Neural Network Based Methods for the Conformational Landscape Determination in High Resolution Cryo-Electron Microscopy

Dissertation

for the award of the degree
“Doctor rerum naturalium” (Dr. rer. nat.)
of the Georg-August-Universität Göttingen

within the doctoral program
“PhD Programme in Computer Science” (PCS)
of the Georg-August University School of Science (GAUSS)

submitted by

Georg Bunzel

from Leipzig, Germany

Göttingen 2022

Thesis Committee

Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Florentin Wörgötter
Computational Neurosciences, University of Göttingen

Prof. Dr. Michael Habeck
Microscopic Image Analysis, Jena University Hospital

Members of the Examination Board

1st Referee: Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

2nd Referee: Prof. Dr. Florentin Wörgötter
Computational Neurosciences, University of Göttingen

Prof. Dr. Michael Habeck
Microscopic Image Analysis, Jena University Hospital

Prof. Dr. Kai Tittmann
Göttingen Center for Molecular Biosciences, University of Göttingen

Dr. Ashwin Chari
Structural Biochemistry and Mechanisms, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Stephan Waack
Theory and Algorithmic Methods, Institute for Computer Science

Date of oral examination: 22th November 2021

Affidavit

I hereby declare that this dissertation with the title “Neural Network Based Methods for the Conformational Landscape Determination in High Resolution Cryo-Electron Microscopy” has been written independently and with no other aids or sources than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Georg Bunzel

Abstract

During the last decades 3D transmission electron cryo-microscopy (cryoEM) has emerged to be the method of choice for the study of motion in larger protein complexes. The aim of Single Particle Analysis (SPA) in cryoEM is to combine a large amount of noisy projection images, usually obtained by a transmission electron microscope (TEM), of the same macromolecular complex into one noise reduced structure (or a finite set of them to study the motion of the respective complex over time). This process is computationally very demanding as the amount of images needed increases tremendously for higher resolution levels. The advances in cryoEM, hence, were backed up by the advances in both, instrumentation (e.g. TEMs, sensors, correctors) but also in computational image processing. Various software tools have been proposed over the years to tackle specific subtasks of the image processing cycle. However, they often rely on certain assumptions (e.g. starting models as reference) or require human input when it comes to the sorting of finite conformational states to describe the dynamics of macromolecular complexes.

In this work a novel software tool based on Artificial Neural Networks (ANNs) is proposed. It does not rely on prior information on the studied dataset. Instead, it processes cryoEM particle images fully autonomously without human intervention. Furthermore, it aims to estimate a continuous conformational space of the studied complex, which can be sampled in order to generate smooth trajectories, instead of a small finite set of conformations.

The software tool is then evaluated with respect to its performance on synthetic test data and also on existing cryoEM datasets featuring two different macromolecular complexes.

Keywords: GAUSS, dissertation, cryoEM, machine learning, variational autoencoder, VAE, ANN

Acknowledgment

First and foremost I want to thank Prof. Holger Stark for his supervision during the last years. While always being extremely helpful and supportive he gave me the opportunity and the freedom to explore. He never got tired of discussing a new idea or evaluating the latest results, sometimes even late at night. His spirit and enthusiasm has always been inspiring to me and I will always be grateful for what I have learned and experienced during the last years. I also want to thank Prof. Florentin Wörgötter for being part of my thesis advisory committee and his helpful remarks during the meetings. I cannot thank Prof. Michael Habeck enough for always supporting me in every aspect, for all the conversations on sometimes pretty abstract topics, his ability to sort and channel my thoughts, his patience and encouragement. I am deeply grateful to him.

Furthermore I want to thank the members of the Stark lab that I met on the way. First I want to thank Dr. David Haselbach, from him I learned most about cryoEM. His ability to interpret between biochemistry and computer science was extremely helpful and made access a lot easier. I want to thank Dr. Niels Fischer who inspired me multiple times, Dr. Ashwin Chari for all the interesting conversations.

I want to thank the members of the development team. I learned a lot from Dr. Boris Busche and Dr. Jan-Martin Kirves, with whom I spent most of the time when I was new to the lab. I want to thank Dr. Lukas Schulte for the fun that we had and the conversations regarding my project. Dr. Sabrina Fiedler definitely deserves a special spot here for all the discussions on maths, noise and the beauty of Matlab, but also for being a friend I could always rely on. I also want to thank Tobias Koske and Dr. Mario Lüttich for maintaining the hardware infrastructure.

I am grateful for all the things that I learned about electron microscopy from Dr. Uwe Lücken and Dr. Dietmar Riedel, all of their stories (and there were a lot!) and their support. I want to thank Dr. Dirk Wenzel for the chats on our shared hobby.

Furthermore I want to thank Dr. Michael Hons. Without him I would maybe never got in touch with the world of cryoEM.

I want to thank Dr. Fabian Henneberg for all the nice and deep conversations. I want to thank Dr. Karl Bertram for all of his enthusiasm and optimism and - of course - the coffee tastings.

I am thankful for all the members of my office. Especially Uma Lakshmi Dakshinamoorthy, Valik Petrychenko and Dr. Jan-Erik Schliep, for their scientific input on the biochemical side of this project, but also for being good friends. I want to thank Dr. Kashish Singh and Benjamin Graf for their expertise and the valuable input on the FAS complex, as well as for the data that I got from Kashish which is part of this thesis.

A special thanks also goes to Juliane Moses, who took care of all the administrative aspects.

Of course I also want to thank my friends Patrick, Michael, Thomas, Anna, Melanie, Katja, Lars, Jan, Sebastian, Ega, Florian for their support. Lena, Karl, Uwe and Sabrina I want to thank for all the highly competitive rounds of Doppelkopf.

I want to thank my family for their support, especially my sister who continuously encouraged me in everything I did.

Last, but not least I want to thank Elke. Without her, I would most likely not be the computer scientist that I am today.

Contents

List of Tables

List of Figures

1 Introduction

Any sufficiently advanced technology is indistinguishable from magic.
– **Arthur C. Clarke**, *Clarke's third law*

Macromolecular machines are the essential building blocks of all living beings on Earth. They perform various tasks to enable cells to grow, reproduce and degrade in an ordered manner. Hence, studying the motion of biological macromolecules is crucial for understanding the complex interactions of these systems. Structural biology has developed lots of different techniques to shed light (or electrons) on the basic mechanisms that drive these machines. The following sections are a brief overview on the main approaches used to date, namely Nuclear Magnetic Resonance spectroscopy (NMR), X-ray Crystallography (XRC) and 3D transmission electron cryo-microscopy (cryoEM).

1.1 Methods in Structural Biology

1.1.1 X-ray Crystallography (XRC)

XRC is the oldest method used in structural biology. Still the vast majority of structures is solved with this method (see Figure 1.1), due to its robustness and its ability to achieve resolutions below 1Å. The aim of XRC is to embed identical shaped molecules in an aligned and regularly patterned lattice of unit cells forming a crystal. This crystal is then introduced into a beam of x-rays and the diffraction pattern is recorded. Gradually rotating the crystal while continuously recording the diffraction pattern allows us to compute a three-dimensional (3D) structure of the crystallized complex. One drawback of XRC is that the recorded diffraction patterns lack phase information that needs to be restored externally. The main limitation, however, is that the crystallization process in itself is very tedious, hard to rationalize and still poorly understood.

The (ideally) uniform structure of the crystal lattice is advantage and disadvantage at the same time. On the one hand it allows very high resolutions, on the other hand it prevents from studying dynamics in the sample.

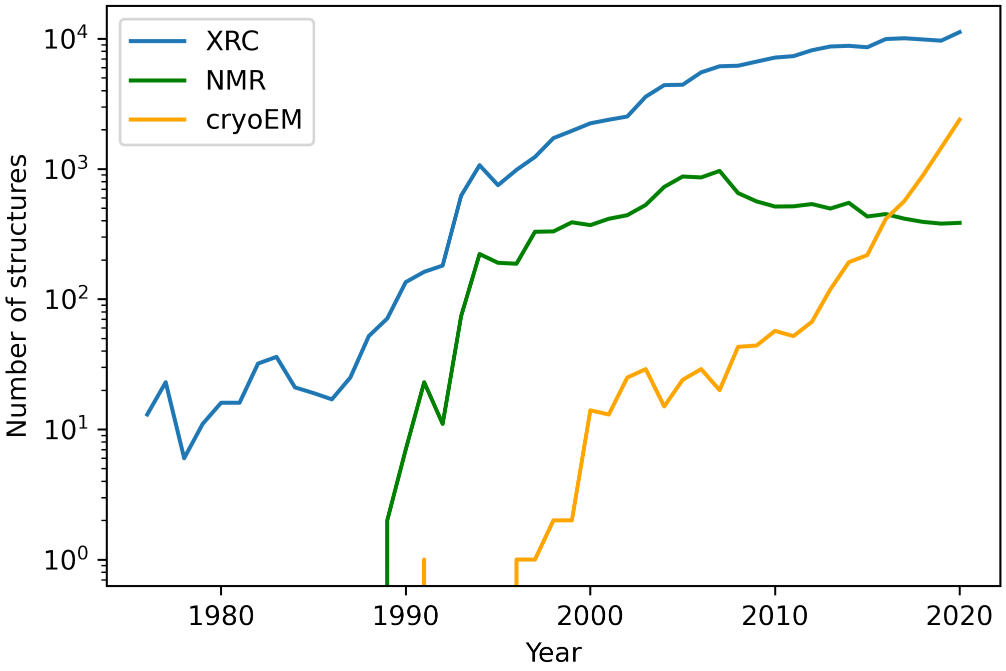


Fig. 1.1: Numbers of published structures in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [13] with their respective imaging method. The overall number of structures is growing rapidly with a sharp increase especially for cryoEM.

1.1.2 Nuclear Magnetic Resonance spectroscopy (NMR)

NMR makes use of the fact that the gyromagnetic ratio of atomic nuclei differ which results in different resonance frequencies. Introduced into a strong magnetic field atomic nuclei align along the direction of the magnetic field. Specific radio wave pulses result in a rotation of the magnetic moment of the nuclei. After the pulse they relax back into their previous orientation but at the same time emit the gained energy as radio wave pulses. The energy can be measured, as well as the relaxation times. This information can be integrated into a spectrum. Specific patterns of radio wave pulses then allow for the extraction of structural information of the molecule and even studying the motion of the molecule over time. However, due to the complexity of the spectrum it is computationally very intricate to extract this information. NMR is therefore limited to studying relatively small molecules [37, 38].

1.1.3 Electron Microscopy

A wide variety of different electron microscope (EM) based methods are used nowadays for different use cases. They differ in the type of used EM, e.g. transmission electron microscope (TEM), Scanning Electron Microscope (SEM) or Scanning Transmission Electron

Microscope (STEM), but also in the method itself. In terms of structure determination, e.g. cryo-tomography or cryoEM based Single Particle Analysis (SPA) are used. The method studied in this thesis is SPA using cryoEM. Compared to XRC and NMR it is a fairly young method, but its popularity for high resolution structure determination of biological macromolecules has increased tremendously over the last decades (see Figure 1.1).

cryoEM makes use of a TEM to record images of thousands (sometimes even millions) of molecules of the same kind that are embedded into a thin layer of vitreous ice. cryoEM is the only single particle method of the aforementioned, because it records the full information spectrum for every individual recorded particle. This makes cryoEM the method of choice for the study of dynamic macromolecular assemblies. Given that all the images show the same macromolecule the information can then be fused into one single, or a set of 3D structures in a computationally expensive process, usually aided by High Performance Computing (HPC).

This computationally demanding process has led to a lot of effort which has been put into the development of sophisticated methods to extract meaningful information from the vast amounts of data produced. This inevitably lead to higher and higher resolved structures, which found its recent climax in the publication of apoferritin at atomic resolution by Yip et al. [78] from the Stark lab (Figure 1.2).

Despite that leap forward in resolution some problems remain unsolved to date, usually linked to low signal-to-noise ratio (SNR) in the images and structural heterogeneity within the macromolecules under investigation. Some macromolecular assemblies, in particular Pyruvate dehydrogenase (PDH), still can't be tackled with todays algorithms, due to their high flexibility.

The following section gives a short overview on the computational side of SPA, a detailed review on the method can be found in Chapter 2.

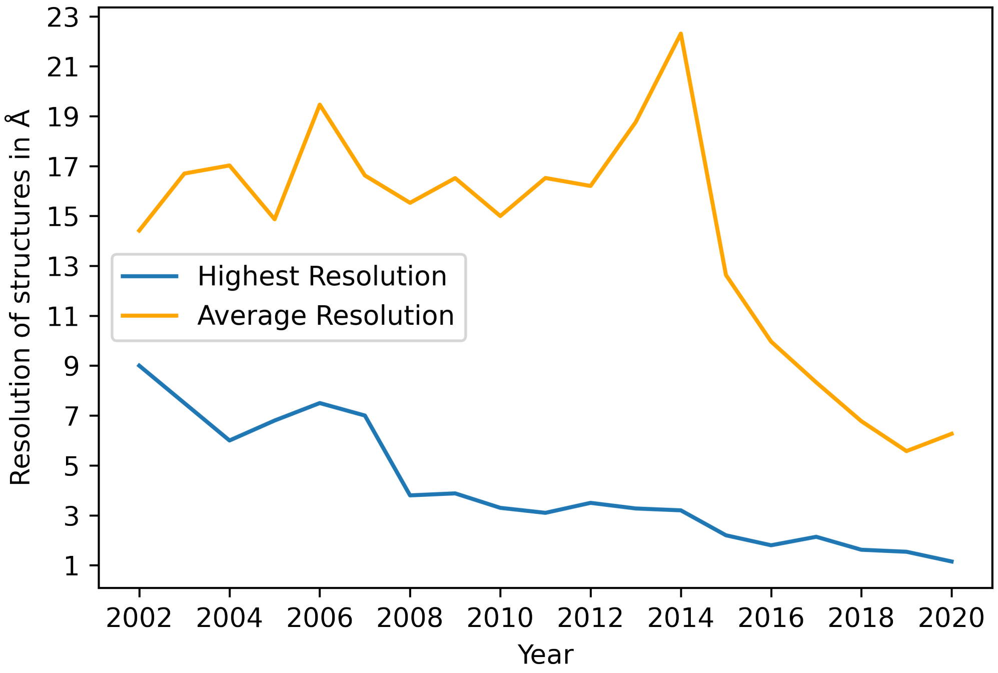


Fig. 1.2: Resolution trend for structures deposited in the RCSB PDB [13]. The overall resolution in cryoEM improves continuously both in everage and also in the maximum resolution achieved with its climax in Yip et al.[78] who published a structure of apoferritin at an atomic resolution of 1.25Å.

1.2 Primer on SPA using cryoEM

The calculation of 3D macromolecular structures is a resource demanding process. It combines biochemical, physical and computational methods. Due to the imaging conditions in cryoEM, especially the extremely low SNR (see Section 2.3.1) it is necessary to record lots of images and combine their information to suppress the noise. To achieve the latter, it is necessary that the specimen only consists of particles of the same kind. The concept of using different computational methods to combine images of the same shape is called SPA. The following section gives an overview (see Figure 1.3) of the basic concepts behind SPA. For a detailed description of the cryoEM work-flow see Section 2.4.

Preparation In a first step the purified sample needs to be prepared for imaging. That means it needs to be modified in a way that it can withstand the conditions in the EM. It is therefore, for example, applied to a thin carbon support film on a metal grid (usually copper or gold) and either stained with a heavy metal salt and air-dried (negative staining) or rapidly frozen in liquid ethane to prevent ice crystals from forming (vitrification/cryo conditions).

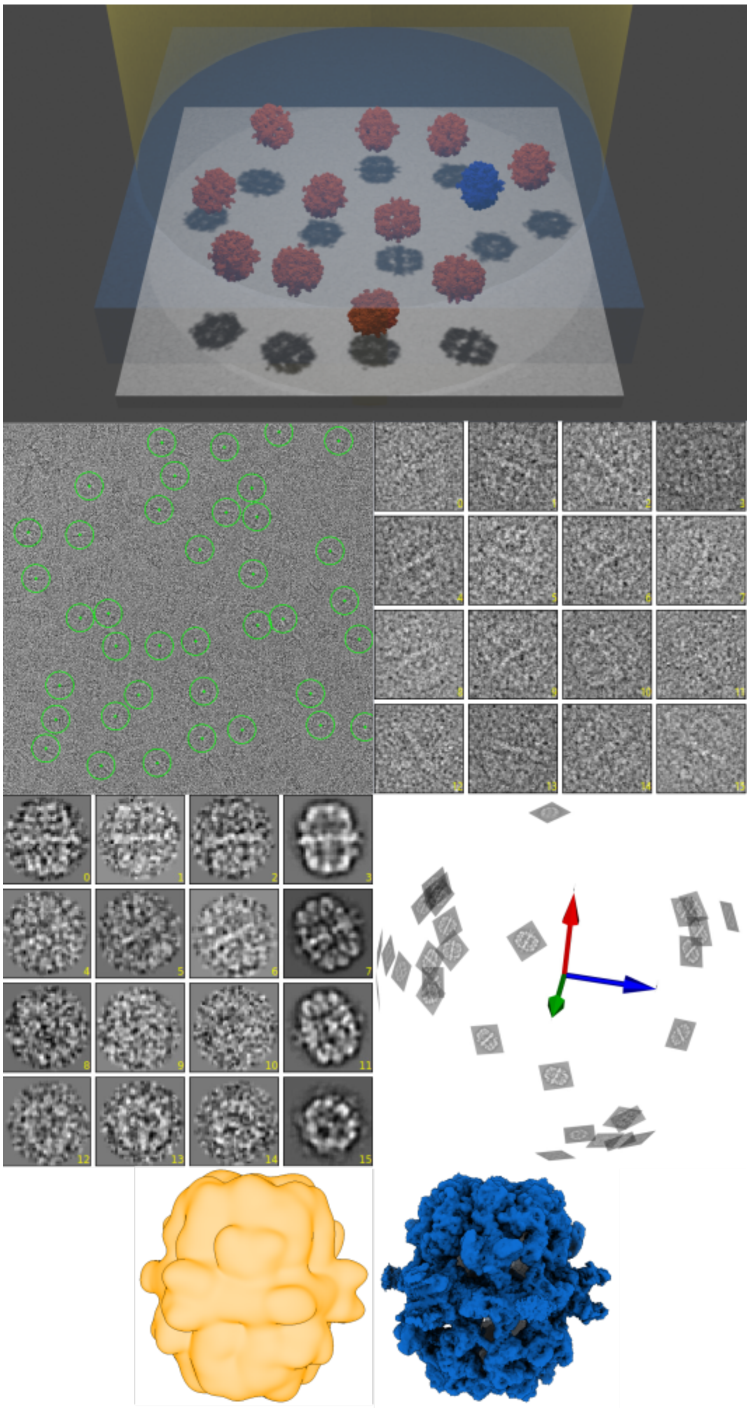


Fig. 1.3: cryoEM work-flow overview. **Top row:** Particles are rapidly frozen in a thin layer of vitreous ice. The particles show different conformations and orientations. An electron beam illuminates the sample creating noisy projection images called micrographs. **Second row:** Particles are selected (picked) on the micrograph (left) and extracted with a certain box size (right). **Third row:** Particles are aligned based on their orientation in space and then classified. Averaging over classes reduces noise (left). The set of of rotation angles is determined for every class sum and can be used to calculate a 3D structure (right). **Bottom row:** Intermediate 3Ds can be used as reference for an iterative refinement cycle which improves the low resolved structure (left) further till convergence (right)

Imaging Once prepared, the specimen is transferred into the TEM, exposed to a high vacuum (i.e. 10^{-7} mbar) and illuminated by a condensed and coherent electron beam. The beam either passes the sample or is scattered by the specimen creating an almost perfect projection image. That means, the image that is nowadays typically recorded on a direct electron detection device (DDD), can be assumed to be the sum of all the information illuminated by the beam. The images recorded are called *micrographs*. Every micrograph typically contains hundreds of particles that are (ideally) randomly oriented in space with five degrees of freedom in total (three rotations and the two in-plane shifts). In this thesis the set of all five degrees of freedom will be called *pose* and is denoted Φ. Due to radiation damage imposed by the electron beam it is crucial that the exposure times are short in order to keep the specimen intact. However, this results in a low SNR. The central aim of SPA is to reconstruct the pose for every noisy particle image and to then merge this information into a final structure.

Particle Selection (Picking) As the particles are randomly distributed it is necessary to first locate them on the micrographs. It can be done using a *reference*, that is filtered (see Figure 2.3) projections of an already known structure, or *de novo* meaning from scratch. Using a reference speeds up the picking, but can induce a selection bias into the process. Afterwards single particle images are cut out with a certain box size, creating a particle stack.

Preprocessing The picked particles now need to be corrected for aberrations that were induced by the microscope during imaging. Furthermore, effects due to strongly underfocusing the images need to be compensated. This is done in a process called Contrast Transfer Function (CTF) correction (see Section 2.4.4).
In a next step the particles in the stack need to be modified in a way that the distribution of their intensity values match as good as possible. This process is called normalization (see Section 2.3.2).
Sometimes the images are also filtered to enhance contrast and SNR for the early processing steps. Since filtering degrades image quality, unfiltered raw images are used later on.

Alignment To compensate for the different orientations of the particle in two-dimensional (2D) space a process called *alignment* is performed. That is finding an estimate for the shifts in the x- and y- plane and also the in-plane rotation. This step disentangles the view onto the 3D (which is the projection image) from its spatial orientation.

Classification Having the particles roughly oriented in the right direction now enables for classification. This is a process which tries to find particles that show the same view of the 3D volume in the stack of all particles with their unique orientations, or at least views that resemble each other. To suppress the noise, similar views are then averaged. The better the classification, the better the SNR in the resulting class averages is. This step follows the assumption that the noise in the images is randomly Gaussian distributed.

Angular Assignment To reconstruct an initial 3D structure, every image (or class sum) needs a set of assigned shifts and angles. To achieve this, there are two possibilities. Either a reference structure is used (which again is prone to bias) or *de novo* in an approach called *angular reconstitution* (see Section 2.4.5.5).

Reconstruction After determining the initial set of angles to the class averages the initial 3D structure is created by smearing out the projections in the direction perpendicular to their respective image plane and applying a ramp function.

Refinement The initial model can then be projected and used as a reference for the alignment and the process continues iteratively until it converges into a (possibly local) resolution minimum.

1.2.1 RELION

An alternative approach for the determination of the five degrees of freedom for each particle image was described by Scheres [59]. REgularized LIkelihood OptiminzatioN (RELION) introduces a *maximum-likelihood* based algorithm. It requires a starting model for the refinement process and features a discrete model for the estimation of conformational dynamics.

1.3 Challenges

The former section briefly describes how to, in principle, reconstruct 3D structures from 2D cryoEM density images. The process itself is very tedious and prone to error and bias. There are multiple reasons for that. The most important aspects are sketched here.

1.3.1 Ill-posed Reconstruction Problem

cryoEM is an ill-posed reconstruction problem. The 2D density images extracted from micrographs lack information. One issue is illustrated in Figure 1.4. Every 2D projection from a 3D volume is inherently ambiguous, therefore, multiple projections are needed for reconstruction. Due to the geometry of the molecule and the sample preparation process it is impossible to cover every projection direction. This lack of certain views of the molecule affects the reconstruction accuracy tremendously. But not just the lack of certain views is a major issue, also the uneven distribution of views is a challenge. In fact, most of the molecules show only a handful of so called *preferred views* that contribute the vast majority of images to the dataset. In practice, this can lead to reconstruction results that are anisotropic, which means that they are smeared out in certain directions.

