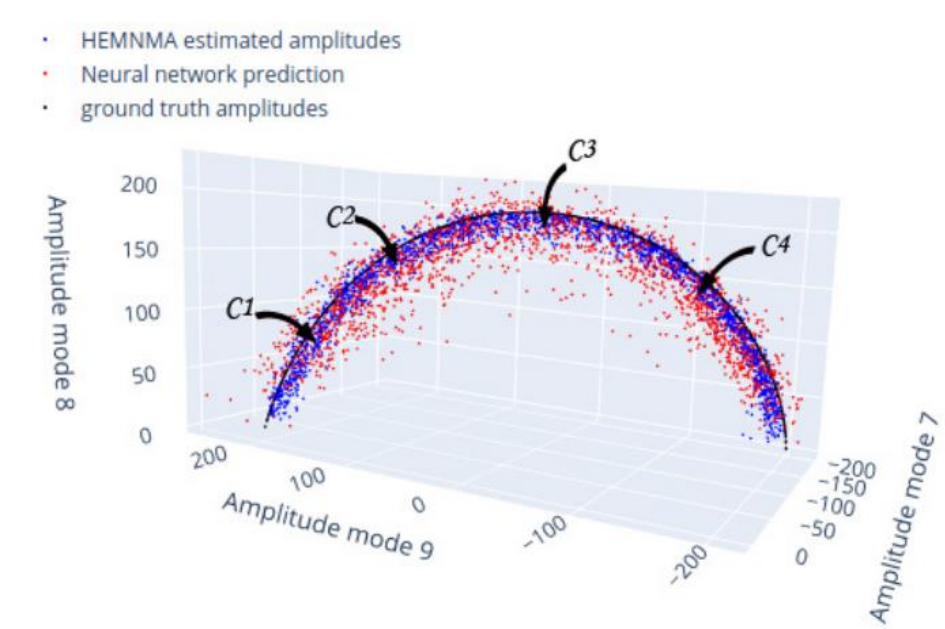


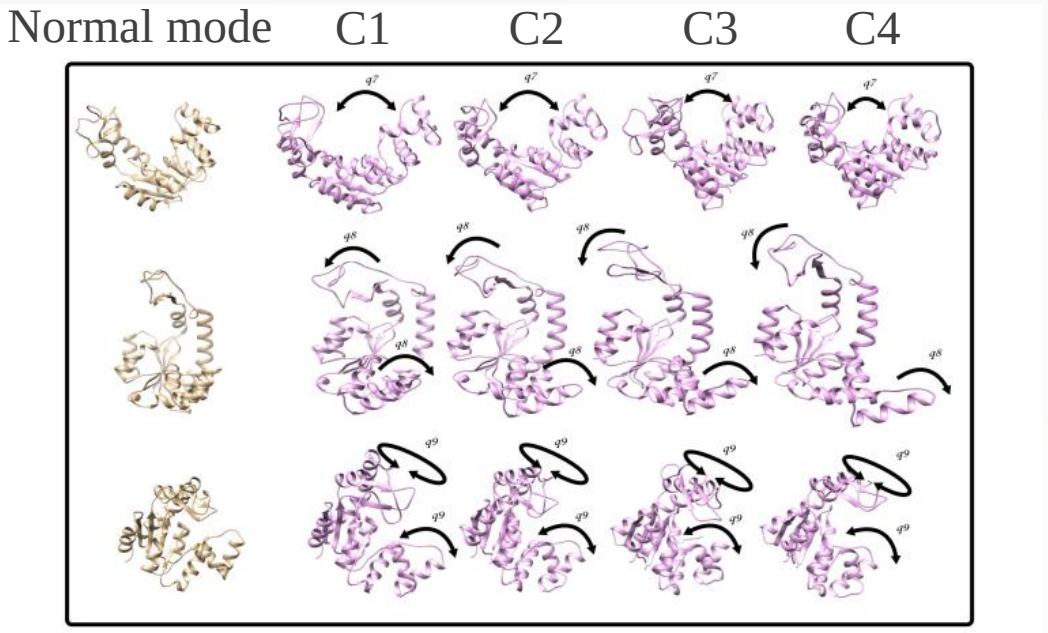
Training：14,000 images
Testing：2,000 images
Validation：2,000 images

Input：18,055 images
(256×256 pixels)

ERRORS	GLOBAL	q_7	q_8	q_9
INFERRED *vs.* HEMNMA-ESTIMATED	19.03	12.35 ± 16.47	19.58 ± 26.15	25.16 ± 34.60
INFERRED *vs.* GROUND-TRUTH	20.22	12.61 ± 16.84	20.94 ± 27.37	27.13 ± 36.82
HEMNMA-ESTIMATED *vs.* GROUND-TRUTH	6.58	5.75 ± 8.42	6.23 ± 7.20	7.76 ± 7.23

Average atomic-coordinate error ： 0.9 Å





Conclusion

1. A method combines neural network and HEMNMA
2. In future, this method can be used to obtain 3D reconstruction from images corresponding to similar conformations.