Another major issue is that the reconstruction algorithms themselves are affected by the low amount of signal in the images. This can potentially allow for multiple possible reconstruction results and therefore the optimization could get stuck in a local minimum. Even the choice of starting conditions and the randomized nature of the processing can lead to differing results.

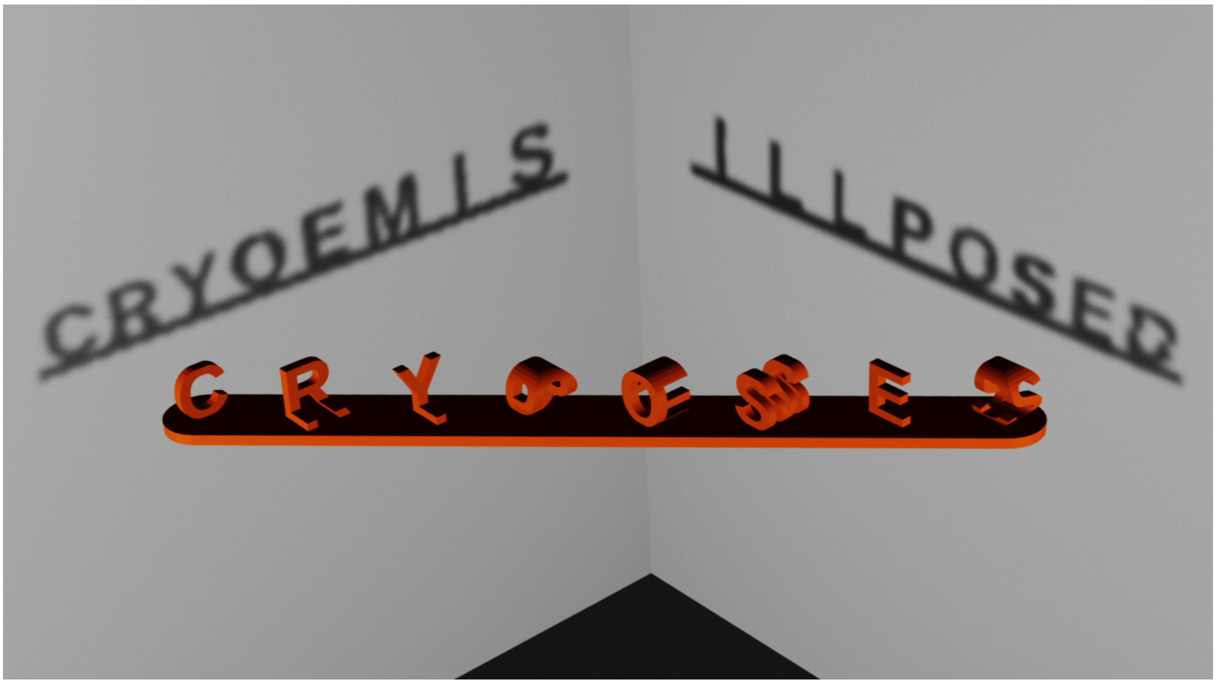


Fig. 1.4: This rendered scene shows a so called ambigram. The projection of every letter (depicted as shadows) into perpendicular projection directions show readable text, but contribute only half of the information needed to infer the over-all shape of the 3D object. This is also an issue cryoEM is affected of. Obviously, the more noise is present in the images, the more ambiguous the projection images become and the more of them are needed to distinguish certain features up to a point where it is impossible to distinguish them.

1.3.2 Noise

The high amount of noise present in cryoEM data induces major issues and has a high impact on the reconstruction accuracy. It contributes to both, over-fitting and under-fitting effects.

1.3.2.1 Over-fitting

Over-fitting occurs when the model fits the data too closely such that it is unable to generalize to unseen future events. Naturally, if the amount of noise present in the data is very high (like in cryoEM density maps) algorithms are prone to over-fit the data. This happens because noise is mistaken for signal and gets incorporated into the actual signal, which deteriorates the results of many integral steps of cryoEM reconstruction. For example it is more likely in 2D classification that two particles, showing different views, end up in the same class if the amount of noise is very high. Also it is worth noting that noise, due to its high-frequent nature, does not affect the reconstructed map equally. The more relevant high resolution information is therefore degraded disproportionally.

In general noise in cryoEM is assumed to be independent and Gaussian distributed. However, it is unclear to date if this assumption is too simplistic and more sophisticated models for the distribution of the noise could potentially be beneficial.

1.3.2.2 Under-fitting

Under-fitting occurs when the model fits the underlying data not well enough. Under-fitting occurs in cryoEM often in form of *model bias*. There are three main points in the processing where model bias is often induced.

The first is in particle picking when a projected 3D reference structure is used. This can result in picking noise patches that show certain characteristics that resemble actual particles. In a famous publication Henderson has shown that he was able to reconstruct a 2D picture of Albert Einstein from pure noise images [33].

The second situation where model bias can occur is the use of a 3D reference for the refinement process. This is illustrated in Figure 1.5.

The third situation is the use of symmetry in the data. Many macromolecular assemblies consist of subunits that resemble each other up to a certain resolution level. Some reconstruction algorithms make use of this fact by reconstructing multiples of these subunits which reduces the computational cost tremendously. However, this assumption is simplified, since at very high resolution levels even those subunits can differ. Applying a symmetry assumption on the molecule can therefore result in the inability to detect important details in the motion of the respective subunits.

In reality, of course, model bias is very subtle. Most of the picked particles are actual particles and contain signal, just a small fraction is pure noise. The bias is therefore much harder to detect. Not using a reference at all would therefore be very desirable.

1.3.3 Heterogeneity

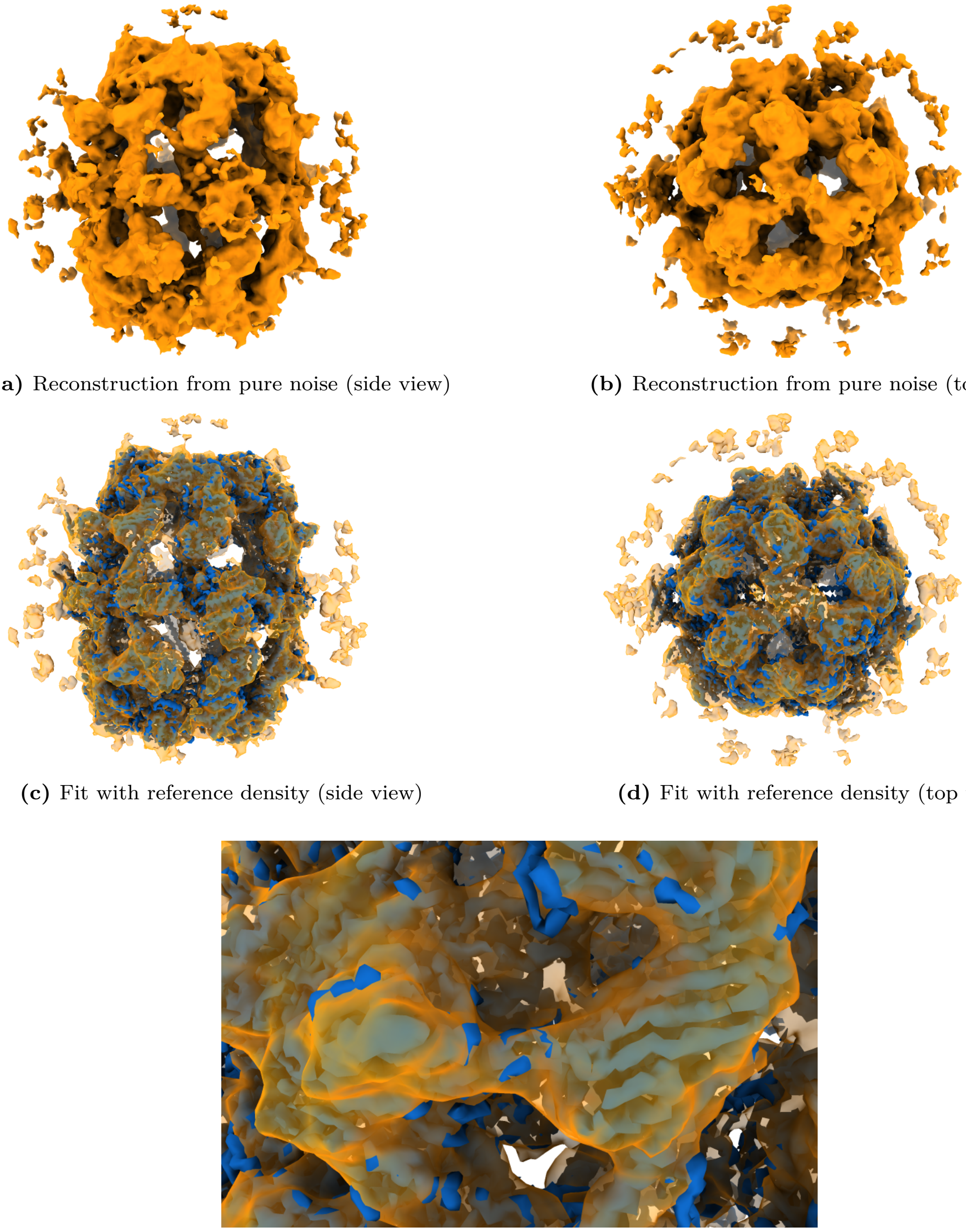
Heterogeneity is one of the main challenges in cryoEM. If present in the data it results in a deterioration of reconstructed 3D volumes. Sometimes, it is just a local phenomenon and can be seen in a blurring of dynamic parts of the molecule. If large domains or the whole molecule move, it can result in a non-convergence of the whole reconstruction workflow.

1.3.3.1 Conformational Heterogeneity

Both, the multivariate statistics and also the maximum likelihood approach from RELION makes use of a simplified image formation model, namely that each projection image from a micrograph is a distinct (but weighted) view of the same or a couple of most relevant *states* (usually less than hundred). They share the idea that the discrete amount of states can be used to approximate the (inherently continuous) so called *conformational landscape*. However, they rely on extensive manual sorting and also on some sort of interpolation between the determined discrete states in order to create a continuous trajectory. This is also difficult, because biological macromolecules tend to spend most of the time in stable states with only very short (that means hard to observe) transitions in-between them. An example for this kind of process can be seen in Figure 1.6.

1.3.3.2 Structural Heterogeneity

While conformational heterogeneity manifests as rigid body motion of biological macromolecules *structural heterogeneity* can be observed when molecules for example differ in their over-all structure, show additional ligands bound to the structure, disintegrate or suffer from structural defects. Being able to sort conformations based on those structural features can be beneficial. On the one hand, particles that show defects can be discarded, on the other hand, biologically relevant features can be extracted for further study.



(a) Reconstruction from pure noise (side view)

(b) Reconstruction from pure noise (top view)

(c) Fit with reference density (side view)

(d) Fit with reference density (top view)

(e) Study of some details

Fig. 1.5: Illustration of model bias. For this illustration grids with carbon foil, but without any particles were imaged in a TEM. Afterwards 20,000 pure noise images were extracted from the micrographs. These simulated "particle" images were used for reconstruction with a reference structure of the Fatty Acid Synthase (FAS) molecule. Note that the reference structure was coarsened by a factor of two in order to reduce computational complexity.
Top row: Reconstruction result from two different perspectives.
Middle row: Fit of the reconstruction result (orange, semi-transparent) with the reference structure (blue).
Bottom row: Zoomed perspective. It becomes apparent that the reconstruction algorithm over-fits towards the reference and not only matches the rough shape, but also reproduces relatively fine details.

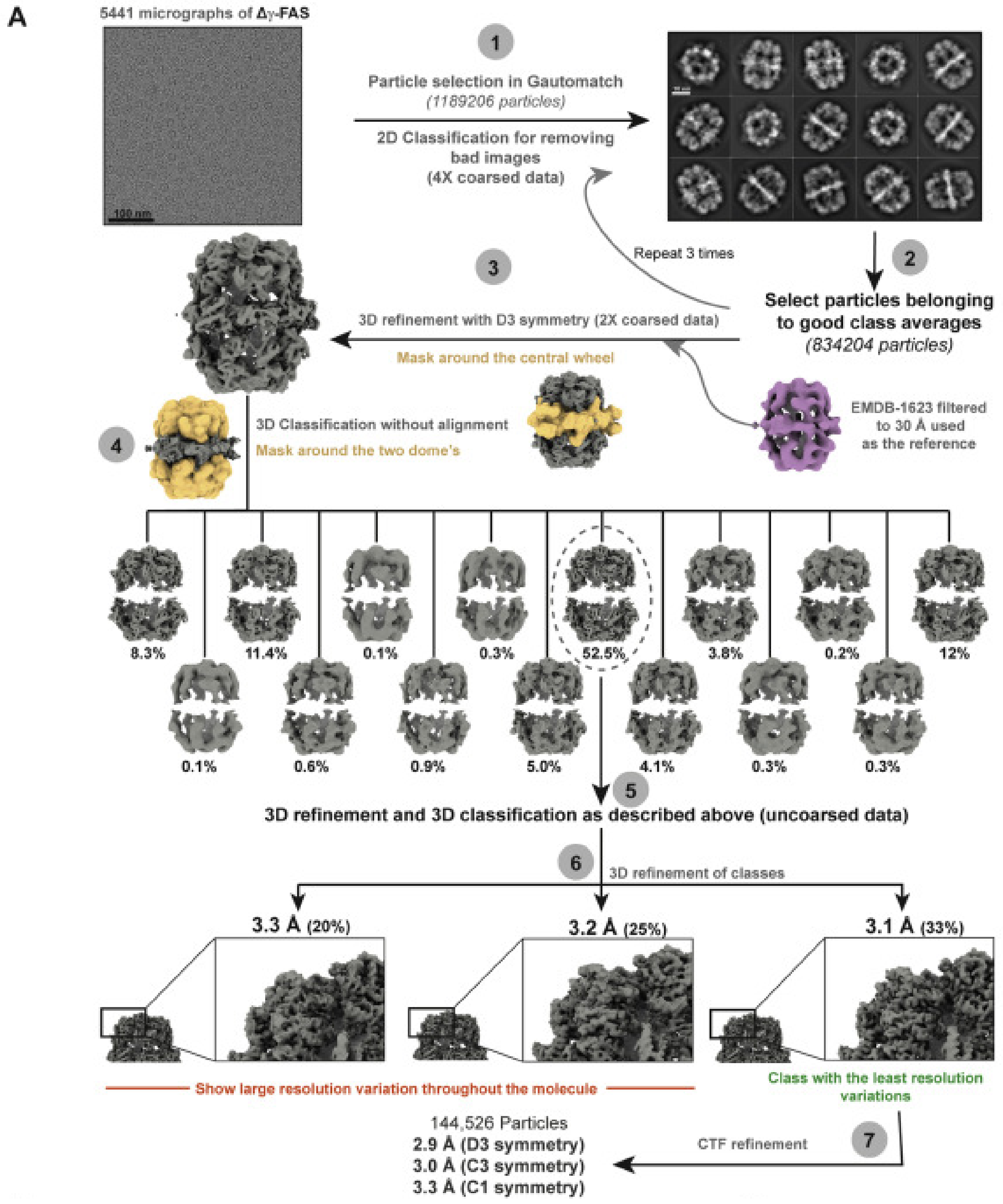

A
5441 micrographs of Δγ-FAS
100 nm
1
Particle selection in Gautomatch
(1189206 particles)
2D Classification for removing
bad images
(4X coarsed data)
Repeat 3 times
2
Select particles belonging
to good class averages
(834204 particles)
3
3D refinement with D3 symmetry (2X coarsed data)
Mask around the central wheel
EMDB-1623 filtered
to 30 Å used
as the reference
4
3D Classification without alignment
Mask around the two dome's
8.3%
11.4%
0.1%
0.3%
52.5%
3.8%
0.2%
12%
0.1%
0.6%
0.9%
5.0%
4.1%
0.3%
0.3%
5
3D refinement and 3D classification as described above (uncoarsed data)
6
3D refinement of classes
3.3 Å (20%)
3.2 Å (25%)
3.1 Å (33%)
Show large resolution variation throughout the molecule
Class with the least resolution
variations
144,526 Particles
2.9 Å (D3 symmetry)
3.0 Å (C3 symmetry)
3.3 Å (C1 symmetry)
CTF refinement
7

Fig. 1.6: Illustration of hierarchical conformational sorting, which involves 2D classification, a low-pass filtered reference structure, selective masking of dominant rigid subunits and refining the selected discrete subclasses. This illustration is taken and adapted from [65] with courtesy of Kashish Singh.

1.3.3.3 Conformational Landscape as Manifold

A less simplified model is to think of every particle image as a distinct entity that is only loosely connected to all the other particles. Every single projection image can then be modeled as a discrete sample of a continuous conformational space of (projected) 3D volumes. Mathematically, this space can be modeled as a *manifold*, that is a low-dimensional subspace embedded into the high dimensional voxel space of all possible 3D volumes. This

manifold locally resembles an Euclidean space. One can think of a d-dimensional manifold in an n-dimensional space as a bend d-dimensional hyperplane. This corresponds to the so called *manifold hypothesis*, which states that high dimensional data tends to concentrate around low dimensional subspaces embedded in the high dimensional space.
Following this approach the objective of cryoEM becomes the approximation of this conformational manifold.

1.3.4 The Zoo of Tools in CryoEM

As described before, cryoEM involves lots of steps to get from the raw stacks of micrographs to finished 3D structures and trajectories in between them. Many of these steps require lots of human intervention and it is very natural that the automation of this process increased over time. However, this led to a variety of specific solutions for specific sub tasks rather than software packages that tried to offer automation for the whole process from start till end.

Without any doubt, RELION (see Section 1.2.1) is the most used software to date, but even this software is usually paired with other tools to e.g. perform particle picking (e.g. gautomatch) or CTF correction (e.g. gctf) and for the creation of starting models (e.g. PRIME, SIMPLE) for the refinement process.

One software package that was designed with the idea in mind that the whole process could be handled in one environment is the so called *Cow Suite* (see Section 2.1), developed in the Department of Structural Dynamics at Max-Planck-Institute for biophysical Chemistry led by Prof. Holger Stark. Unlike many other software packages on the market it offers advanced project management capabilities and also features interfaces for external software tools to be integrated into the workflow. The software suite offers e. g. tools for particle picking [15], the assessment of micrograph quality [62] and tools to automatize the whole cryoEM processing cycle including alignment and classification [48], visualization and validation [42], for sorting of heterogeneous datasets [46], visual programming and much more. Cow Suite was designed with accessibility in mind, therefore it is a platform independent software package (which means it runs on Windows, Linux and macOS), heterogeneous infrastructures (like HPC, but also local stand-alone machines).

This set of tools, often used in combination, makes it hard to reproduce and validate the results obtained. It would be desirable to streamline the process in a way that the amount of tools can be reduced further.

1.4 The Rise of Machine Learning

The roots of machine learning date back to the late 50s of the 20^{th} century. In a famous publication Samuel stated that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program [58], which is often turned into a definition of machine learning assigned to him: "Field of study that gives computers the ability to learn without being explicitly programmed.“

Machine learning is usually classified into three main approaches. *Supervised learning*, *unsupervised learning* and *reinforcement learning*. The former will be sketched briefly in the following sections.

1.4.1 General approaches

1.4.1.1 Unsupervised Learning

Unsupervised learning uses unlabeled training data. The task of the machine learning algorithm is then to autonomously find patterns in the data. A very common scenario in unsupervised learning is the clustering of data, that is grouping the data based on certain features (e.g. k-means clustering [47]).

1.4.1.2 Supervised Learning

Supervised learning uses training data that was previously labeled, which means the desired output with respect to the input is known in advance. This data is then used to train a model, which is capable of finding a mapping between training data and its respective labels. A typical application is *classification*, e.g. the process of particle picking in cryoEM.

However, it is often impractical to have access to labeled training data, due to the complexity of manual labeling. In many cases, it is also virtually impossible to manually assign labels to training data, especially in the presence of vast amounts of noise. Hence, one needs to resort to *semi-supervised learning* (i.e. partially labeling data) or *self-supervised learning* (i.e. training the network with unlabeled data, but also using the data as label). Technically, the latter is a mix of supervised and unsupervised approaches.

1.4.2 Representation- and Deep Learning

Representation learning has emerged to be a field on its own in machine learning, as it touches supervised and unsupervised learning. Bengio et al. define representation learning as learning representations of the data that make it easier to extract useful information when building classifiers or other predictors [7]. That means it is not only important to find an internal representation which describes the data sufficiently, but also to organize it in a way that it is (potentially human) interpretable.

Multiple approaches exist to approximate good representations. *Deep learning* based methods have become more and more popular in recent years. Deep learning makes use of the fact that Artificial Neural Networks (ANNs) (see Section 2.5.2) are universal functional approximators [36], which means that given an input x and an expected output y a neural network is (ideally) capable of finding a mapping f such that $y = f(x)$.[1]. In practice it turned out that stacking ANNs deeper enables them to learn better internal representations that are more robust, and it makes them easier to train.

1.4.3 (Variational) Autoencoder

Autoencoder are ANNs that consist of three main parts. An *encoder*, a *bottleneck* and a *decoder*. They aim to learn efficient representations (encodings) of the training data. The encoder thereby tries to find a mapping of the input data into the so called *coding* or *latent* space such that it preserves most of the information, the decoder on the other hand tries to reproduce the input data based on the latent space as good as possible. It can be shown that a linear autoencoder can produce a latent space that resembles the space of Principle Component Analysis (PCA) (see Section 2.3.8).

Due to the capability of the decoder to generate output data that resembles the input data they are considered so called *generative models*. However, the ability to generate output data is limited in vanilla autoencoder.

Variational Autoencoder (VAE) solve this problem by parameterizing a mixture of distributions rather than a fixed encoding. This enables to sample from this mixture model and to thereby generate data that resembles data of the input space.

[1] This especially means that if one thinks of cryoEM as a mapping f of an arbitrary stack of 2D input images to a continuous conformational space of 3D structures, there must be an ANN which parametrizes f (but it might be hard to find/train).

1.5 Aim of the work

cryoEM is a method in structural biology that has proven to be of tremendous importance for the structure determination of biological macromolecular machines. Although reconstruction algorithms in cryoEM are becoming more sophisticated in practice there are still lots of hurdles to overcome, especially when it comes to the processing of highly dynamic or in general heterogeneous data, due to the discrete image formation model used to date.

cryoEM also suffers from the problem that there are lots of specialized tools available that do not integrate very well and produce varying results, such that the 3D density maps obtained by the work-flow depend a lot on the tools chosen and also on the respective user. Often references for picking, masks for rigid regions and starting models are being used that potentially introduce bias.

Aim of this thesis is to introduce a fully autonomous tool that can process a stack of particle images and does not rely on any prior information of the molecule, which especially means it does not rely on starting structures, assumptions on symmetry or the distribution of noise. It is designed to unify the work-flow such that it becomes less dependent on the user and the choice of tools while still remaining compatible with other software such as Cow and RELION. The image formation model will be extended in a way that it features an inherently continuous conformational space that can be used to draw samples from. To approximate this space machine learning is used in form of self-supervised ANNs called VAEs. Although being specifically designed for conformational heterogeneity the algorithm is also evaluated for its capability to generalizes to structural heterogeneity, like broken particles.

Furthermore the tool presented in this thesis is supposed to run on heterogeneous hardware, such as single user workstations, but also in an HPC environment on multiple graphics processing units (GPUs) and also multiple computing nodes simultaneously.

Outline Chapter 2 describes the fundamental concepts used in this thesis. It starts with introducing mathematical preliminaries and how they apply to cryoEM. Section 2.5 elucidates fundamental ideas of machine learning. Results exhibits how the methods presented in Chapter 2 can be combined to create a tool which is able to estimate conformational landscapes from 2D projection images of cryoEM density maps, while significantly simplifying the overall workflow. This tool is then applied to several datasets. Chapter 4 is an evaluation of the results obtained in the previous chapter. Chapter 5 finally summarizes the work and points towards questions that could be addressed in future work.

2 Materials and Methods

Computer science is no more about computers than astronomy is about telescopes, biology is about microscopes or chemistry is about beakers and test tubes. Science is not about tools. It is about how we use them, and what we find out when we do.
– Edsger W. Dijkstra

2.1 Software

Software	Link	Reference
Python	https://www.python.org/downloads/	[74]
Tensorflow	https://www.tensorflow.org	[2]
Horovod	https://github.com/horovod/horovod	[63]
mrcfile	https://github.com/ccpem/mrcfile	[14]
Relion	https://github.com/3dem/relion	[59]
CowSuite	https://www.cow-em.de	[15, 42, 46, 48, 62]
Matlab	https://www.mathworks.com	[50]
ChimeraX	https://www.rbvi.ucsf.edu/chimerax/	[53]
Singularity	https://syslabs.io	[45]
Blender	https://www.blender.org	[18]

2.2 Hardware

All calculations were performed on the High Performance Computing (HPC) cluster of the Department of Structural Dynamics at Max-Planck-Institute for biophysical Chemistry. The computers have the following configuration:

- CPUs: 24x Intel(R) Xeon(R) Gold 6128 CPU 3.40GHz
- RAM: 376GB
- GPUs: 4x NVIDIA Tesla V100 SXM2 32GB

Up to 24 nodes with this configuration are available for calculation and can be used simultaneously.

2.3 Mathematical Preliminaries

This chapter introduces the fundamental mathematical concepts for understanding the cryoEM work-flow Section 2.4 as well as for the proposed algorithm. It starts with basic image statistics and normalization strategies ...

2.3.1 Image Statistics

In the context of this thesis a two-dimensional (2D) image is defined as $x \in \mathbb{R}^{n \times n}$, whereas a three-dimensional (3D) image is defined as $x \in \mathbb{R}^{n \times n \times n}$. Images are generally assumed to be in squared or cubic shape, however they can be reshaped into their one-dimensional (1D) representation denoted by $\mathbf{x} \in \mathbb{R}^{n^2}$ or $\mathbf{x} \in \mathbb{R}^{n^3}$ respectively.

Arithmetic Mean The image mean is defined by the sum over all the intensity values x_i in an image divided by the total number of values.

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.1}$$

Variance The variance of an image can be expressed by the squared difference of every intensity value x_i of the image from its mean μ divided by the total number of intensity values in the image and is denoted σ^2. The variance is always ≥ 0

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{2.2}$$

Standard Deviation The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\sigma^2} \tag{2.3}$$

Euclidean Norm The euclidean norm is often called *L2 norm.* It quantifies the length of a vector (which is equivalent to the distance from the origin of the vector space):

$$\|x\| = \sqrt{\sum_{i}^{n} x_i^2} \tag{2.4}$$

It can also be used to express the euclidean distance between two vectors in the same vector space:

$$\|(x_1 - x_2)\| = d(x_1, x_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \tag{2.5}$$

The euclidean distance between two vectors is strongly affected by noise and therefore performs poorly in many scenarios in 3D transmission electron cryo-microscopy (cryoEM).

Cross Correlation Let x_1, x_2 be two images with $x_i \in \mathbb{R}^n$ then the cross correlation coefficient (CCC) is defined as:

$$\mathrm{CCC} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{x}_1\left(r_{\mathrm{i}}\right) - \overline{\mathrm{x}}_1\right)\left(\mathrm{x}_2\left(r_{\mathrm{i}}\right) - \overline{\mathrm{x}}_2\right)}{\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{x}_1\left(r_{\mathrm{i}}\right) - \overline{\mathrm{x}}_1\right)^2 \sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{x}_2\left(r_{\mathrm{i}}\right) - \overline{\mathrm{x}}_2\right)^2}} \tag{2.6}$$

With:

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_i\left(r_j\right); \quad i = 1, 2 \tag{2.7}$$

The CCC is a metric which is often used in cryoEM to measure the similarity of images.

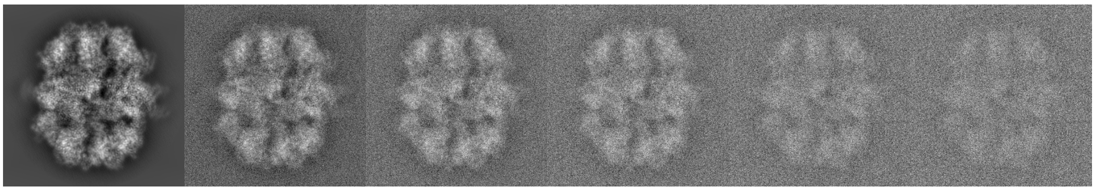
SNR The signal-to-noise ratio (SNR) quantifies the amount of signal present in an image compared to the amount of noise. It is therefore defined by the ratio of the power of the signal and the power of the noise:

$$SNR = \frac{P_{Signal}}{P_{Noise}} \tag{2.8}$$

Another very common definition, especially used in automated image processing is:

$$SNR = \frac{\mu}{\sigma} \tag{2.9}$$

Here, μ denotes the mean of an image and σ is the standard deviation of the noise. In typical image processing scenarios the SNR is assumed to be 5 or higher to be able to distinguish image features. This is known as the Rose criterion [56]. However the SNR in cryoEM is usually 0.1 or even below that limit. Which means that the amount of noise present in the images exceeds the signal by far. Figure 2.1 shows projection images with added synthetic gaussian noise. It becomes apparent that the lower the SNR is the harder it becomes to distinguish especially small features in the images. This is due to the fact that small features in the projection images resemble the intensity distribution of high frequent noise, making it significantly harder to reconstruct these features compared to large domains.



(a) Particle (b) SNR Level 1 (c) SNR Level 0.5 (d) SNR Level 0.25 (e) SNR Level 0.1 (f) SNR Level 0.05

Fig. 2.1: Comparison of different SNR levels. a shows the raw particle image(EMD-4577 in EMDB). The other images show a decrease in SNR due to added synthetic gaussian noise. While the overall shape remains barely visible to the human eye, small features quickly become indistinguishable the more noise is added. In real-world conditions SNR levels of particle images are usually in the range of **(e)** or **(f)**

2.3.2 Normalization Strategies

The distribution of intensity values in cryoEM images is heavily influenced by various recording conditions, e.g. the thickness of the ice, electron dose, isotropy of the electron beam). It varies between different datasets, but also within one dataset and even in the same micrograph. Therefore, it is necessary to normalize them in order to make them comparable. This section compares basic normalization strategies and introduces a technique not yet used in cryoEM. For an in-depth review on different normalization strategies see [67].

Min-Max Normalization The first normalization technique that comes to mind is the so called min-max normalization, which is defined by:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2.10}$$

Its advantage is that it squashes all the values of x into an interval of $[0, 1]$. However, due to the presence of noise it is heavily influenced by outliers.

z-score Standardization The most used normalization strategy in cryoEM is the z-score standardization. It is used to equalize mean μ and standard deviation σ of the images. It is defined by:

$$x' = \frac{x - \mu}{\sigma} \tag{2.11}$$

Since the intensity values of an image are assumed to be gaussian distributed this results in a centering of the normal distributions of the images and an equalization of their respective standard deviations.

Tanh Estimator The tanh Estimator is a normalization strategy that on the one hand is more robust to outliers compared to the min-max normalization, but also squashes the intensity values into the interval $[0, 1]$. It is defined by:

$$x' = 0.5 \left[\tanh \left(\lambda \cdot \frac{x - \mu}{\sigma} \right) + 1 \right] \tag{2.12}$$

Where μ again is the mean and σ the standard deviation. λ is called the spread.

2.3.3 Fourier Transformation

Fourier Transformation is a technique which is widely used in signal processing. It was developed in the 19th century [5, 27]. Fourier transformation decomposes a signal into a weighted sum of sine and cosine components. In image processing that means to split the information in the image that is mainly characterized by the spatial distribution of its intensity values into its frequency components.

The Fourier transform $\mathcal{F}$ of a continuous and integrable function $f(x)$ can be expressed by:

$$\mathcal{F}(y) = \int_{\mathbb{R}^n} f(x) \mathrm{e}^{-\mathrm{i}yx} \, dx \tag{2.13}$$

This transformation is fully reversible up to some numerical error. The inverse transformation is called *inverse Fourier transform* and can be calculated such that:

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathcal{F}(y) \mathrm{e}^{\mathrm{i}yx} \, dy \tag{2.14}$$

In image processing, however, the signals are of discrete nature and the integrals become sums:

$$\mathcal{F}(y) = \sum_{k=0}^{N-1} f(x) \mathrm{e}^{\frac{-\mathrm{i}yx}{N}} \tag{2.15}$$

Analog for the inverse discrete Fourier transform:

$$f(x) = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{F}(y) \mathrm{e}^{\frac{\mathrm{i}yx}{N}} \tag{2.16}$$

The discrete Fourier transform of a 2D image can then be calculated by Fourier transforming along one dimension and then again Fourier transforming the result in the other direction. High performance algorithms for the calculation of the Fourier transform exist, of which the most well known is the so called Fast Fourier transformation (FFT) (and inverse Fast Fourier transformation (iFFT)) [12].

2.3.4 Central Slice Theorem

The central slice theorem (CST) is essential to many 3D reconstruction algorithms in cryoEM. It states that the Fourier transform of a two-dimensional projection of a three-dimensional object is identical with the corresponding central section of the three-dimensional transform of the object [20]. It is therefore possible to construct a 3D volume slice by slice using the Fourier transformed projection views of the object.

$$\tilde{X}(u,v,w) = \iiint_{\text{object}} \Big(X(x,y,z) \ \cdot \ \exp(2\pi \mathrm{i}(xu+yv+zw)) \Big)\, dx\, dy\, dz \tag{2.17}$$

The central section ($z = 0$) of the transform then becomes:

$$\tilde{X}(u,v,0) = \iint \Big(p(x,y) \ \cdot \ \exp(2\pi \mathrm{i}(xu+yv)) \Big)\, dx\, dy \tag{2.18}$$

Where

$$p(x,y) = \int X(x,y,z)\, dz \tag{2.19}$$

is the projection of the density X parallel to the z-axis. The concept of the CST is illustrated in Figure 2.2.

For a comparison of different Fourier related transforms and how they apply to the Central Slice theorem see Bracewell [10].

2.3.5 Filtering

Multiple filtering algorithms exist in image processing. They are often used for feature extraction or to pronounce/suppress certain features. The most prominent filtering techniques in cryoEM are the *gaussian low pass* and the *gaussian high pass filter*. The principle is depicted in Figure 2.3. The low pass filter eliminates high frequency information (noise, but also small features in the images). The high pass filter on the other hand suppresses low frequency information and thereby pronounces high resolution features. The combination of both of them is also very common and called *gaussian band pass filter*.

All of the mentioned filter make use of the Fourier transform (see Section 2.3.3), by selectively masking certain areas in the frequency domain. The masks applied to the fourier transformed images are also shown in Figure 2.3. Low pass filtering is often used in cryoEM to artificially lower the resolution of reference structures (resp. starting models) to reduce over-fitting effects (Section 1.3.2.1).

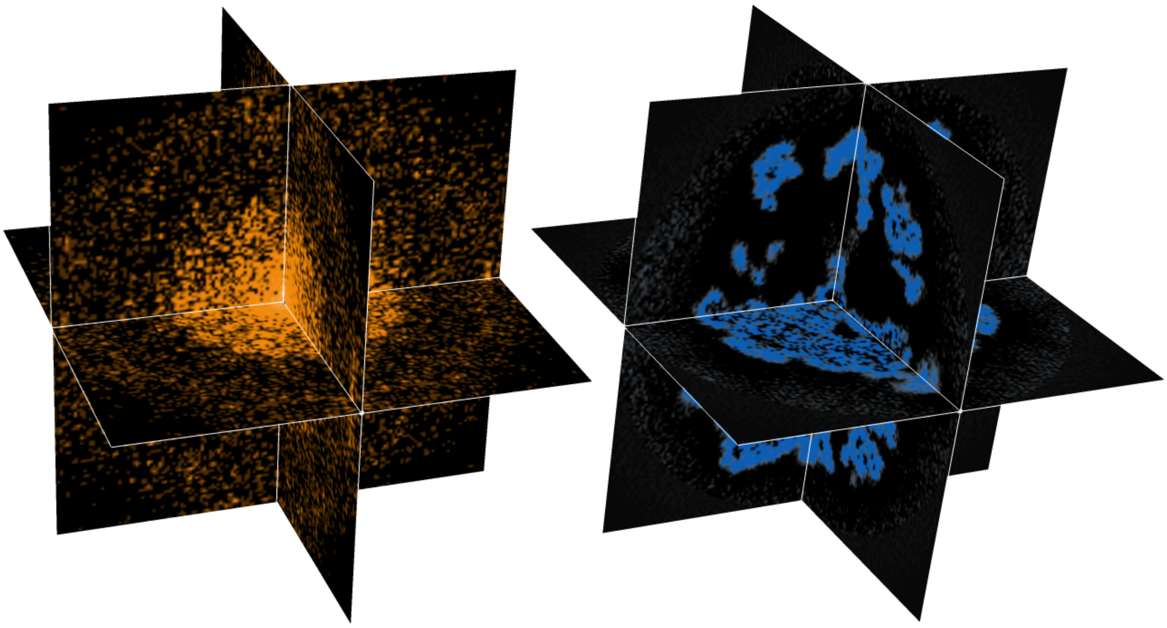


Fig. 2.2: Illustration of the CST. Fourier transformed projections (orange) of the same volume share a common line. In the absence of noise, three perpendicular projections are enough to describe the whole molecule entirely. For comparison the corresponding slices in real space are depicted in blue on the right.

2.3.6 Coarsening

Coarsening is a down-sampling operation often used in image processing. Coarsening merges blocks of neighboring pixels into one single pixel by averaging them. This has two advantages. On the one hand the image is reduced in size and can therefore be processed more easily. On the other hand the SNR in the image is boosted due to the fact that noise, unlike signal, is assumed to be random and independent and identically distributed (i.i.d.). The coarsened image X_c can be calculated by the following formula:

$$X_c(x,y) = \frac{1}{c}\sum_{i=0}^{c}\sum_{j=0}^{c} X(x+j,y+i) \tag{2.20}$$

Where X is the original image and c is the *coarsening factor*. The number of pixels n of the original image X is reduced to $n_c = \frac{n}{c^2}$ in the coarsened image X_c. The maximum obtainable resolution (see Section 2.4.6.1) for the reconstruction from coarsened images, however, is also reduced by the coarsening factor.

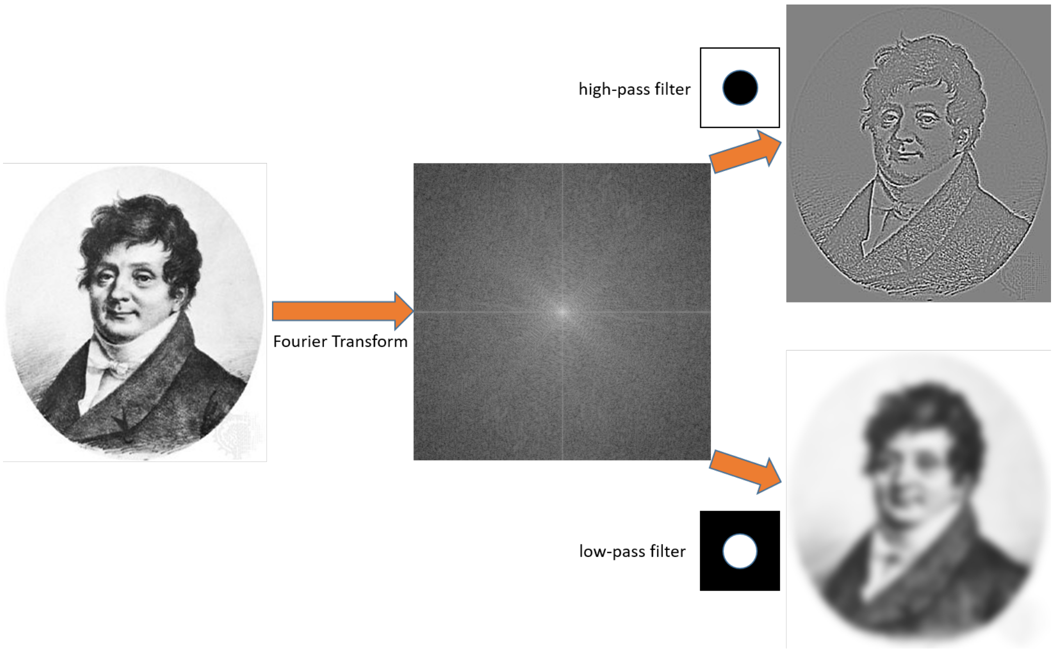


Fig. 2.3: Effects of filtering in Fourier space. **left:** Jean-Baptiste Joseph Fourier, **middle:** Fourier Transform of the image, **right:** after inverse fourier transformation low-pass filtered image shows less detail, whereas high-pass filtered image emphasizes on details. The portrait of Joseph Fourier was taken from [17] and was released by the author into public domain.

2.3.7 Image Rotation and Projection

2.3.7.1 Rotation in 2D

A rotation of a vector x in 2D space around an angle γ can be described by a matrix vector multiplication with a rotation matrix R.

$$x' = Rx \tag{2.21}$$

With:

$$R = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \tag{2.22}$$

2.3.7.2 Rotation in 3D

Analog to Equation 2.22 a 3D rotation can be expressed as three single rotations around predefined axis. For the convention ZYZ' this is:

Euler Angles

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} R_y = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} R_z = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.23}$$

$$R_{zyz'} = R_z(\alpha) \cdot R_y(\beta) \cdot R_{z'}(\gamma) \tag{2.24}$$

The triple (α, β, γ) is called a set of *Euler angles*. Although being a very intuitive description of 3D rotations it has some drawbacks. First it is ambiguous, which means that there are multiple sets of Euler angles that describe the same rotation. Second it suffers from the so called *gimbal lock*. Gimbal lock occurs when two of the rotation axis become parallel to each other which means that the space of possible rotations degenerates into a 2D space.

(Unit) Quaternions A quaternion $q \in \mathbb{H}$ is a hypercomplex number of the form

$$q = a + bi + cj + dk \tag{2.25}$$

with $a, b, c, d \in \mathbb{R}$ and i, j, k as the *basic quaternions*. The latter especially means that:

$$i^2 = j^2 = k^2 = ijk = -1 \tag{2.26}$$

The conjugate of a quaternion q, denoted q^* is defined as:

$$q^* = a - bi - cj - dk \tag{2.27}$$

A unit quaternion is a normalized quaternion (norm 1) that can be used to describe rotations in $\mathbb{R}^3$:

$$q_u = \frac{q}{\|q\|} \tag{2.28}$$

with

$$\|q\| = \sqrt{qq^*} = \sqrt{a^2 + b^2 + c^2 + d^2} \tag{2.29}$$

Unit quaternions solve the issues Euler angles suffer from, but have the disadvantage of double covering the space of rotations, because the complex conjugate q^* of a quaternion q describes the same rotation as the quaternion itself.

2.3.7.3 Projection

The projection p of a volume $X \in \mathbb{R}^{n \times n \times n}$ in z direction can be expressed as a line integral:

$$p(x, y) = \int X(x, y, z)\, dz \tag{2.30}$$

which in the discrete case means:

$$p(x, y) = \sum_{z=1}^{n} X(x, y, z) \tag{2.31}$$

Note that every projection of a volume X in an arbitrary direction can be expressed by first rotating X in the respective direction and then projecting it along z.

2.3.8 Dimensionality Reduction

Dimensionality reduction is an important concept to make sense out of high dimensional data. It follows the idea that high dimensional data tends to concentrate along low dimensional subspaces that are embedded into the high dimensional space. A technique which is often used in cryoEM is the so called Principle Component Analysis (PCA), since it can be calculated very efficiently and the results are visually interpretable to a certain extent (see Figure 2.4).

PCA The key idea behind PCA is to change the basis of the input data such that the basis vectors are oriented towards the directions of the highest variance. To achieve this in a first step the covariance matrix is being calculated:

$$cov(X, Y) = Q = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ & \ddots & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \tag{2.32}$$

Afterwards the covariance matrix is decomposed by eigendecomposition into a product of three components:

$$Q = P \Lambda P^{-1} \tag{2.33}$$

Where P is the matrix of all (column) eigenvectors ω_i and Λ is a diagonal matrix of corresponding eigenvalues λ_i:

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \tag{2.34}$$

Since Q is a positive semi definite matrix the inverse of P is equivalent to its transpose:

$$P^{-1} = P^T \tag{2.35}$$

Sorting the pairs (ω_i, λ_i) with respect to the value of λ_i yields a matrix $\hat{P}$ that can be easily used for dimensionality reduction. The contribution of each eigenvector ω_i to the total variance in the data can be calculated by the ratio:

$$\frac{\lambda_i}{\sum_{(i)} \lambda_i} \tag{2.36}$$

The actual dimensionality reduction is then performed by skipping eigenvectors with a low contribution to the overall variance:

$$\hat{P}^* = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_k & 0 & \cdots & 0 \end{bmatrix}, k < n \tag{2.37}$$

and projecting X into the lower k dimensional subspace X_{PCA}:

$$X_{PCA} = X\hat{P}^* \tag{2.38}$$

Figure 2.4 depicts the first 10 eigenvectors of a stack of 10,000 particles. In the top row PCA is applied to the raw images. In the bottom row particles are roughly aligned before being transformed. The first components in each row contribute more to the overall shape of the particle, whereas the later components are associated with finer details, but are also more noisy. In cryoEM eigenvectors are sometimes called *eigenimages*. It is clearly visible that the quality of the results obtained by PCA is strongly influenced by the orientation (that is in-plane rotation and shifts in x- and y-direction) of the particles. While the bottom row shows eigenimages that match the overall shape of the particle fairly well, the results in the top row remain poor. For details on the alignment process see Section 2.4.5.3.

t-distributed stochastic neighbor embedding (t-SNE) t-distributed stochastic neighbor embedding (t-SNE) is a statistical tool for non-linear dimensionality reduction. SNE was first introduced by Hinton and Roweis in [35] and then modified to use a t-distribution

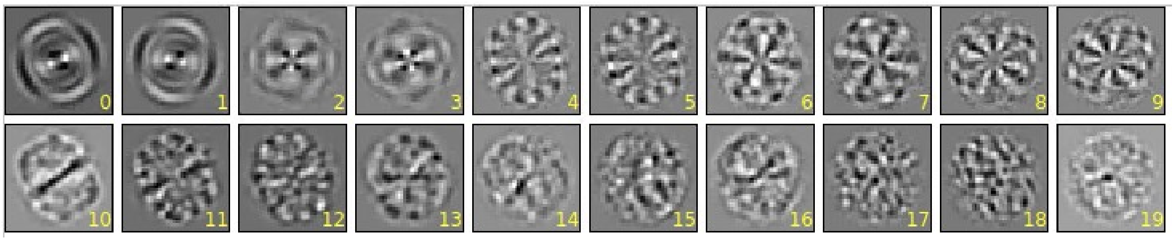


Fig. 2.4: *Top row:* First 10 eigenimages from a dataset of 10,000 randomly oriented particle images. The eigenimages show some features but are heavily influenced by the orientation of the particles. *Bottom row:* First 10 eigenimages of the same 10,000 images that were roughly aligned before being transformed by PCA. The resulting eigenimages clearly show features of the particles. The first vectors show the main shape of the particles, while the less important eigenvectors show small features.

by Van der Maaten and Hinton [69]. It is often used in machine learning to visualize high dimensional data in low dimensional subspaces. t-SNE tries to group similar vectors together. While the distances between the clusters are distorted the distances within a cluster are preserved.

The first step is to assign a probability for each element that describes its similarity to every other element $i \neq j$:

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k\neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2\right)} \tag{2.39}$$

Since only distances between non equal elements are interesting here, $p_i i = 0$. Since $p_{j|i}$ is a probability distribution it follows: $\sum_{(j)} p_{j|i} = 1$ for all j.

A normalization step yields:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \tag{2.40}$$

Now t-SNE finds a mapping of a lower dimensional t-distribution q such that:

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_k \sum_{l\neq k} \left(1 + \left\|y_k - y_l\right\|^2\right)^{-1}} \tag{2.41}$$

This is done by performing *gradient descent* optimization on the Kullback–Leibler (KL)-divergence of p and q:

$$\mathrm{KL}(P\|Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{2.42}$$

2.4 Fundamentals of Single Particle Electron Microscopy

2.4.1 Instrumentation

It was Ernst Abbe who first stated that the resolution of a microscope is limited by the wavelength. That is called *Abbe's diffraction limit* [3]:

$$r = \frac{\lambda}{2n\sin(\theta)} \tag{2.43}$$

Here r is the maximum obtainable resolution, λ is the wavelength, and $n\sin(\theta)$ is the so called *Numerical Aperture* of the objective lens. In light-microscopy, the latter (in theory) results in an optimal resolution of half the wavelength. In practice that means a maximum resolution for light microscopes of around 200 Å. Some techniques were introduced to reach higher resolution levels by selectively illuminating certain fluorophors (e.g. STED [32]). However, their resolution power is still far away from what is needed to study biological macromolecules at atomic or near atomic resolution.

Back in 1924 de Broglie realized that not just visible light (photons), but every moving particle with a mass features a wavelength [22]:

$$\lambda = \frac{h}{mv} \tag{2.44}$$

With m and v being mass and velocity of the particle and h being *Planck's constant*. Since electrons exhibit much shorter wavelengths than visible light, theoretically, and practically, that much higher resolutions may be achieved in a microscopy application (see Section 2.4.3). The first electron microscope (EM) was then introduced by Ernst Ruska and Max Knoll in 1931 [43], featuring a lens invented by Knoll. The fundamental principle is still the same to date.

2.4.1.1 Transmission Electron Microscope (TEM)

Figure 2.5 depicts the structure of a TEM. The beam of electrons is created in the Field Emission Gun (FEG) by applying a strong electric field to an electron source, typically in the range of 100- 300keV. In high resolution EMs this source, the cathode, is often a

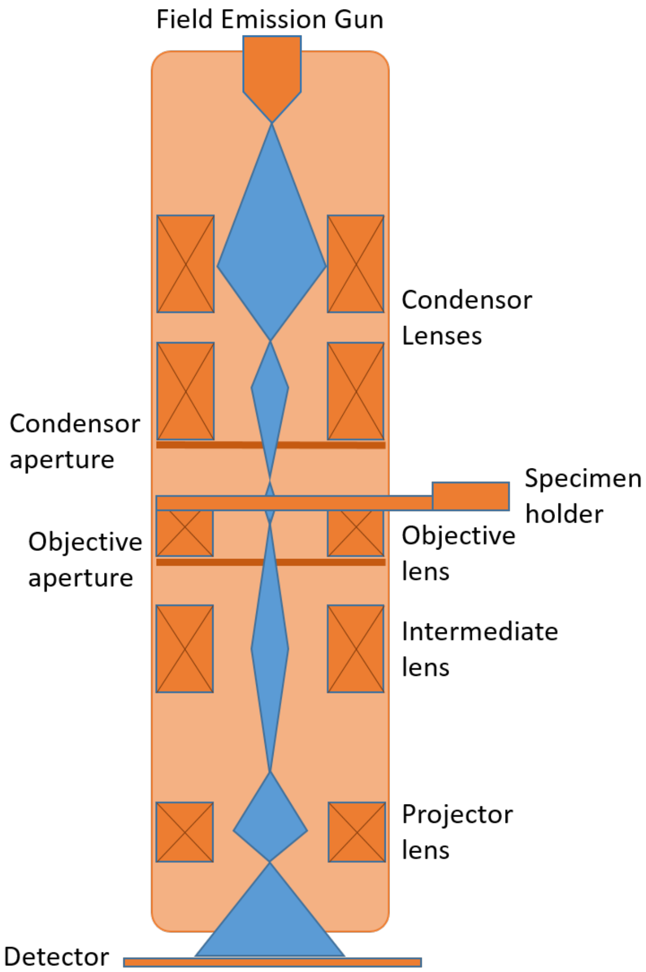


Fig. 2.5: Schematic of a transmission electron microscope (TEM). From top to bottom: Field Emission Gun (FEG) emits the electron beam because of a high acceleration voltage. Condensor lenses are shaping the beam and can adjust its spot size. Condensor apertures are blocking electrons that are too far away from the optical axis. The specimen holder is used to load the sample into the microscope. The objective lens is the first image forming lens. It creates a diffraction pattern of the sample in its back focal plane and a real space image in the image plane. Again the objective aperture blocks electrons that are scattered in undesired directions. Intermediate lenses are further magnifying the image which then in a last step is projected by the projector lenses onto the detector.

tungsten crystal coated with zirconium dioxide [77]. The high current in the flat anode underneath results in an extraction of electrons from the pointy tip of the crystal. This can be supported by heating the crystal (warm FEG) or leaving it at lower/room temperatures (cold FEG). The aim is to create a beam which is both spatially and temporally coherent, which means that all electrons are ideally extracted from exactly the same spatial position of the tip and at the same time feature the same speed (energy) with respect to every individual electron. It is important that the column of the microscope maintains a vacuum, since contaminations with impurity atoms (such as molecules contained in air or water) reduce the free path length and result in unwanted scattering effects, hence, increasing image noise[26, 77].

In a next step the electrons are accelerated in an electric field and the beam is shaped by the *condensor lens system.* The condensor lens system consists of electromagnetic lenses that control the shape of the beam (i.e. elliptically illuminated area) and its intensity (spot size). In principle, an electromagnetic lens system consists of four parts. *Deflectors* moves the beam towards the optical axis of the actual lens. The lens itself consists of a coiled conductor, in combination with a current, acting as an electromagnet, which applies a *Lorentz force* to focus the beam of electrons. A *stigmator* corrects the beam i.e. for astigmatism, which means it compensates for the effect that the lens does not focus the beam evenly over its whole area. The last part of the lens is the so called *lens aperture* which is basically a pinhole that blocks all the electrons that are too far away from the optical axis which would not contribute to the image information but rather result in blurring [26, 77].

The beam is then shaped by further electromagnetic lenses of the objective lens and eventually penetrates the sample, resulting in electron scattering. The beam is then reshaped by the *objective lens.* This lens focuses electrons scattered in the same direction in the same point, thus creating a diffraction pattern of the specimen in its back focal plane. The *intermediate lens* magnifies either the diffraction pattern from the back focal plane (which is the Fourier transformed image) or the actual (real space) image. The task of the *projector lens* is to project the final image onto the *detector.* Depending on the type of microscope there are different kinds of detectors available. charge coupled device (CCD) sensors convert electrons using scintillators into photons that then can be detected. More expensive direct electron detection device (DDD) sensors are able to directly detect electrons resulting in a much better SNR (see Section 2.3.1). A much more exhaustive description of the anatomy of modern TEMs can be found in [26, 77].

2.4.1.2 Aberrations in the TEM

The imaging quality of a TEM is strongly influenced by various factors. The most important aberrations that can arise during imaging are sketched here briefly. While it is best to compensate for those aberrations directly in the microscope, some effects are often corrected *in silico* during image processing (see Section 2.4.4).

Spherical Aberration (Cs) Spherical aberration is the most common aberration in EM. It is induced due to a longer focal length for beams parallel to the optical axis compared those at an angle. The result is a blurring of a point in the image into a disc. The effect can be reduced by under-focusing the image and also by the use of specialized devices called Cs-correctors.

Chromatic Aberration Chromatic aberrations in EM can be observed when an incoherent beam of electrons (hence, electrons that differ slightly in energy (speed)) are focused by the lens in different points. The difference in energy can e.g. occur due to inelastic scattering events or an incoherent electron source. Especially for high resolution imaging it is beneficial that a small energy loss (i. e. due to inelastic scattering events) does not result in an electron being out of focus. Chromatic aberration can be reduced by using dedicated devices like monochromators. However, they are not yet widespread.

Astigmatism Astigmatism is an optical effect that is commonly caused by the electromagnetic lens. If the magnetic field in the lens is not perfectly round, it results in differing focal lengths for (e.g.) perpendicular electron waves. Which means that a circular structure in the image would appear to be elliptic. Low order astigmatism can be corrected fairly easy in the lens by the use of stigmators.

2.4.2 Sample Preparation

Image recording requires preparation steps to enable the biological sample to withstand the conditions in the TEM. That is the high radiation damage induced by the electron beam and the vacuum in the column. There are two approaches used to achieve this. *Negative Staining* amplifies the image contrast while sacrificing high resolution information, whereas the *Cryo Preparation* produces lower contrast, but preserves high resolution information. The former is primarily used for screening the latter for the actual reconstruction work-flow.

Negative Staining Preparation The Negative Staining Preparation was invented in 1959 by Brenner and Horne[11]. This sample preparation technique uses a heavy metal salt (e.g. Uranyl acetate) as a so called *stain*.
During the staining process the specimen is first applied to a carbon coated copper grid and then embedded in the heavy-metal salt which serves as a contrast agent. While recording the images in the TEM the stain amplifies the interaction of the electrons with the sample. This enhances the typically low amplitude contrast. However, the process limits the recording of sample features to low resolution details and is known to distort the molecules due to air-drying. Despite these shortcomings negative staining is often used in the early phase of a project to gain a first impression of the behavior of the sample and to assess its quality.

Cryo Preparation In contrast, the cryo preparation process deals with the vacuum and high radiation environment in the TEM by embedding the specimen into a thin layer of *amorphous*, hence, vitreous, ice. This layer of ice is created by rapidly freezing the specimen in an appropriate coolant like liquid ethane at -180°C. This important step mostly prevents ice crystals from forming, which otherwise would damage the specimen. Furthermore, the non-crystalline structure of the vitreous ice ideally does not substantially contribute information to the recorded images which could distort the output. The process of embedding the sample into ice is called *vitrification* and was first described by Dubochet and McDowall [23, 24].

This process of plunge-freezing mostly keeps the sample intact and does not alter its structure. Hence, it is suitable for high resolution structure determination.

2.4.3 Image Formation

To make features in an image distinguishable, they need to have different intensity values (hence, contrast). The larger the difference between the intensity values in an image are, the higher its contrast is. In EM the contrast is mainly determined by the differences in amplitude and phase shifts of the scattered with respect to the unscattered portion of the imaging electron wave. The following section describes the influence of electron scattering on the formation of contrast in EM images.

As mentioned in Section 2.4.1.1, the sample inside a TEM is illuminated by a coherent beam of incident electrons. This results in scattering effects. *Inelastic scattering* occurs when atoms of the specimen are hit by the incident electrons, which results in an energy transfer to the respective atom of the sample. This introduces several unwanted effects. First, it can lift electrons of the sample atoms into higher energy shells, which results in the emission of x-rays or other electromagnetic radiation, once the created vacancies in lower electron shells are filled again by electrons from higher shells. Other effects are ionization of the sample or secondary electron scattering [26]. These effects deteriorate the sample quality and contribute to noise in EM images. Inelastic scattering often results in high scattering angles of the electrons, such that they can be partially removed by apertures. Due to the energy loss of inelastically scattered electrons, the latter can also effectively be removed by electron optical devices called energy filters [77]. Since biological samples mostly consist of lightweight atoms (e.g. nitrogen, hydrogen, carbon) the amount of inelastic scattering is low. Inelastic scattering is associated with *amplitude contrast*. *Elastic scattering* on the other hand occurs when electrons are deflected by the nuclei in the specimen without loosing much of their energy. This alters their scattering angles and hence, also their path lengths in the optical setup of the microscope. It, thus, induces a phase shift in the interfering phase-shifted and unscattered incident wave. This *phase contrast* is much more relevant for the image formation and therefore described here in detail.

Let Ψ_0 be the incident electron wave. As it hits the sample it becomes altered and carries the information in the form of its 3D *Coloumb potential*:

$$\Phi(r) = \int C(r, z)\, dz \tag{2.45}$$

Where r is a 2D vector of interaction positions and z is the thickness of the sample. Now that the incident wave is scattered by the specimen it transforms into an exit wave Ψ_{ex}:

$$\Psi_{ex}(r) = \Psi_0 \ \cdot \ \exp(i\sigma\Phi(r)), \text{ with } \sigma = \frac{m_c\lambda}{2\pi h^2} \tag{2.46}$$

Since σ is a constant factor its omitted in the following for the ease of readability.

As mentioned before, biological samples mostly contain lightweight atoms therefore Equation 2.46 can be simplified using the so called *weak phase approximation* ($\Phi(r) << 1$) [26]. First the exponential is expressed by a taylor series:

$$e^x = \sum_0^\infty \frac{x^n}{n!} \tag{2.47}$$

which leads to:

$$\Psi(r)_{ex} = \psi_0 \ \cdot \ \sum_{n=0}^{\infty} \frac{i\phi(r)^n}{n!} \tag{2.48}$$

And then expanded:

$$\Psi(r)_{ex} = \psi_0 \ \cdot \ \left[1 + i\Phi(r) - \frac{1}{2}\Phi(r)^2 + \frac{1}{6}i\Phi(r)^3 - \ldots\right] \tag{2.49}$$

This is then truncated after the first two terms, because of the weak phase assumption. This yields

$$\Psi(r)_{ex} \approx \psi_0 \ \cdot \ [1 + i\Phi(r)] \tag{2.50}$$

In other words, the exit wave $\Psi(r)$ corresponds to the sum of the incoming unscattered wave Ψ_0 and scattered wave Ψ_{ex} phase shifted by $\frac{\pi}{2}$. The intensity distribution can then be expressed by:

$$I(r) \approx \Psi_{ex}(r) \ \cdot \ \overline{\Psi_{ex}(r)} = 1 + \Phi(r)^2 \tag{2.51}$$

With $\overline{\Psi_{ex}(r)}$ being the complex conjugate of $\Psi_{ex}(r)$. Now the main issue here is that since $\Phi(r) << 1$ the contrast can practically not be detected. Therefore another phase shift of $\frac{\pi}{2}$ needs to be introduced. Thus the exit wave becomes:

$$\Psi(r)_{ex} \approx \psi_0 \ \cdot \ [1 - \Phi(r)] \tag{2.52}$$

Which yields the updated intensity distribution:

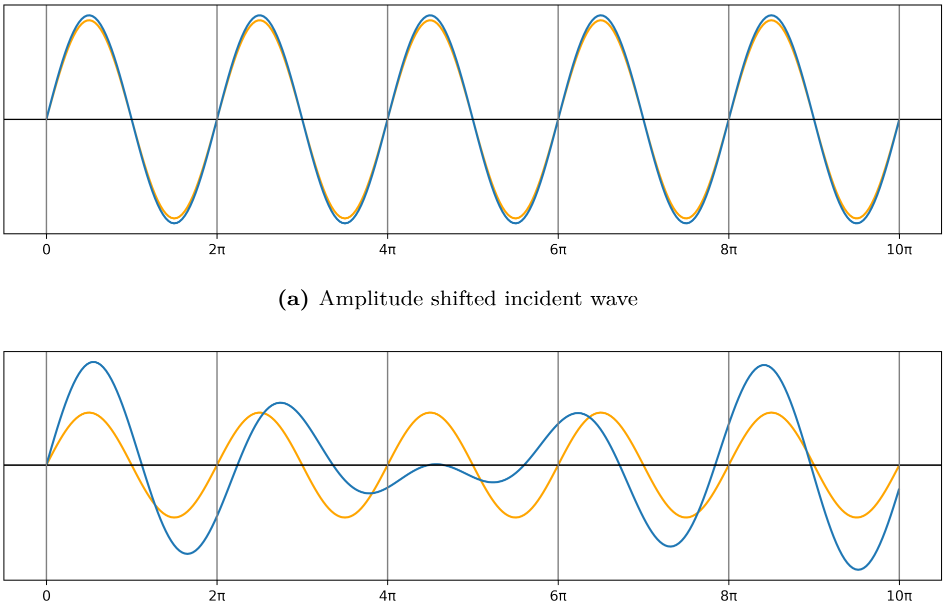
$$I(r) \approx \Psi_{ex}(r) \cdot \overline{\Psi_{ex}(r)} = 1 - 2\Phi(r) \tag{2.53}$$

In contrast to Equation 2.51 this is now a linear dependence on the phase shift.

However, this is just the ideal case which in practice never occurs. Due to aberrations in the microscope (see Section 2.4.1.2) the real intensity distribution is different. One can think of it as a convolution of $\Psi(r)_{ex}$ with the specific Point Spread Function (PSF) of the microscope:

$$\Psi(r)_{real} = \Psi(r)_{ex} * PSF(r) \tag{2.54}$$

The PSF essentially characterizes how much and in which way a single point is distorted by the TEM.



(a) Amplitude shifted incident wave

(b) Phase shifted incident wave

Fig. 2.6: Simulation of how a phase shift can result in detectable amplitude contrast. *Top:* The small shift in amplitude of the incident electron wave (orange) compared to the resulting wave (blue) is barely detectable. *Bottom:* If the wave additionally experiences a phase shift (e.g. by spherical aberrations in the lenses or by defocussing the beam) the amplitude contrast is amplified and becomes visible.

2.4.4 Contrast Transfer Function (CTF) Correction

In Section 2.4.3 a short introduction in image formation in cryoEM was given and the concept of PSF was introduced. The PSF is specific to the used TEM and needs to be corrected in order to achieve high resolutions (see [49, 79]). This is usually done in Fourier Space. The fourier transform of the PSF is called Contrast Transfer Function (CTF). Given that convolution in real space becomes multiplication in fourier space Equation 2.54 turns into:

$$\mathcal{F}(\Psi(r)) = \mathcal{F}(\Psi(r)) \cdot CTF(r) \cdot E \tag{2.55}$$

Note that E is an additional term introduced by Trueblood et al. [68]. It is an envelope function that accounts for the decay of signal in higher spatial frequencies, due to optical aberrations, incoherence of the beam and other limiting factors [25]. Given a scattering angle θ, E is defined by:

$$E(\theta) = e^{-2B\theta^2} \tag{2.56}$$

With B being the so called *experimental B-factor*.
The CTF itself can be approximated by a wave aberration function $W(\theta)$ by the Scherzer formula [61]:

$$W(\theta) = \frac{\pi}{2\lambda}\left(C_s\theta^4 - 2\Delta z\lambda\theta^2\right) \tag{2.57}$$

Where θ is the scattering angle, λ is the wavelength (see Equation 2.44), C_s is the spherical aberration (see Section 2.4.1.2) and Δz the current defocus. Applying the sine yields:

$$CTF(\theta) = \sin(W(\theta)) = \sin\left[\frac{\pi}{2\lambda}\left(C_s\theta^4 - 2\Delta z\lambda\theta^2\right)\right] \tag{2.58}$$

This fully describes the observed wave as a product of the fourier transformed emerging wave, the CTF and a dampening function E as seen in Equation 2.55.

2.4.5 Image Processing

The former section described in detail how an actual image is formed in the TEM. This section is supposed to give insight into the fundamental image processing work-flow to create a final 3D (resp. a set of final 3Ds) from these images.

Figure 2.7 illustrates in a simplified rendering how the sample embedded in a thin layer of ice could look like. The sample molecules (here Fatty Acid Synthase (FAS)) are randomly oriented in space (orange) and can have different conformations (blue). They are embedded in a thin layer of ice (light blue) and illuminated by a coherent beam of electrons (yellow), creating noisy projection images (depicted as shadows) in the imaging plane. Images containing projections of particles are called *micrographs*. Particles that have the same orientation but different conformations can result in very similar projection images on the micrograph (dark blue FAS molecule and the orange one close-by).

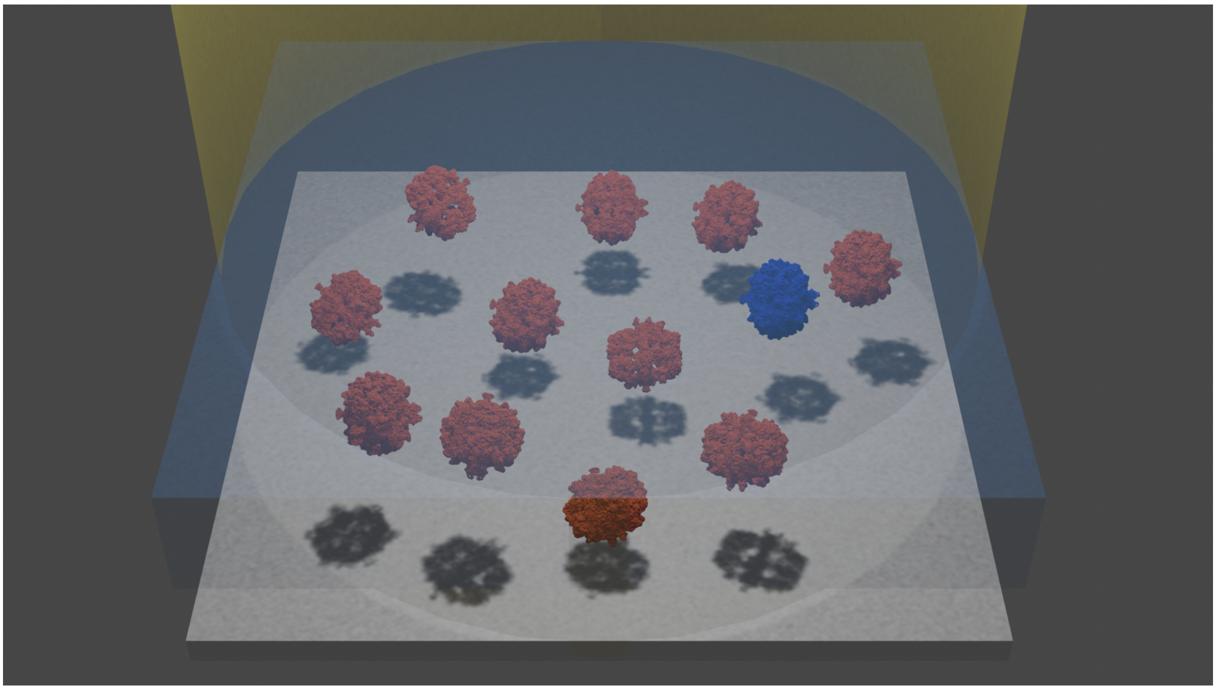


Fig. 2.7: The rendering illustrates imaging in cryoEM

2.4.5.1 Particle Picking

The first crucial step in cryoEM image processing is the extraction of the 2D images from the micrographs recorded in the TEM. The amount of particles depicted per micrograph depends on the quality of the sample and the size of the macromolecule but is usually around hundred or even less. To reconstruct high-resolution structures, a vast amount of projection images is required, usually in the range of several hundred thousands to even millions. Therefore, it is impractical to perform this task manually and a lot of tools were published to tackle it.

In general, there are two approaches used with their respective advantages and disadvantages. The *based* approach uses a (typically low-pass filtered) reference structure that is projected into 2D space. The projection angles are sampled from a regular grid in a pre-defined angular distance. The 2D reference images are then compared to patches of the micrograph. Various metrics have been established to reliably detect particles, of which the most common is the CCC (see Section 2.3.1). Although being very quick and reliable, this approach has the main issue that it is prone to over-fitting. This mainly happens when random noise patches coincidentally resemble actual particles. The closer the images are recorded to focus (that means the lower the SNR(Section 2.3.1)) the more affected the picking tools are.

The second approach is the so called *reference free* approach. These tools use basic image statistics to match particles in the micrograph. These approaches are often less reliable, but also less prone to over-fitting.

In recent times also machine learning has been established as a tool to find particles in micrographs (e.g. [52, 75, 76]). These tools are often based on convolutional neural networks and feature denoising capabilities.

All of the mentioned tools are strongly influenced by the amount of noise present in the micrographs, but also contaminations like ethane crystals, particle agglomerations, inhomogeneous carbon foil, varying ice thickness, ice crystals and so on. This makes particle picking a very challenging task. The consequences can be that images extracted from the micrograph may not contain particles, contain defective particles, have a strong bias towards certain views of the particle, or contain artifacts (sometimes even with characteristics that resemble actual particles). This of course has a very strong influence on the speed of the reconstruction and moreover: its quality. An example micrograph from the FAS dataset is shown in Figure 2.8. The green circles illustrate picked particles selected by a reference free automated picking algorithm (see [15]) in the *Cow Suite*.

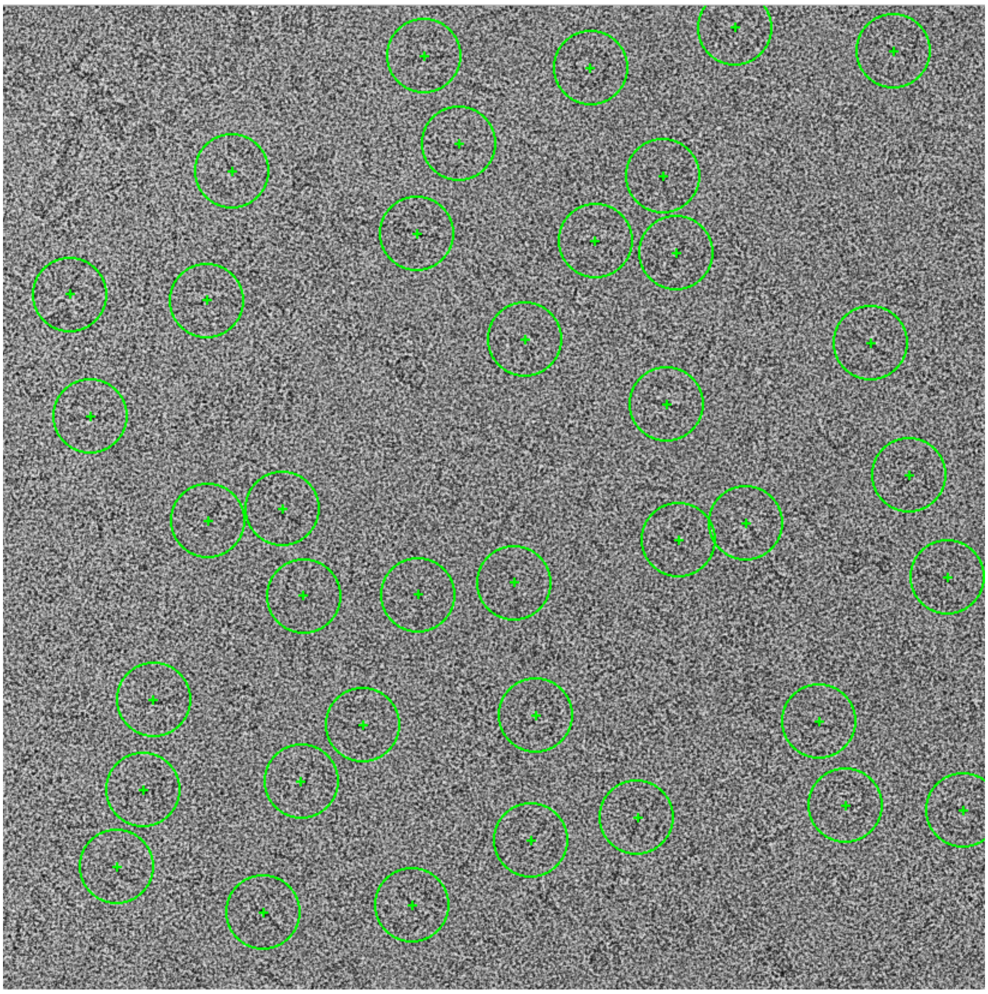


Fig. 2.8: Illustration of a micrograph with pre-picked particles (green circles). Note that not all particles in the micrograph were found.

2.4.5.2 Preprocessing

The next step prepares the stack of picked particles for the reconstruction. Based on the micrograph that the respective particles were picked from they are now CTF corrected (see Section 2.4.4). Since micrographs differ in the distribution of their intensity values (e.g. influenced by the thickness of the ice, characteristics of the beam, etc.) they need to be *normalized* (see Section 2.3.2). This means that all the particle images are modified in a way that their intensity distributions are equalized. To improve the performance of the subsequent algorithms particles are then also filtered (see Figure 2.3) to increase SNR and masked to decrease the importance of pixels that do not belong to the particles themselves.

2.4.5.3 Alignment

As illustrated in Figure 2.7, the imaged particles are randomly oriented in space and also suffer from a low SNR. The main goal is therefore to first align them in a way that similar particles (which means similar views from the corresponding 3D structure) match in terms of their 2D spatial orientation. This involves compensation for three degrees of freedom, namely the shifts in x- and y- direction and also the in-plane rotation ϕ (see Section 2.3.7.1). To perform alignment there are two main approaches the *reference based* and the *reference free.*

Reference free alignment If no starting model exists or the usage wants to be avoided, reference free alignment is performed. This makes use of the so called *rotational average* and thereby shifts the particles towards the center of the box.

$$\sigma_{rot} = \frac{1}{360} \sum_{\phi=1}^{360} R_{\phi} \sigma_X \tag{2.59}$$

With σ_X being the arithmetic mean of the set of particle images X and R being a 2D transformation matrix (see Section 2.3.7.1). σ_{rot} is then used as a reference for the alignment, as described in the next paragraph.

Reference based alignment If a starting model already exists it can be used for alignment. This is done by projecting the reference structure in a defined angular sampling based on a regular grid. A finer grid results in more projection angles and better alignment results, however the complexity increases tremendously since every projection needs to be compared to every particle image in the dataset. The determination of the optimal set of parameters is usually performed in an exhaustive manner.

Let $x_i \in X$ be an image from the stack of particles and $y_j \in Y$ the j-th reference then the set of alignment parameters is determined by the optimization problem:

$$\min_d \Big(d(Tx_i, y_j) \Big) \quad \forall T(\phi, s_x, s_y), i, j \tag{2.60}$$

Where T is:

$$T(\phi, s_x, s_y) = \begin{bmatrix} \cos\phi & -\sin\phi & s_x \\ \sin\phi & \cos\phi & s_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.61}$$

The distance d is often set as the euclidean distance or the normalized cross correlation (see Section 2.3.1).

2.4.5.4 Classification

Once the best fit for the translational and rotational parameters is found, the particles need to be classified and averaged in order to increase the SNR. This can, for example, be done with PCA based classification (see Section 2.3.8). Here the dataset is transformed from a pixel space into a space of independent covariances. This space can then be clustered by e.g. *k-means clustering* [47]. The better the particles are already aligned, the better the classification is. It is therefore crucial to iteratively improve the alignment based on the classification and vice versa to make classification as independent as possible from the actual orientation of the particle. After classification the next step is to average

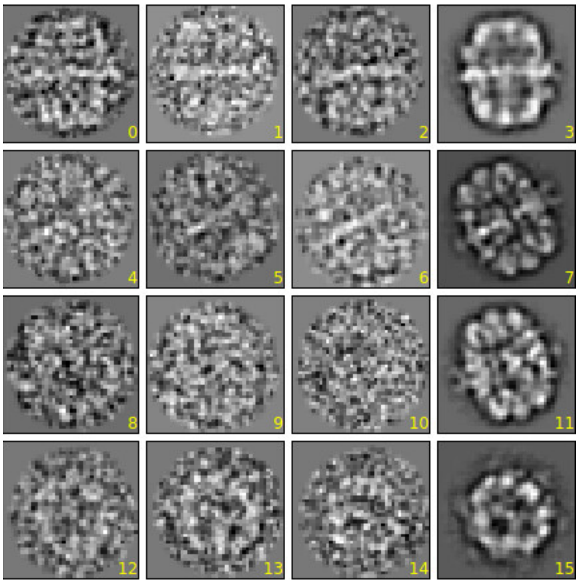


Fig. 2.9: Illustration of the classification process. The four rows show three members of arbitrary classes and their class sum (last column). All four classes show distinct features and orientations. The class sums have an increased SNR

all members of the respective classes. This is done in order to suppress the amount of noise in the images. The resulting average particle images are called *class sums*. Figure 2.9 illustrates the effect of classification and averaging.

Class sums can either be used in a next iteration of alignment/classification or if they are already well enough defined they can be used for 3D reconstruction.

2.4.5.5 Angular Assignment

Multiple approaches exist to assign angular information to the 2D projection images. This section describes the so called *angular reconstitution* by Van Heel [72] and *projection matching*.

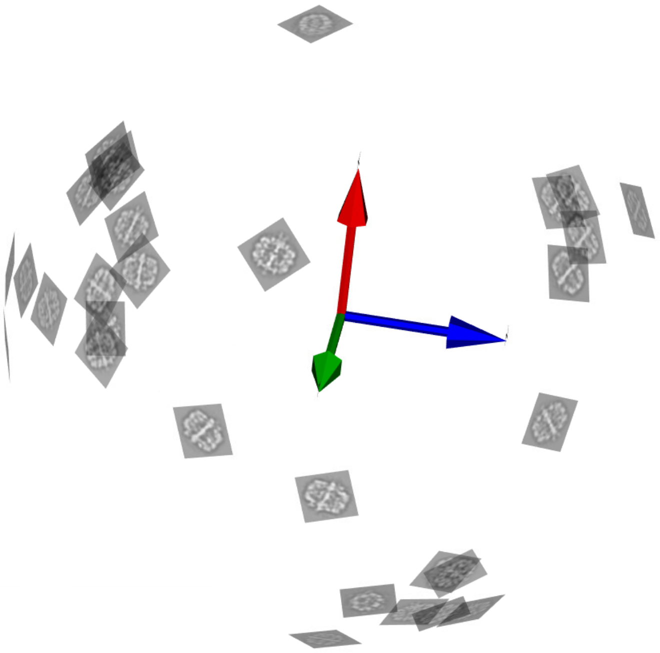


Fig. 2.10: Illustration of the class sums arranged on the Euler sphere based on their individual set of Euler angles. Some directions lack corresponding classes which in result would lead to a smeared out 3D structure. This can be improved by better alignment and classification. However it is also influenced by the imbalanced amount of particles (preferred views) in the dataset. The goal is to cover the Euler sphere as evenly as possible.

Angular Reconstitution The concept is based on the *Central Slice theorem* [10] (see Section 2.3.4) with the idea that every Fourier transformed projections of the same volume are slices in the Fourier transform of this volume and share a common line [28]. It can be used to create structures *de novo*, that is if no knowledge about the molecule exists.

The goal of angular reconstitution is now to determine the common lines between the set of particles. In general two particles are not enough to fully describe the orientation of the volume in space. At least three projections are needed. Figure 2.10 illustrates the result of an angular reconstitution. All projections are sorted onto the outer shell of a 2D sphere which describes the set of all possible Euler angles.

In practice the determination of the common lines between a pair of images can also be done in real space by calculating the so called *Radon transform* [55]. Radon stated that in an Euclidean space that is n-dimensional any real function can be expressed by its integral over all of the $(n-1)$-dimensional hyperplanes. This means given all its 1D projections a 2D function can be described and given all the 2D projections the corresponding 3D volume can be determined. With this in mind the 1D radon transformation of every 2D image can be calculated. That is the set of all 1D projections for every projection angle ϕ. This is called a *sinogram*. The common line between two images can then be calculated by e.g. calculating the cross correlation of each set of sinograms. This can be applied in an analog manner in 3D space.

Projection Matching Projection matching can be used if a starting structure already exists. The starting structure is projected with a defined (usually regular) angular sampling and the particle images can be sorted based on their similarity to them. Since the projection angle is known from the starting structure it can be assigned to the matching particle image.

2.4.5.6 Reconstruction

Using the angular information tied to every class sum or projection one can now finally reconstruct a 3D structure. Again there are multiple ways to achieve it with the easiest one being the so called *filtered back-projection*. This is done by smearing out every projection image in the direction perpendicular to its image plane. This results in a structure which is very blurry and therefore needs to be filtered. For filtering usually a *ramp* function is used.

Other, more sophisticated approaches are e.g. Simultaneous Iterative Reconstruction Technique (SIRT), Scale Invariant Feature Transform (SIFT) [4] and Speeded Up Robust Features (SURF) [6]. While often producing better results they are much slower. Especially in early iterations of the refinement cycle it can therefore be reasonable to sacrifice some of the quality for an increase in speed.

2.4.5.7 Iterative Refinement

The structure that is reconstructed in the previous steps can be used as a reference for the next cycle. That means it will be projected in a constant angular distance and the resulting images are used either for alignment or for projection matching. This iteratively improves 2D classes and therefore also the resulting 3D density map.

RELION additionally features the so called *multi-body-refinement*. In this process an arbitrary, but constant and discrete amount of volumes is randomly generated and used as references for refinement. It thereby tries to account for conformational heterogeneity.

2.4.6 Quality Assessment

The assessment of the reconstruction quality is an essential step in cryoEM. Multiple ways exist to evaluate the quality of a given map. The most common measure is the resolution. It is usually expressed by using the Fourier shell correlation (FSC), which measures the consistency of two (half) maps. It is crucial to understand that an absolute measure for resolution in cryoEM does not exist.

2.4.6.1 Resolution

The maximum obtainable resolution for cryoEM density maps is determined by the *Nyquist-Shannon-sampling-theorem* [64].

$$f = \frac{1}{2s} \tag{2.62}$$

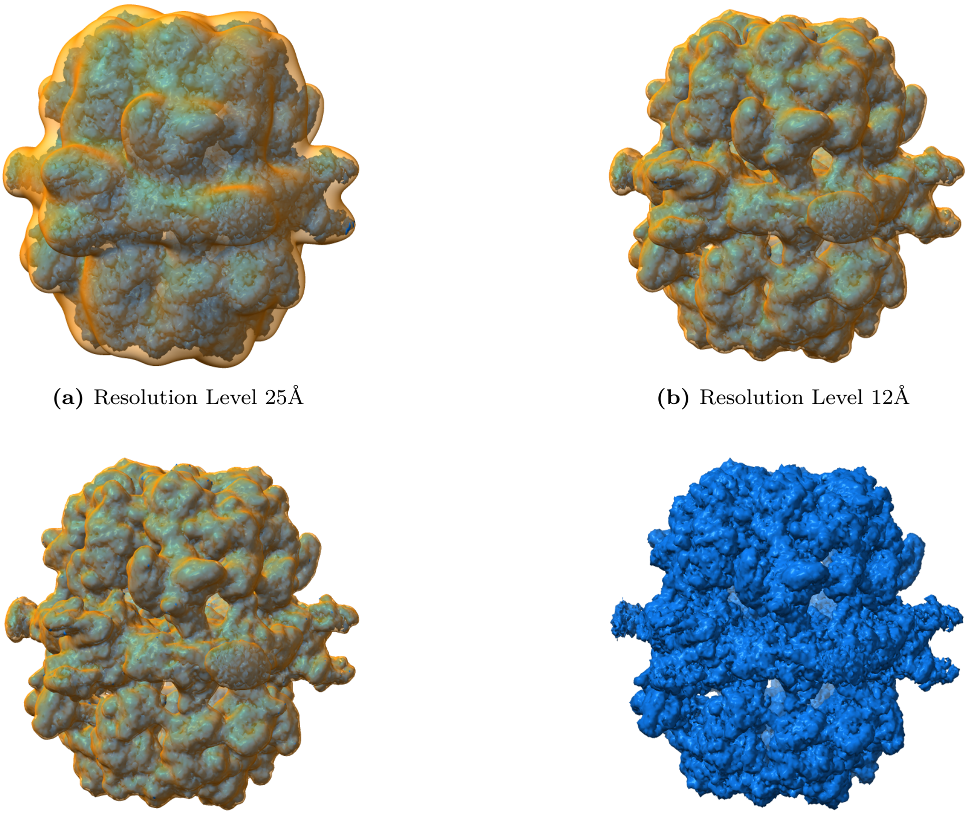
With f being the Nyquist frequency and s being the pixel size. That means for the maximum obtainable resolution d_{max} of the molecule:

$$d_{max} = \frac{1}{f} \tag{2.63}$$

In other words: The maximum obtainable resolution is twice the pixel size. This intuitively makes sense, because at least two pixels are needed to distinguish a change within intensity values. All filters that alter the pixel size (e.g. coarsening) also alter the maximum obtainable resolution. Figure 2.11 shows a comparison of different resolution levels for the FAS molecule.

2.4.6.2 FSC

The most widely used measure for the quality of reconstructed 3D maps in cryoEM is the so called FSC [70]. It is the 3D extension of the 2D Fourier Ring Correlation (FRC) first described by Heel et al.[31]. It evaluates the normalized cross-correlation (see Section 2.3.1) between three-dimensional shells in Fourier space (see Section 2.3.3) of two



(a) Resolution Level 25Å

(b) Resolution Level 12Å

(c) Resolution Level 9Å

(d) Resolution Level 2.8Å

Fig. 2.11: Different resolution levels (orange) compared to the 2.8ÅFAS map (EMD-4577 in EMDB, blue).

volumes.

Usually this is combined with the so called *gold standard* [60, 71] refinement. That is splitting the dataset in two half sets and processing them independently. Thereby, the risk of over-fitting (see Section 1.3.2.1) is reduced. However, FSC is not an absolute measure of the resolution of two volumes, but rather of how consistent the two of them are with respect to their spatial features.

Let X_1 and X_2 be two half maps and $\tilde{X_1}, \tilde{X_2}$ their respective Fourier transforms then:

$$FSC(X_1, X_2, r) = \frac{\sum_{r_i \in r} \tilde{X_1}(r_i) \cdot \tilde{X_2}(r_i)^*}{\sqrt{\sum_{r_i \in r} \left|\tilde{X_1}(r_i)\right|^2 \cdot \sum_{r_i \in r} \left|\tilde{X_2}(r_i)\right|^2}} \tag{2.64}$$

Here, $\tilde{X_2}^*$ denotes the complex conjugate of $\tilde{X_2}$. The FSC essentially measures to which spatial frequency both volumes match. Various thresholds have been applied to this curve and are discussed controversially. The most common threshold values are the 0.5 and the 0.143 criteria. A comparison of different thresholds can be found in [73]. The 0.5 criterion was suggested by [9]. The less conservative 0.143 criterion was first described by Rosenthal and Henderson[57]. It is often used for structures refined using the *gold standard*.

Figure 2.12 shows an example of an FSC curve. The 0.5 and the 0.143 thresholds are marked at the y-Axis.

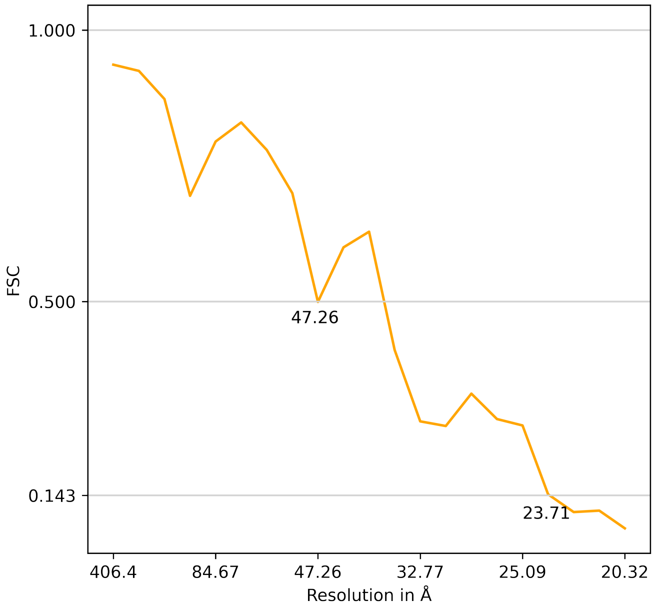


Fig. 2.12: Sample FSC curve. The FSC values for the 0.5 and the 0.143 criteria are marked in the plot.

2.5 Machine Learning

2.5.1 Artificial Neurons

Artificial neurons are entities largely inspired by their biological counterparts (see Figure 2.13). They are described by an input vector x, the corresponding weight vector w, the bias term b, an activation function f and produce an output vector o. The output is calculated such that:

$$o = f(wx + b) = f\left(\sum_{(i)} w_i x_i + b\right) \tag{2.65}$$

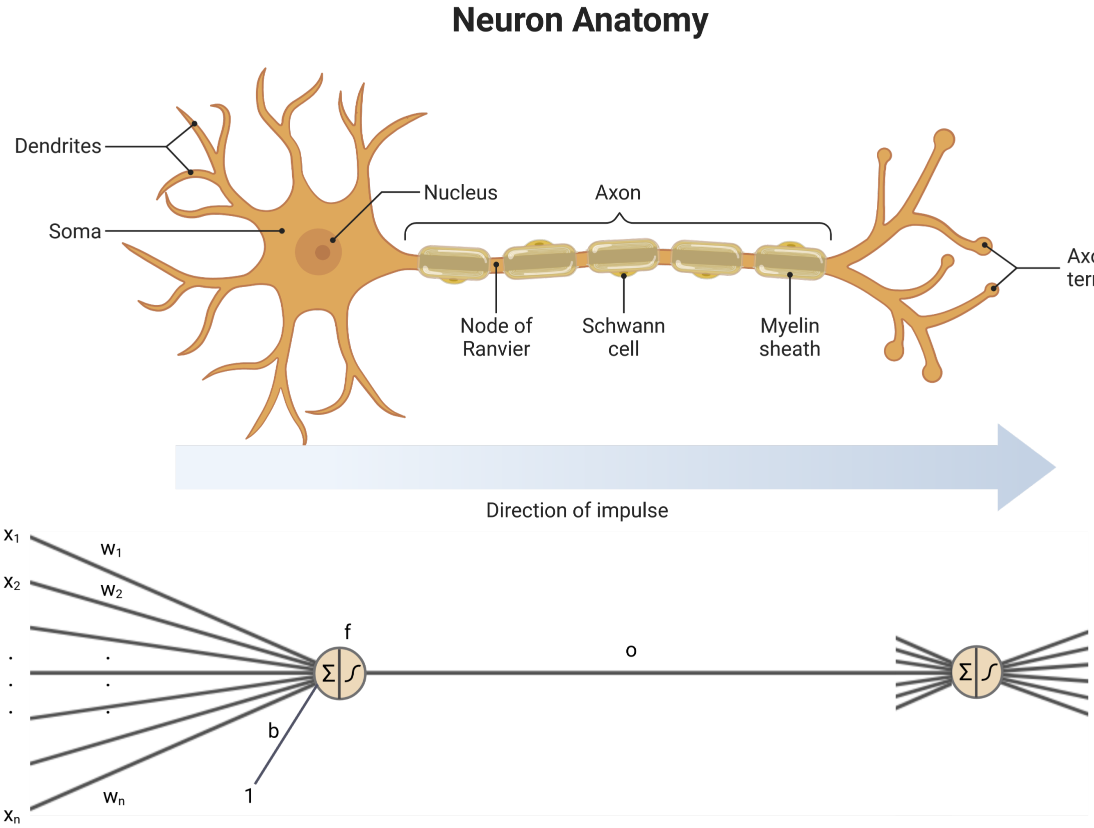


Fig. 2.13: Artificial neuron vs. neuron. The illustration of the neuron was adapted from BioRender (2021). Neuron Anatomy. Retrieved from https://app.biorender.com/biorender-templates/t-5f5b7e6139954000b2bde860-neuron-anatomy

As can be seen in Equation 2.65 the neuron gathers information from its input channels, weights them and calculates the output accordingly, based on the activation function. While w modifies the shape of the output function, b is independent of the actual input and just shifts the curve. This allows for more complex responses of the neuron.

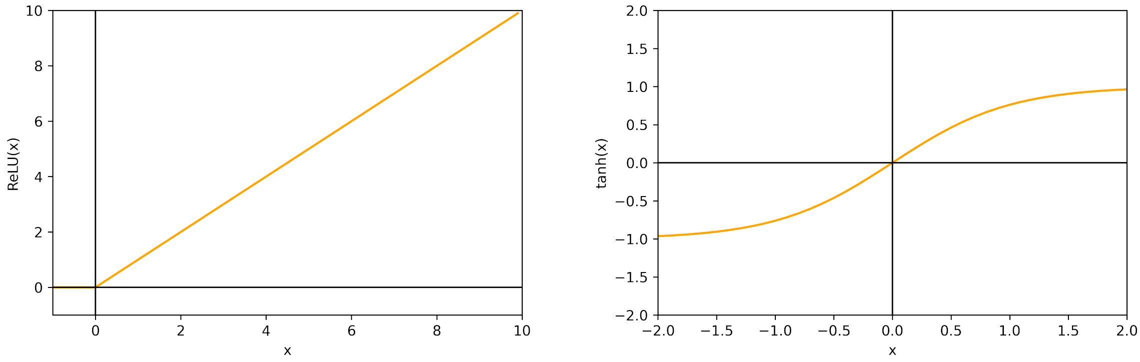
A wide variety of functions f are used as an activation function. The functions used in this thesis are described below.

Rectified Linear Unit (ReLU) The Rectified Linear Unit (ReLU) activation was introduced by Nair and Hinton[51]:

$$f(x) = x^+ = \max(0, x) \tag{2.66}$$

Hyperbolic Tangent (tanh)

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.67}$$



(a) ReLU activation function

(b) tanh activation function

Fig. 2.14: Comparison of different activation functions.

2.5.2 Artificial Neural Networks (ANNs)

As sketched in Section 1.4 ANNs are the basis of todays deep learning algorithms. ANNs group artificial neurons in networks. Figure 2.15 illustrates a so called *fully connected* ANN. In this network, neurons are organised in *layers*. The first and the last layers are typically called *input* and *output* layer, the layers in-between cannot be accessed from outside the network and are therefore called *hidden* layers. The amount of neurons per layer and also the amount of layers in the network can vary. The input layer corresponds in size to the number of *features* that the ANN has access to. In image processing this is usually one neuron per pixel. The number of output neurons depends strongly on the respective task. For *classification* problems the amount can be low (e.g. number of classes), for *regression* problems it can be fairly large (e.g. number of pixels in an inferred image). In general layers with fewer neurons than the previous layer are condensing

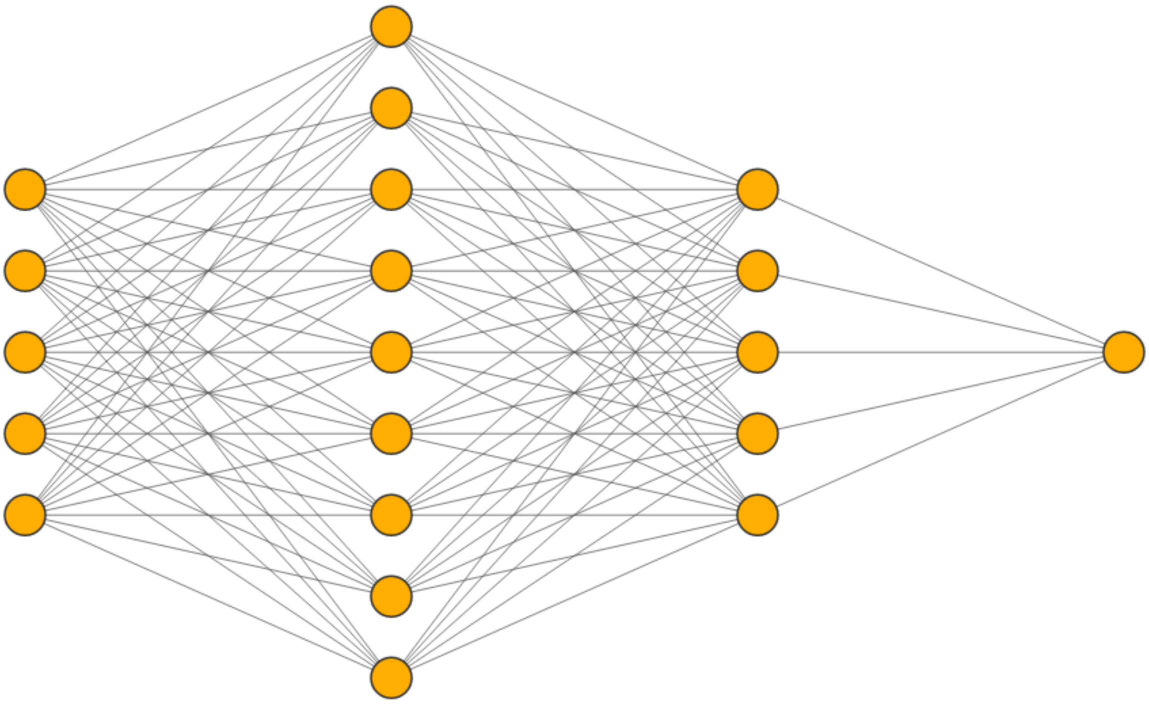


Fig. 2.15: Illustration of a fully connected ANN with an input layer, two hidden layers and one output layer.

information. Layers with more neurons than in the previous layer allow us to inflate previously condensed information.

As the information flows from the input to the output layer of the network it gets weighted. The more layers the ANN has, the more abstract the internal representation of the data becomes. In abstract terms the ANN can be expressed as a composition of functions:

$$\begin{aligned} o(x) &= h_n(\dots h_2(h_1(h_0(x)))) \\ &= f_n(w_n \dots (f_2(w_2 f_1(w_1 f_0(w_0 x + b_0) + b_1) + b_2) \dots) + b_n) \end{aligned} \tag{2.68}$$

With n being the number of layers, h_i being the i-th hidden layer and the other variables as defined previously.

The weights in the ANN now need to be adapted such that for an input x the output o matches the expected result. This is achieved in a process called *training*.

2.5.3 Training Neural Networks

To understand the training process of an ANN it is crucial to understand the training of a single neuron. At the beginning of the training the weights w of the neuron are initialized randomly. In the so called *forward pass* the output of the neuron is calculated by applying Equation 2.65. Which results in an output vector o.

In a supervised setup the neuron has access to an expected output value called *teacher* signal. The deviation of the output from the teacher can be expressed by an arbitrary distance metric, e.g. the difference, the sum of squared errors (SSE), the mean squared error (MSE) and so on. This distance metric is often referred to as *cost metric* or *loss*. In the second step, the so called *backward pass*, an *optimizer* tries to minimize the distance $\mathcal{L}$ between the predicted output vector o of the network and the teacher signal t. E.g.:

$$\min \mathcal{L}_{SSE}(o,t) = \min SSE(o,t) = \min \frac{1}{2}(t-o)^2 \tag{2.69}$$

In simple setups like the single neuron this problem can be solved analytically. In practice one tries to find the optimal values numerically, e.g. by *gradient descent*. The solution can then be approximated by iteratively adapting the weights:

$$w^{(s+1)} = w^{(s)} + \Delta w \tag{2.70}$$

with $w^{(s)}$ being the weights at the current time step s and

$$\Delta w = -\eta \nabla \mathcal{L}, \tag{2.71}$$

with:

$$\Delta w_i = -\eta \frac{\partial \mathcal{L}}{\partial w_i}. \tag{2.72}$$

The parameter η is typically called *learning rate*. Intuitively this means that since the gradient points towards the direction of the steepest ascent, a step of size η is taken into the opposite direction in order to minimize the loss.

In an analog manner, training of networks with multiple neurons and layers can be achieved. Basically, the contribution of every neuron to the total loss is evaluated from the output layer into the direction of the input layer and the weights are adapted accordingly. A very detailed analysis of this process can be found in [8].

2.5.4 Autoencoder

Autoencoder are a type of ANN which is capable of compressing data, while trying to preserve the meaning behind it [44]. From the compressed data (code) it tries to reconstruct the input data as well as possible. The basic structure of an autoencoder is depicted in Figure 2.16. Autoencoder with a single hidden layer, linear activation functions and a squared error cost function span the same space as PCA (see Section 2.3.8) does [54].

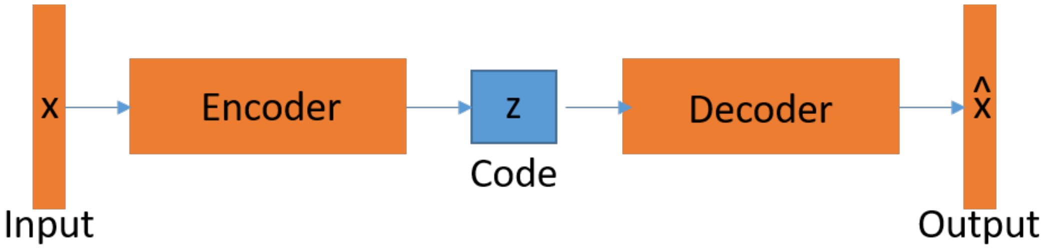


Fig. 2.16: Autoencoder overview.

Autoencoder can be used in combination with other types of ANNs, e.g. as preprocessing step, to reduce noise. The main disadvantages of autoencoders, however, are that there is no way to influence how the data is encoded and that they cannot be used to generate data.

2.5.5 Variational Autoencoder (VAE)

VAE resemble vanilla autoencoders in their basic principle. However, VAE are more powerful in many aspects. They were first described by Kingma and Welling [39]. The basic principle behind VAEs is depicted in Figure 2.17. The VAE consists of two separate ANN that are trained simultaneously. The input space features a complicated distribution, which can only partly be observed in form of the dataset. The latent space, however, can have a simple distribution. The generative model learns a joint distribution $p_\theta(x, z)$ which can be expressed by $p_\theta(z)p_\theta(x|z)$ with $p_\theta(z)$ being the prior over the latent space and $p_\theta(x|z)$ the stochastic decoder network. Task of the stochastic encoder $q_\Phi(z|x)$ is to approximate the intractable true posterior $p_\theta(z|x)$ [40].

In practice often Gaussian prior/posterior pairs are used. To deal with the problem that error backpropagation through stochastic nodes in ANNs is impossible, the so called *reparametrization trick* [41] is used. This trick follows the idea that instead of sampling a (e.g.) Gaussian distribution it can be parametrized deterministically in terms of its mean and standard deviation (which then can be learned by the ANN). The stochastic aspect is then introduced by multiplying a random sample from a standard Gaussian distribution to it. This avoids backpropagation through stochastic nodes.

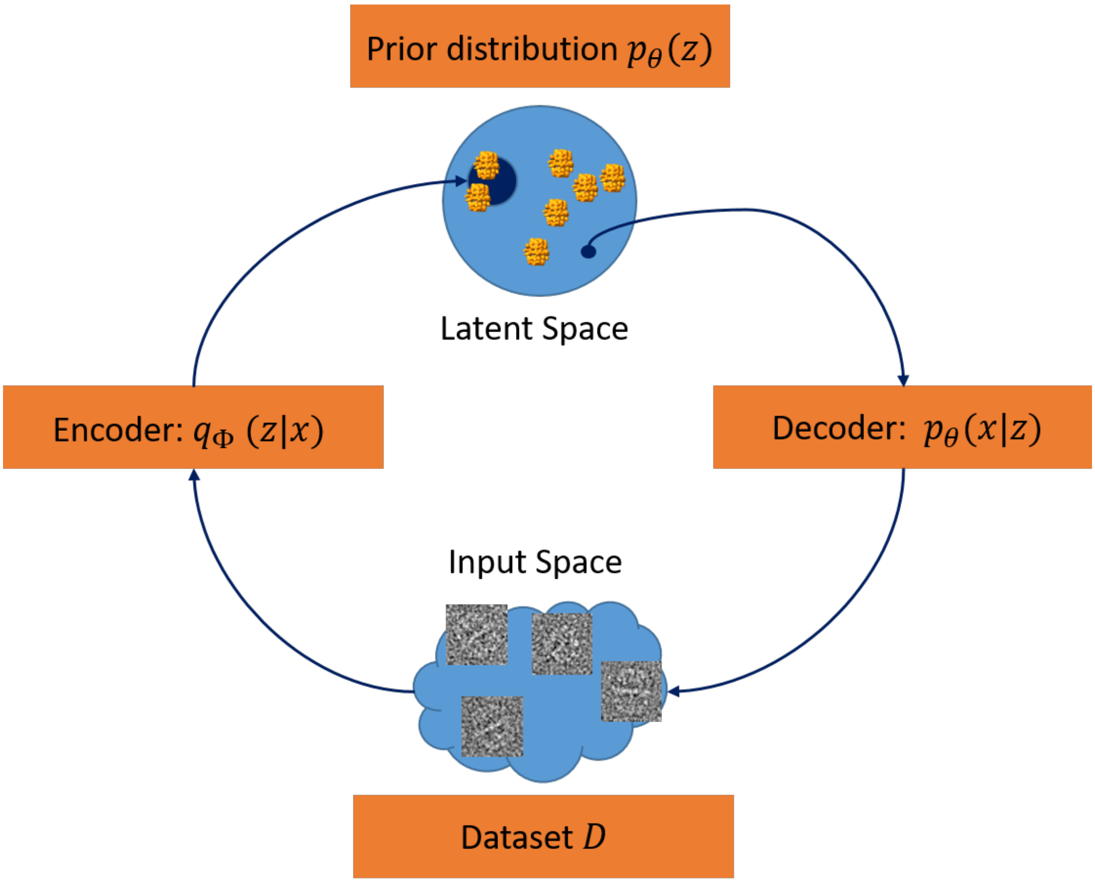


Fig. 2.17: Overview of the main components of a VAE

3 Results

Microscopy and machine learning
Linking to still the little systems
Slow and strong the earth is returning,
Flying to the shore of white silence.
– **Google Verse by Verse AI**[1] [1]

In this chapter a novel software tool is introduced. Its main purpose is to process a stack of 2D cryoEM density maps with minimal user input in a way that it can approximate a continuous space of 3D conformations of the studied biological macromolecule in contrast to many state of the art methods that assume a discrete conformational space. The algorithm does not make any assumptions on the nature, shape, symmetry, distribution and motion of the molecule and therefore minimizes bias. It shows robustness against structural heterogeneity (see Section 1.3.3.2) and can thereby hint towards broken particles. The software is designed to run on graphics processing units (GPUs) in a HPC environment, but can also be used on single workstations. In theory it could be extended to run on Central Processing Units (CPUs). However, due to the limitations in computation speed it seems not reasonable at this point.

3.1 Setup

To enable the software to run in an HPC environment it was integrated into a *Singularity* container. Singularity [45] is a platform, which allows the user to spawn container which create an encapsulated, standardized environment. This ensures reproducibility and reduces side effects due to incompatible software.

The algorithm itself is implemented in python and based on the *Tensorflow* platform by Google [2]. Tensorflow provides an Application Programming Interface (API) for many machine learning related tasks, especially the design and training of ANNs. The Tensorflow API is written in Python, which makes it convenient to use, the computational core, however, is written in C++, which makes it efficient.

[1] This poem was created by a generative ANN. First line was used as a seed. Chosen seed authors were Edgar Allan Poe, Walt Whitman, James Russell Lowell with a little help in assembling from the author of this thesis.

The distribution of the computational load onto different GPUs and computers (nodes) is done via the *Horovod* [63] library, which uses Message Passing Interface (MPI).

3.2 Preprocessing of the Input Data

The algorithm expects the data as a .mrc stack [19] and to be normalized with the tanh estimator (see Equation 2.12), such that the range of intensity values is in the interval $[0, 1]$. During various experiments it has been shown that it is beneficial to also apply a circular mask to the data after normalization, with e.g. 95% of the diameter of the box. However, this is not required. Not masking the images, though, can lead to artifacts caused by transformation operations in later steps of the algorithm (e.g. shift/rotation). More manual preprocessing steps are not required.

In the next step the software splits the data into a training and a test set. The latter is used to evaluate the performance during runtime. The data is then randomized and split into chunks, so called mini-batches. As the algorithm performs calculations on a per-batch level, it is crucial that the data gets shuffled further during the training process, to prevent images from ending up in the same batch again.

3.3 Design of the Algorithm

3.3.1 Overview of the Architecture

The machine learning algorithm is based on a VAE architecture [40] (see Section 2.5.5). VAE are trained in a self-supervised manner, which means that they evaluate the accuracy of their reconstruction based on a similarity metric between input and output data. Input and output usually are of the same dimension.

The VAE presented here features an asymmetric pair of encoder and decoder. The general structure of the ANN is depicted in Figure 3.1. The key concept is the idea that even though the VAE is trained on 2D input images and outputs 2D images (which resemble the input images) it can additionally estimate 3D structures as an intermediate step, which in the end, can be used as the desired output. For an image of size $n \times n$ the decoder therefore outputs $n \times n \times n$ values. To again match the dimension of the input shape, the output of the decoder needs to be reduced in size. This is done by projection, i.e. calculating the sum along the z-axis (see Section 2.3.7.3) of the neurons in the 3D grid. The projection operation itself is a deterministic operation and not subject to the training of the ANN. The growth and shrinkage in dimension is only beneficial if the grid can be transformed in terms of its spatial orientation. The grid transformation follows the concept of spatial transformer networks (see [66]). The set of transformation parameters (pose) Φ can then be inferred by the ANN based on the observed input 2D image. This (ideally) makes the space of 3D structures represented by the latent space of the neural network invariant with respect to the pose Φ. This disentanglement of pose and structure enables the ANN to learn a meaningful representation of the data, based on conformational heterogeneity of the macromolecular complex rather than its orientation in space.

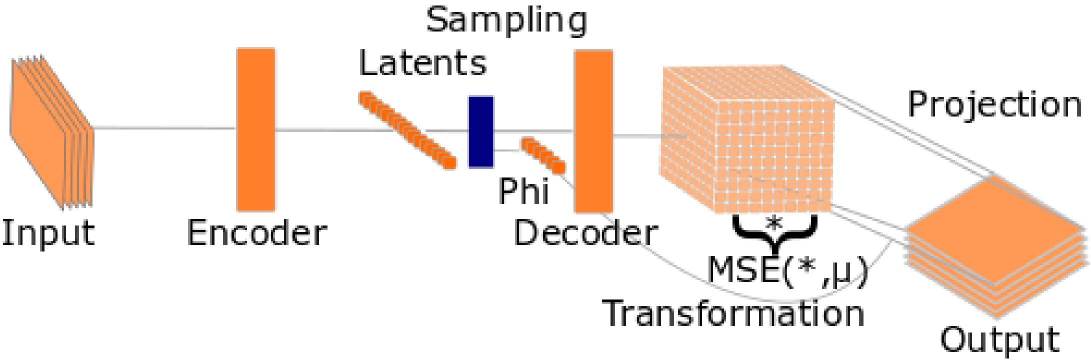


Fig. 3.1: Simplified structure of the ANN.

Encoder The input size of the decoder is $n \times n$ where n is length and width of cryoEM density maps. The encoder network is a fully connected ANN. More sophisticated structures were studied during the process of designing, including convolutional neural networks, resnets, densenets. However, the vanilla architecture has shown the best results. Especially convolutional layers often resulted in the convergence of the ANN into local optimization minima and heavy over-fitting.

The encoder was therefore assembled from 10 fully connected layers of 512 neurons each with the leaky ReLU activation function. The results of the last layer are then fed into two separate layers which parametrize mean and variance of a set of Gaussian functions. Both layers are of the same dimension, namely the dimension of the latent space. Multiple different priors were studied here, e.g. the von Mises-Fisher distribution as suggested by [21]. However, Gaussian priors have shown the best results.

Sampling As mentioned previously VAE use the re-parametrization trick [39, 40] in order to make it possible to perform backpropagation. This is not possible for random nodes, therefore the distributions are parametrized by deterministic nodes multiplied by random samples from Gaussian distributions. That way, the weights of the neurons that parametrize the actual Gaussian distributions can be optimized during the training process, while the stochastic nodes are not trained at all.

Decoder The samples drawn from the Gaussian distributions are then decoded in the decoder network. The decoder network consists of 20 layers with 200 neurons each and ends with another layer which features one neuron per voxel of the 3D structure. The activation function for all layers is again leaky ReLU. After reshaping these n^3 neurons into an $n \times n \times n$ shape it can optionally be masked. This has turned out to be beneficial, because the ANN otherwise starts to put a lot of structural density outside the center of the image, which in cryoEM images is rarely the case.

Pose Estimation As visible in Figure 3.1, there is a second layer in the ANN, which is connected to the latents. This layer handles the estimation of the pose. This is done by using four of the components of the layer to parametrize unit quaternions (see Section 2.3.7.2) and two more for the shifts in x- and y- direction. Multiple different ways to express rotations were tested here, like Euler angles, rotation matrices, axis angles. The only one that has proven to be working was the quaternion representation. The inferior performance of Euler angles might be caused by some of the intrinsic issues that they suffer from and that were discussed before (see Section 2.3.7.2).

To reduce bias induced by irregular sampling of rotation angles it was crucial to use an activation function, which could output reasonable values for the components of the quaternion. Hence, the tanh activation function is used, as it outputs values in the range of $[-1, 1]$.

Transforming the Grid Based on the estimated parameters for the pose, the generated volume from the latent space needs to be transformed, such that it (when projected along the z-direction) matches the input image as close as possible. The volume is (due to the discrete number of neurons) represented as intensity values distributed on a regular 3D grid. This regular voxel grid needs to be transformed into a target voxel grid. Naturally, this requires interpolation, which induces artifacts in the resulting images. Also it has a major impact on the performance of the algorithm, because every point in the grid needs to be rotated in every training step. This could potentially be improved by modeling volumes as point clouds, rather than discrete entities.

3.3.2 Calculating the Loss Function

As mentioned previously, the loss of a VAE is composed of the reconstruction loss between the input image and the output image and the KL divergence between the prior and the aggregated posterior. The loss has a regularizing effect on the ANN. In order to be suitable for cryoEM data, the loss was modified and extended. The reconstruction loss was chosen to be the sigmoid cross-entropy. Hence, the data was normalized using the tanh estimator. The reconstruction loss is calculated between the input image and the projection image of the rotated 3D structure. Furthermore a new loss term was introduced called *averaging loss*. This loss term is a soft regularizer which encourages the network to learn a relationship between the structures in the same batch. This is done by averaging over all the volumes and their reflected versions in the batch and summing up all the deviations from the mean in terms of the mean squared error.

The loss is then weighted based on different weighting factors [34]. It turned out that the averaging loss needed to be weighted 10,000 times higher than the KL divergence. The weighting factors have proven to work for the examples that are shown in this thesis. However, it is not yet clear if there might be a better parameter set.

3.3.3 Hyperparameter

There are a few hyperparameters that can be tuned in the algorithm. This section gives an overview of them.

Learning Rate The learning rate is crucial for the ANN to learn a meaningful representation of the data. If it is too high, then the network is unable to capture fine details. On the other hand, if it is too low, then the training is extremely slow. A value of 10^{-5} worked well in many scenarios.

Dimension of the Latent Space The next parameter is the dimension of the latent space. If the dimension is too low, then the network is unable to capture complex relationships between different samples of the training set. If the dimension is too high, then (usually) the additional dimensions are not used by the VAE. However, the reconstruction accuracy is worse compared to a lower dimension of the latent space. In most cases a dimension around 20 produced good results.

Batch Size As the averaging loss is influenced by the mean and the deviation from the mean, the batch size has a strong influence on the training process. However, in most cases 120 seems to be a reasonable value.

3.3.4 Training the ANN

During training, the software provides the user with the relevant scores for the test set, i.e. the KL divergence, the projection and averaging loss and it outputs sample reconstructions as .png and .mrc files. The user can then decide, based on this output, when to stop the training process.

3.3.5 Inference

After the training of the network has finished it can later on be loaded and used for inference. This is done by either feeding stacks of 2D images into the network or stacks of vectors with a dimension that match the dimension of the latent space. The latter can be used to e.g. interpolate trajectories along the latent space, as will be demonstrated later on.

3.3.6 Performance/Complexity

The speed of the algorithm is mainly determined by the grid transformation operation. Since it is necessary to evaluate the pose for every volume during every training step this is very costly. As the volume scales with the power of 3 even a small increase in box size has a strong influence on the training times. Bigger batch sizes also slow down the training process. Realistic values are between 1 and 2 training steps per second. Which results (depending on the task) in training times of several days till even weeks on one node with 4 GPUs.

3.4 Testing the Algorithm

This section is divided into four main parts. The first part studies synthetic test data, while in the other parts the algorithm is applied to different macromolecular complexes.

3.4.1 Simulated Data

This section provides a proof of concept of the algorithm on simulated test data.

3.4.1.1 Design of the Experiment

For this experiment a simulated dataset was created using the Cow Suite (see Section 2.1). First, a solid L-shaped structure Figure 3.2 was designed in 3D. The box size for this experiment was 40 pixel by 40 pixel. The L-shaped structure was then projected based on a regular HEALpix grid [29] with an angular sampling of 0.92 which resulted in $\approx 50,000$ projection images. A random subset of these projections can be seen in Figure 3.3. To mimic real cryoEM datasets, which (ideally) contain multiple copies of different projection angles, every image was cloned two times. In the next step white Gaussian noise was added to the images such that the resulting SNR resembled a realistic (i.e. in cryoEM standards) level of 0.1. In the last step the resulting dataset of roughly 150.000 images was normalized using tanh estimator with a spread of 0.3.

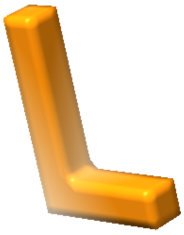


Fig. 3.2: Synthetic L-shaped test volume

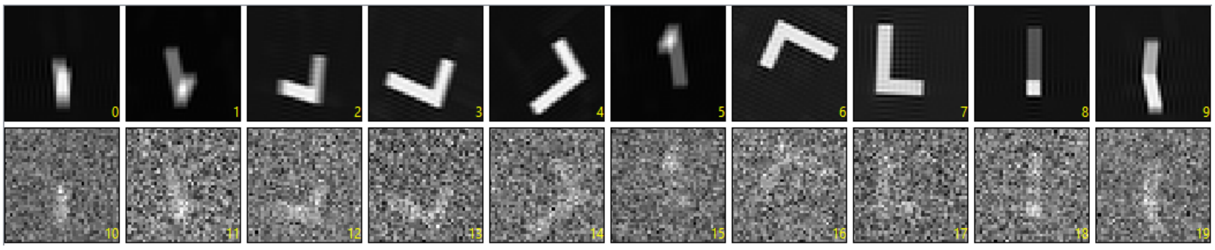
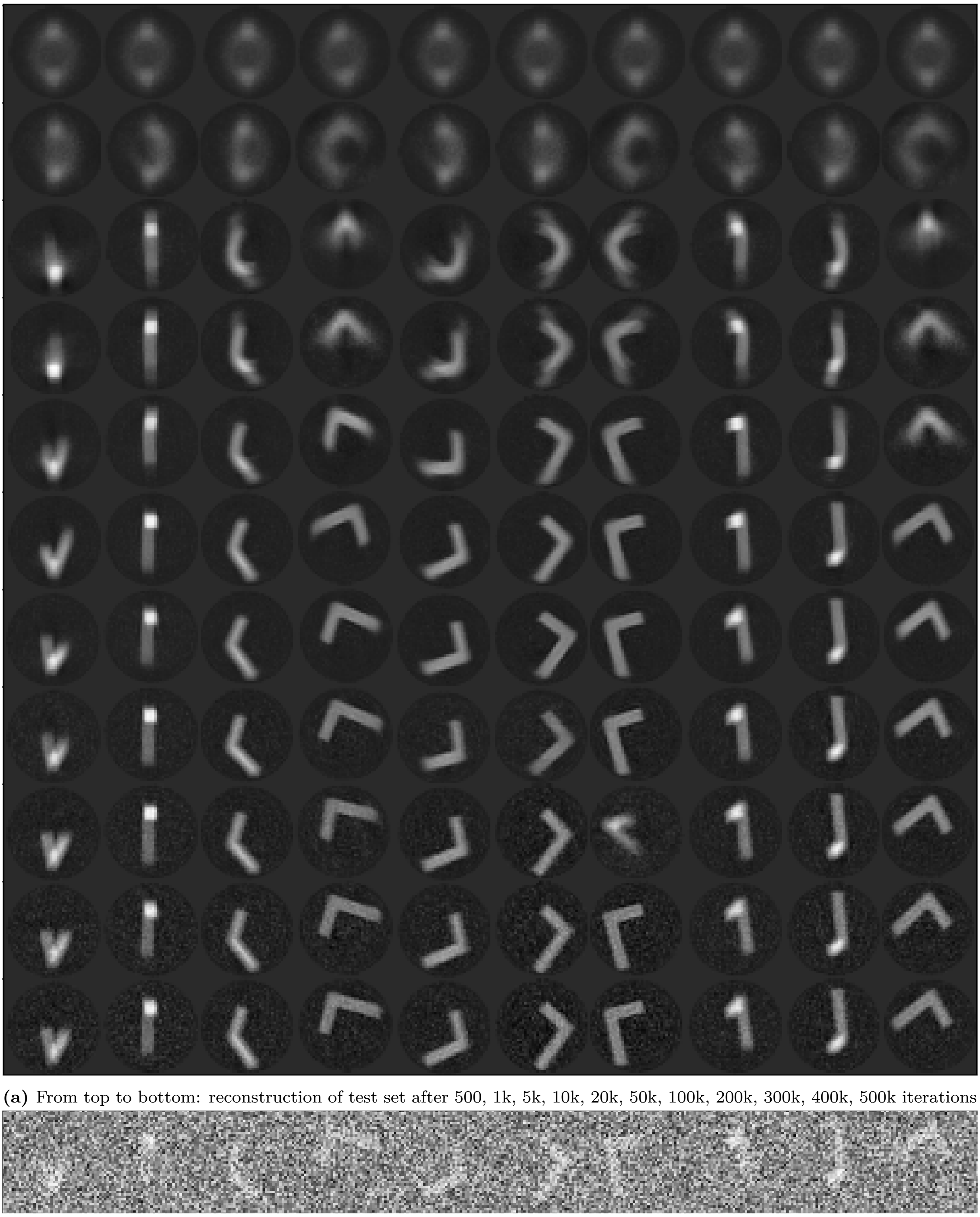


Fig. 3.3: A selection of different projection angles of the simulated test set before and after adding Gaussian noise up to a SNR level of 0.1

The dataset was then fed into the neural network. The chosen hyper parameters were 21 dimensions for the latent space and a mini batch size of 30 images per GPU with a total of 4 GPUs on one of the previously described nodes and a combined total learning rate of 10^{-5}. A spherical mask with a diameter of 95% of the box size was applied to each reconstructed 3D structure during training.

3.4.1.2 Observations

Training While trained on the synthetic dataset the algorithm converged to a reasonable solution. Some intermediate results from the training process are illustrated in Figure 3.4, Figure 3.5 shows the corresponding loss curves. Figure 3.4 monitors the training progress of the network for ten random samples from the test set. The depicted images are 3D structures projected along the z-dimension (which means perpendicular to the image plane)



(a) From top to bottom: reconstruction of test set after 500, 1k, 5k, 10k, 20k, 50k, 100k, 200k, 300k, 400k, 500k iterations

(b) Illustrated are 10 random samples from the test set.

Fig. 3.4: Reconstruction of simulated data. Subfigure (b) shows a random subset of the 2D test set. Subfigure (a) illustrates the training progress of the network. The illustrated images show projections of the reconstructed 3D structures that correspond to the 2D input images.

with the predicted angular information that would correspond to the angular information of the input image.

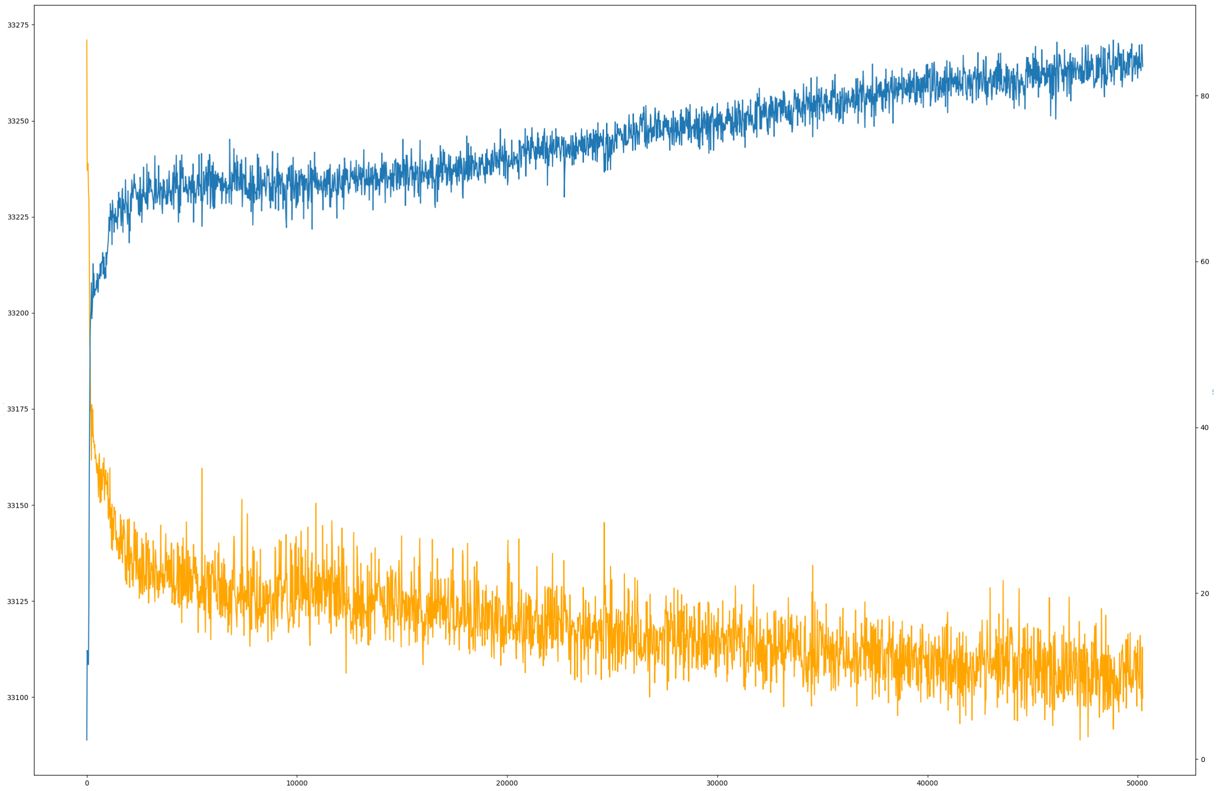


Fig. 3.5: Loss curves for synthetic test data. The loss was evaluated for every tenth iteration.

Already after a few hundred training steps the ANN approximated a mean structure, which provided a good fit for the given spectrum of rotations. Further training refined the L-shape. It became apparent that during training (e.g. steps 5k and 10k) the network learns a representation which incorporates two different chiralities (i.e. an overlap of two different L-shaped structures). During further training it rejects this handedness invariant representation towards the desired shape. In step 50k the overall shape is already very clear, the orientation of the object, however can be totally off. This means that shape and orientation of the object are disentangled to a certain degree. After around 300k steps the ANN clearly starts to overfit the background noise of the input images, which becomes apparent in noise around the L-shaped object in the reconstruction.

The loss curves correspond to that observations. The reconstruction error shows a sudden drop at the beginning of the training (when the mean structure is found) and then decreases gradually over time while still fluctuating heavily. The KL divergence antagonizes this development. While being low at the beginning it increases first sharply and then gradually. This can be explained by the complex interaction of encoder and decoder in the network. First, the posterior carries no information, which makes it unable to distinguish the currently presented sample from any other sample. Essentially, this is like drawing a random sample from the prior. As the training progresses the network learns to to put more information in the posterior (increased cost for the KL term) to gain a decreased cost in the likelihood term, by applying knowledge about the distribution of the data from the posterior.

Hyperparameter Tuning The selection of hyperparameters was done empirically and based on experience, as the training speed did not allow to explore a huge space of combinations. However, it turned out that the network is fairly robust with respect to the choice of the dimension of the latent space. Too few dimensions (i.e. below 10) resulted in the scenario that the network got stuck in a local optimum, comparable to the top line of Figure 3.4. Too many dimensions (i.e. more than 30) were less problematic, but often resulted in overfitting which manifested in smeared out 3D structures, rather than a structure that resembled the desired ground truth volume.

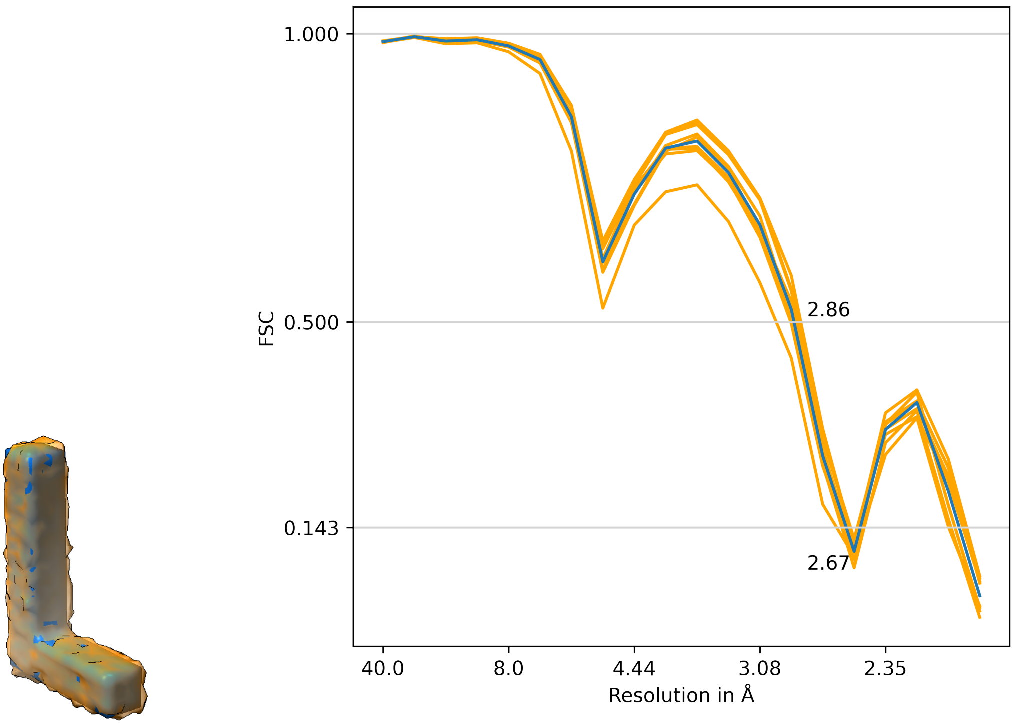
Another important hyperparameter has proven to be the learning rate. The selection of this parameter was crucial. A learning rate which was too high resulted in massive overfitting and prevented the network from learning any reasonable structure. In fact it just learned to reproduce the input images by smearing them out in the projection direction, by neglecting the higher order relationship between the images. On the other hand a too low learning rate resulted in extremely slow training, which, of course, is undesirable as well. 10^{-5} turned out to be a good value in practice and was used in all the experiments in this thesis.

Reconstruction Quality Assessment To assess the quality of construction performed by the ANN, multiple 3D structures were generated based on observed noisy 2D images. One reconstructed sample is depicted in Figure 3.6a in comparison with the ground truth volume (see Figure 3.2).

In Figure 3.6b the FSC curves for 10 reconstructed volumes were studied. To calculate the curves, the reconstructed volumes were aligned with the reference (ground truth volume). Noisy parts (i.e. reconstruction artifacts) that did not belong to the L-structures were removed beforehand. The alignment was realized based on the normalized cross correlation.

The FSC curves for the respective volumes show only slight deviations. The measured threshold values for the 0.5 criterion (2.86) and the 0.143 criterion (2.67) were identical due to the discrete nature of the FSC. Slight deviations are also caused by interpolation artifacts during alignment and slight misalignment of reference with respect to the reconstructed map for the evaluation.

Visually, the reconstructed volumes match the shape of the ground truth volume well. However, they show some softness, especially in the edges.



(a) Reconstructed map vs ground truth

(b) FSC curves (orange) for 10 volumes generated based on 2D images of the noisy input set compared with the ground truth volume. The mean of the curves is indicated in blue.

Fig. 3.6: Reconstruction quality of simulated data. Subfigure (a) shows a reconstructed 3D structure (orange, semi-transparent) with the ground truth volume (blue). Subfigure (b) depicts 10 FSC curves for an assumed pixel size of 1Å. The FSC values for the 0.143 and the 0.5 threshold are labeled in the diagram. The measured deviations in the FSC curves are small. Since the FSC is measured on a discrete shell level, the resulting values for the 0.5 and 0.143 criterion are identical for all curves.

Generating volumes As described previously, the ANN is capable of generating one estimated 3D structure per 2D input image. However, it can also generate valid 3D structures from unobserved regions of the latent space. To achieve this it is sufficient to decode a sample from the latent space with the trained decoder. It is therefore possible to generate samples based on e.g. a regular grid, or by random sampling.

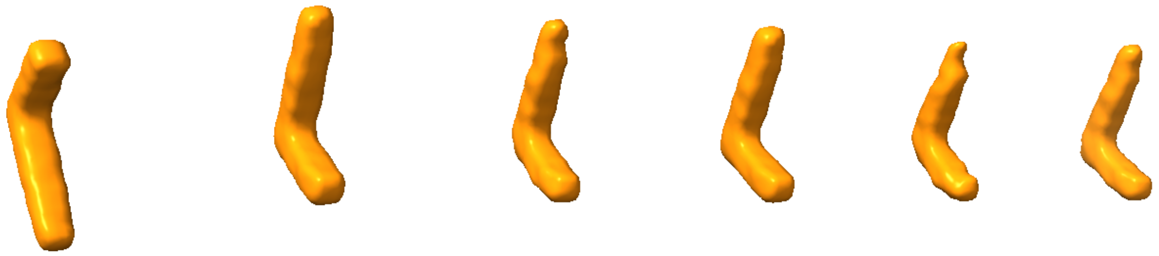


Fig. 3.7: Randomly generated samples from the latent space.

Figure 3.7 depicts 6 random sample 3D structures generated from the latent space without any previously observed corresponding 2D input images. The first sample shows a different handedness than the other five. The spatial orientation of volumes with the same handedness shows implicit alignment, which indicates that the neural network was capable of learning an orientation invariant representation of the L-structure.

3.4.2 Proof of Concept - 26S Proteasome

In real world scenarios cryoEM data often behaves differently compared to synthetic test data. There are various reasons for that. Varying imaging conditions, the distribution of noise, optical effects like aberrations in the TEM, an anisotropic distribution of projection angles and so on can deteriorate image quality.

In this experiment the performance of the ANN on a real world dataset was studied. The protein complex used was the human *26S proteasome*. The 26S proteasome is an essential macromolecular machine which plays an important role in protein degradation in the human body.

It consists of two main parts. A lower, socket like structure, the *20S proteasome* which is fairly rigid and a dynamic top region, the *19S lid*. The 20S part weights 700 kDa while the 19S lid has a weight of 900 kDa. The molecule measures around 315 Å in height and 200 Å in width. The 20S proteasome is D7 symmetric, while the whole 26S proteasome does not feature any symmetry. The shape of the protein complex is depicted in Figure 3.8.

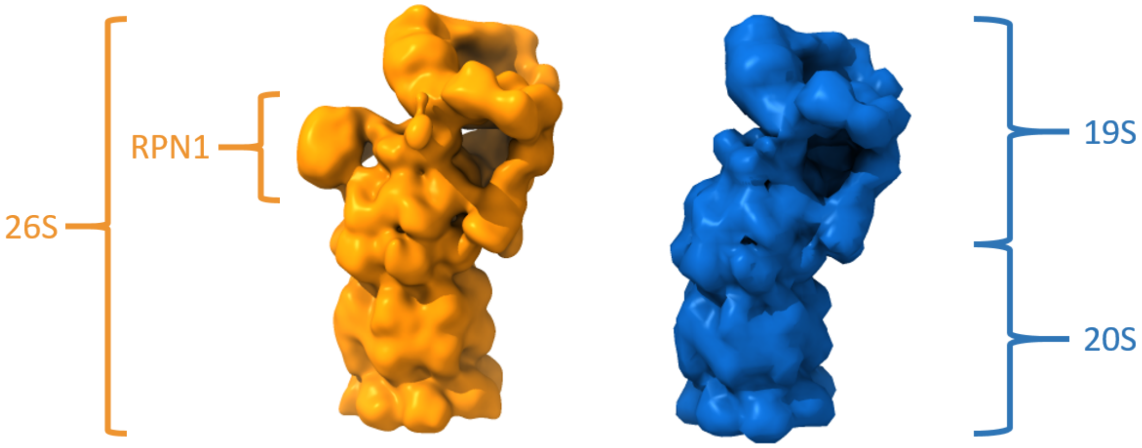


Fig. 3.8: 26S proteasome. The 26S proteasome consists of a fairly rigid lower part, the 20S proteasome and a dynamic part, the 19S lid. The highly dynamic Ribophorin I (RPN1) subunit, which is present in the orange structure [16], was masked in the blue structure[30]. Both structures were filtered to a resolution of 20.32 Å. The blue structure was used as a reference in the experiment.

3.4.2.1 Design of the Experiment

The dataset used for this experiment was the same as used in [30]. It consists of 640,000 images, recorded with a pixel size of 1.27 Å and a box size of 336 by 336 pixel. In order to reach reasonable computation times the dataset was coarsened by a factor of 8. The resulting box size was therefore 42 by 42 pixel, with a pixel size of 10.16 and based on that a (theoretical) maximum resolution of 20.32 Å could be obtained. Albeit present in the data, Haselbach et al. removed the RPN1 subunit to improve the alignment during the refinement process. Hence, it was missing in the reference structure, which needed to be considered during the experiment.

After coarsening the data was normalized with the tanh estimator (spread 0.3) and masked with a circular mask with a diameter of 95% of the box size. The data was then fed into the ANN without further modifications.

The training was performed on one of the previously described nodes with a total of 4 GPUs. The hyperparameters for the training were 21 dimensions for the latent space, a batch size of 120 images per GPU and a combined total learning rate of 10^{-5}.

3.4.2.2 Observations

Training A rough overview of the training progress can be found in Figure 3.9. It becomes apparent that the ANN quickly converged into a local optimum which resembled the dominant top view in the data. It took the network 10,000 training steps till it started to distinguish different views of the molecule, which was considerably longer compared to the synthetic test data (see Figure 3.4). This correlated with the drop in the projection loss and the increase in the KL divergence (see Figure 3.10).

It is worth noting here that the ratio between the projection loss and the KL divergence played an important role. If the weight for the KL divergence was too low, it resulted in a posterior collapse, which means that the KL divergence remained (close to) zero and the VAE was unable to learn a relationship between the observed samples of the dataset. This could be observed by either the reproduction of a spherical mean objection by the ANN for every projection or by overfitting against certain views (mostly top and side views), which resulted in smeared out 3D structures.

Further investigation of Figure 3.9 shows that the ANN first tries to determine a rough estimate for the orientation of the macromolecule in 3D space and afterwards tries to refines the shape. This hints towards the assumption that orientation is perceived as a distinct feature in early iterations of the training while later on the transfer of knowledge from other volumes results in a more orientation invariant representation and implicitly a less important role of the orientation of the molecule in space. In Figure 3.11 36 random sample particle images are depicted (Figure 3.11a) with their respective reconstruction (Figure 3.11b). In Figure 3.11c the reconstructed images were used as a colored overlay over the input images to illustrate that the orientation matches. In principle this could also be used for segmentation tasks (e.g. masking to reduce background noise, improving 2D classification etc.)

Naturally, the symmetric and fairly rigid 20S proteasome was approximated quicker than the dynamic 19S lid. This could be observed by a sharper, less blurry shape. The 19S lid remained blurry till the training process was aborted. Comparing iteration 1,000,000 and 2,000,000 the increase in reconstruction quality was insignificant. However, the amount of noise in the background of the reconstructed images increased which led to the assumption that the ANN started to overfit against background noise, rather than trying to further refine the shape of the molecule. Subsequently, the training process was aborted.

Another observation could be made regarding image five in Figure 3.9. During training the orientation and also the shape of the reconstructed object changed often and it remained blurry and unspecific till the end. Visual evaluation indicated that no actual particle was present in the input image. This means that minor overfitting effects against pure noise images can occur, but pure noise was not mistaken for actual particles.

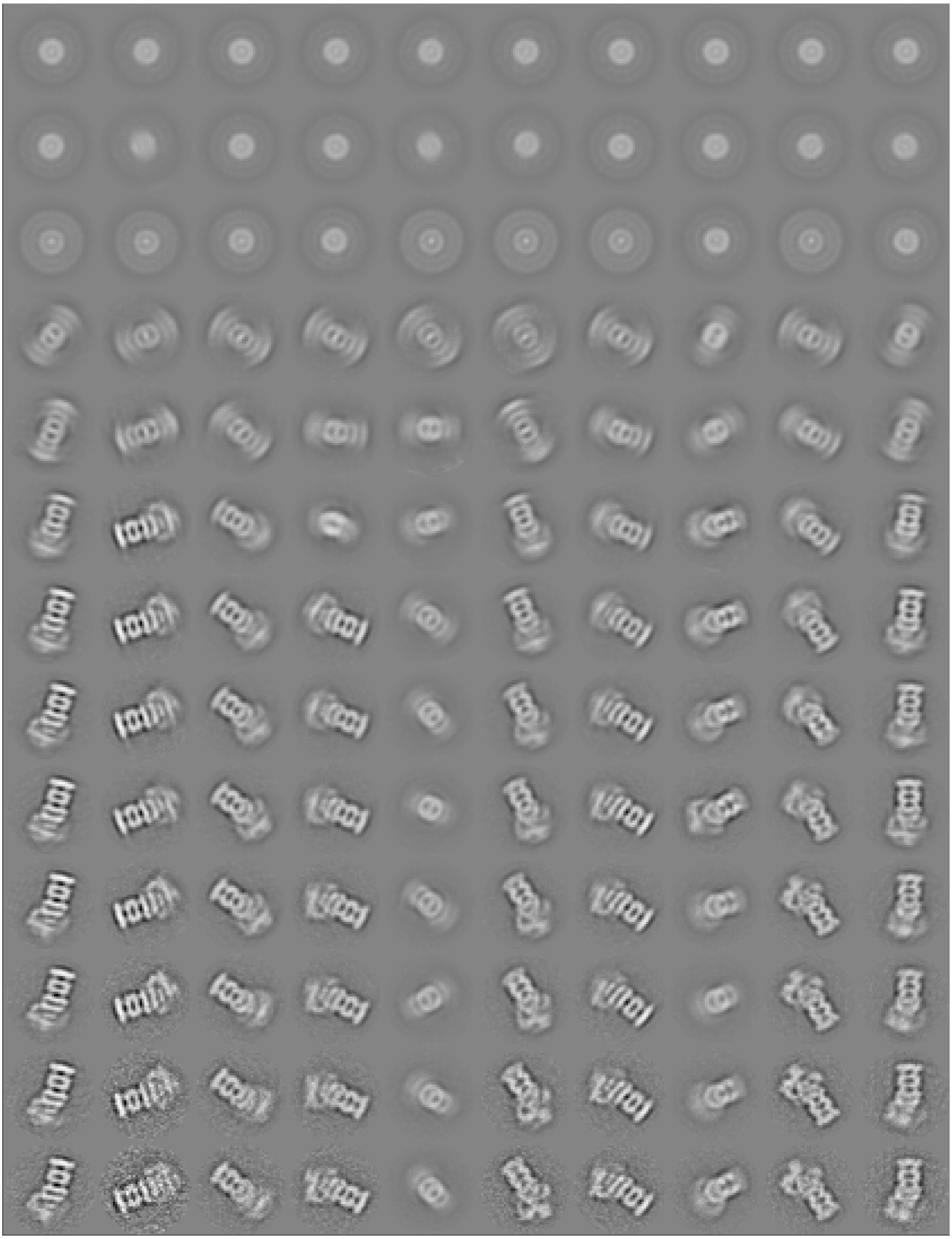


Fig. 3.9: Training progress of the 26S proteasome dataset.

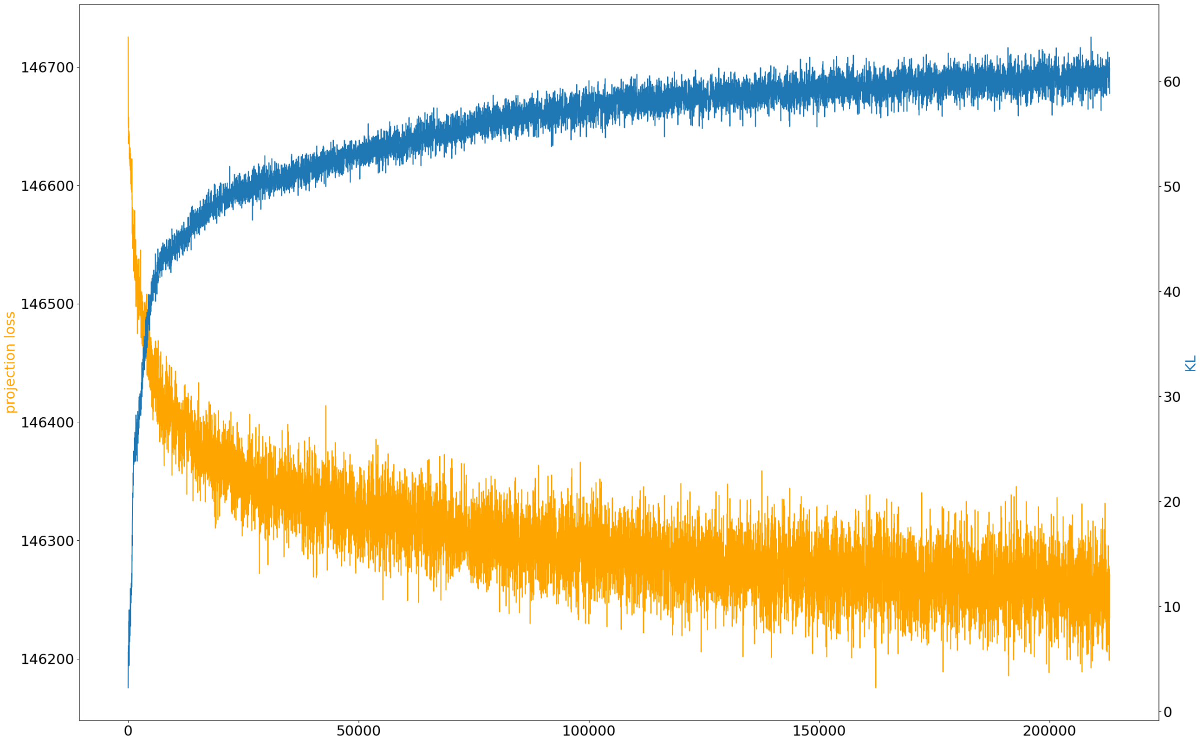
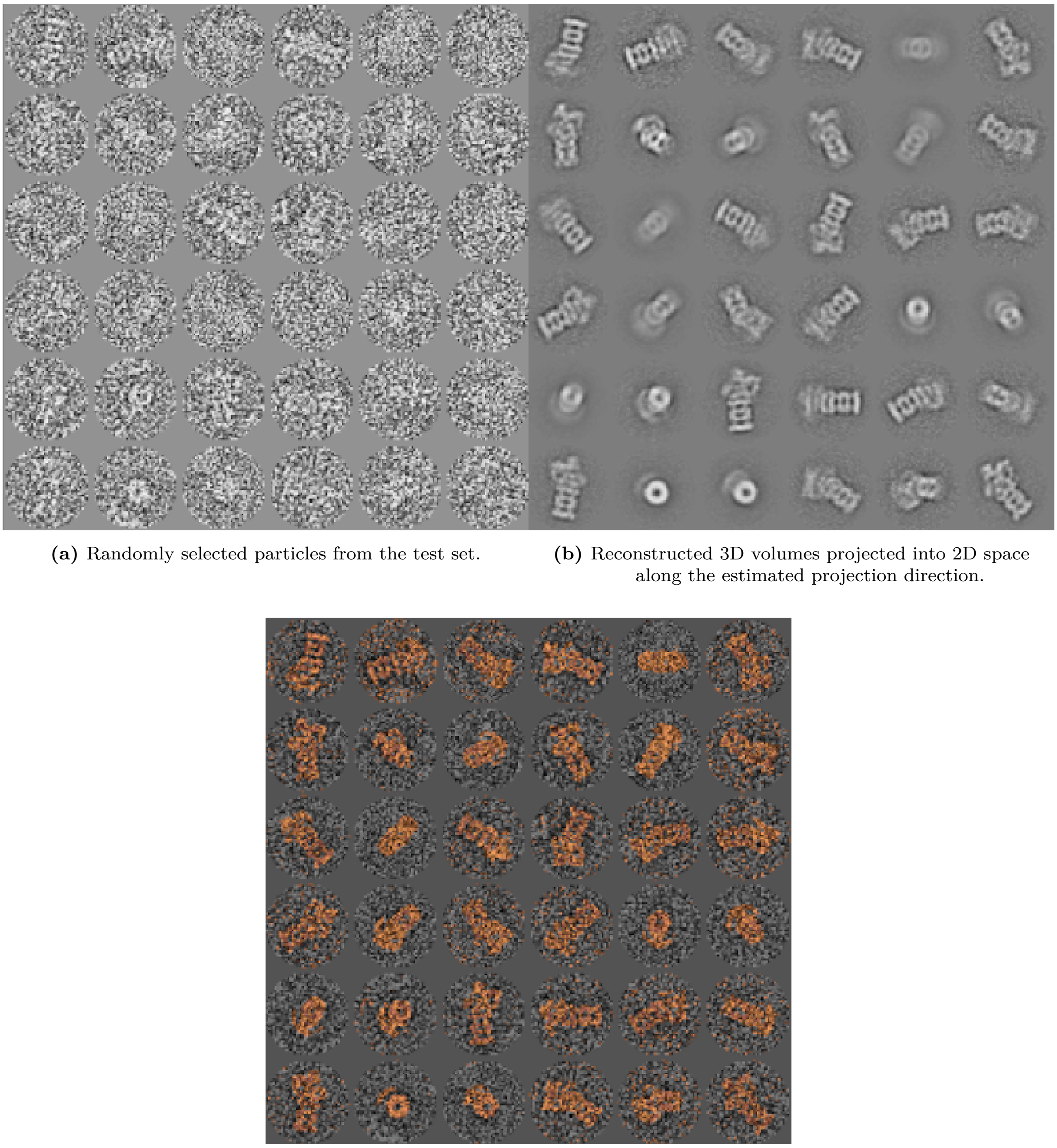


Fig. 3.10: Loss curves for the training on the 26S proteasome dataset. The loss was evaluated for every tenth iteration.



(a) Randomly selected particles from the test set.

(b) Reconstructed 3D volumes projected into 2D space along the estimated projection direction.

(c) Reconstruction as an orange overlay over the test images to illustrate the tightness of the fit.

Fig. 3.11: Reconstruction quality for 26S data. The reconstruction showed a tight fit when compared visually to the original particles. Especially in the steady and symmetric 20S part of the macromolecule. The dynamic, asymmetric 19S lid showed some blurriness. Images without any particles (e.g. first row, fifth image) didn't show particles in the reconstruction which indicated that the ANN is not very prone to overfitting. Artifacts in the test images (e.g. ice contaminations) were also neglected in the reconstruction, which is desirable. A larger set of test images can be found in Figure B.1

Reconstruction Quality Assessment The reconstruction quality must be assessed both visually (see Figure 3.12) and in terms of the FSC (see Figure 3.13), for the theoretical background of the FSC see Section 2.4.6.2.

As described previously, the ANN is capable of reconstructing one 3D structure per 2D input particle image. In Figure 3.12 six sample structures that were reconstructed by the ANN are depicted. In the 20S proteasome the reconstruction accuracy was high. Even in the degenerated cases Figure 3.12b and Figure 3.12f the reconstruction matched the reference fairly well. However, in the dynamic 19S lid the fit was less tight. Structures comparable to case **(b)** could sometimes be observed for top views, that resemble circles. The ANN then started to overfit these views and thereby ignored that also top views belong to valid structures. Case **(f)** can be result of too much dynamics in the area of RPN1, a mixture of two different cheiralities, or a circular average around the longitudinal axis of the molecule.

When evaluating the FSC curves it became apparent that structures **(a)**, **(c)**, **(d)** and **(e)** resembled each other which matched the visual evaluation. The FSC value of 23.71 Å for the 0.143 criterion was close to the maximum obtainable resolution, which equaled (based on the pixel size of 10.16 Å) 20.32 Å. As case **(f)** still roughly matched the shape of the reference in terms of low resolution features, the FSC value of 32.71 Å was not too far off. Case **(b)**, however, deviated too much from the reference, such that the resolution was poor.

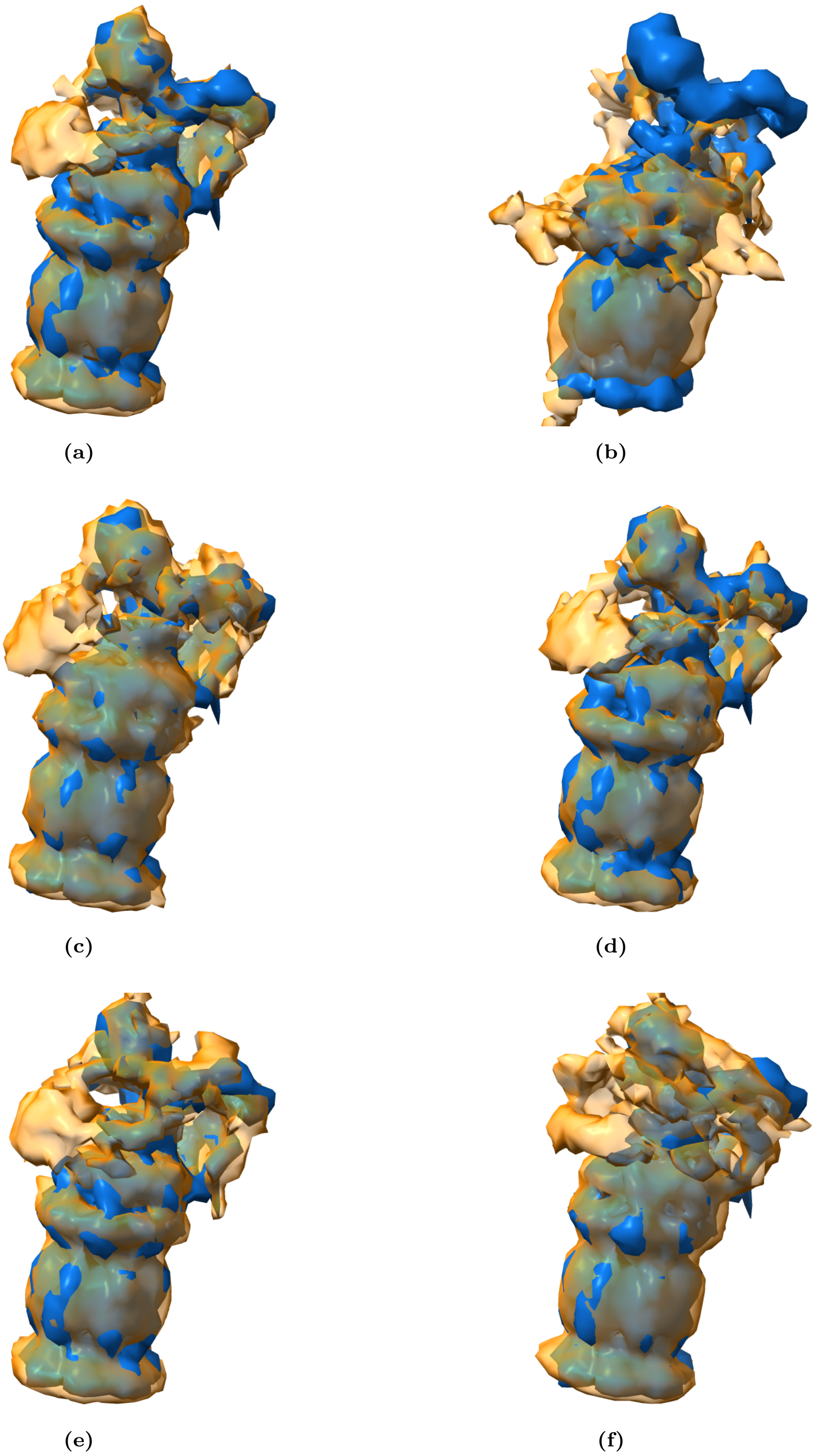


Fig. 3.12: Illustration of 6 reconstructed 3D structures (orange, semitransparent) in comparison with the reference structure (blue). Note that in [30] the highly dynamic RPN1 subunit was removed in order to improve 3D alignment and is therefore missing in the blue reference structure. Furthermore, some of the reconstructed volumes showed the wrong handedness and were therefore flipped along the z-axis to match the handedness of the reference. Structure **(b)** is a typical example for overfitting against a certain projection direction (e.g. top view). In structure **(f)** both chiralities are mixed, which results in a ring-like structure instead of the distinctive RPN1 subunit.

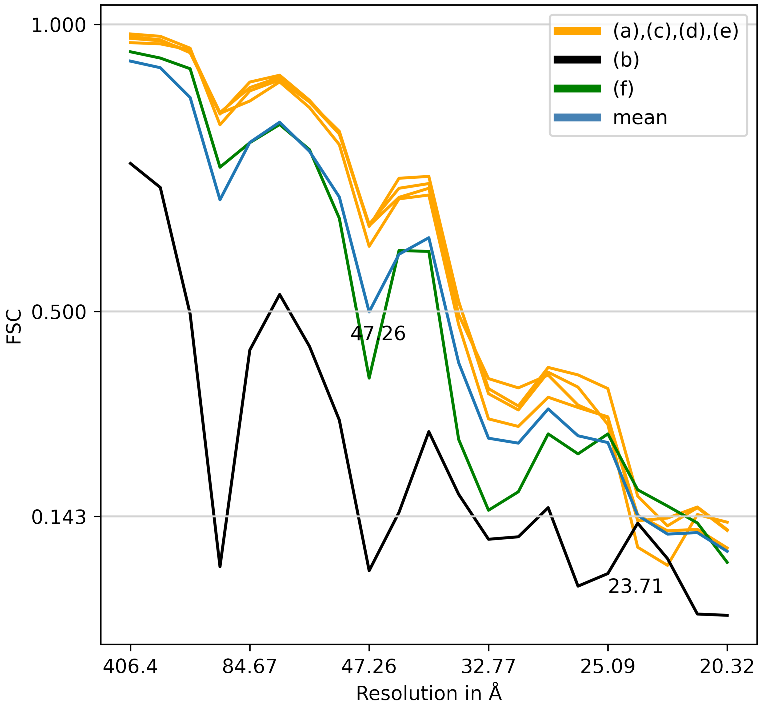


Fig. 3.13: The plot illustrates the FSC curves for the reconstructed volumes in Figure 3.12. The mean FSC values for the 0.5 and the 0.143 criteria are indicated in the plot. For the ease of readability, volumes **(a)**, **(c)**, **(d)** and **(e)** share the same color as their FSC curves resemble each other. Since RPN1 was removed in the reference structure, it was also removed in the reconstructed volumes before calculating the FSC curves.

3.4.3 Conformational Heterogeneity - Fatty Acid Synthase (FAS)

In the previous experiment it was demonstrated that the ANN is capable of reconstructing the 3D shape of a protein complex based on 2D cryoEM input images without prior knowledge about the data, especially, no reference structure was needed. In the following experiment it was studied how well it was able to deal with conformational dynamics within the data.

To demonstrate that the ANN can handle different macromolecules, another protein complex was introduced, the FAS. The FAS complex plays an important role in the synthesis of fatty acids in the human body. It can only fulfill its task by moving in an ordered manner. The FAS complex features a D3 (i.e. a sixfold) symmetry and weights around 2.6 MDa. It measures 270 Å in height and 250 Å in width. Singh et al. [65] have already studied the dynamics of this complex, such that two ground truth reference structures were available for comparison. Both of them show different states that were obtained by 3D classification and hierarchical sorting in RELION (see Figure 1.6). Both states and a superposition of the two of them can be found in Figure 3.14.

The two particle stacks used for the final reconstruction of the volumes were available, such that they could be merged into one dataset which predominantly showed both states when fed into the ANN. However, even after manual and computer aided sorting, particle stacks still include improper images (e.g. without particles, different conformations, contaminations).

3.4.3.1 Design of the Experiment

The dataset used for this experiment was composed of in total 255,000 images. 111,000 images belonged to the rotated state (blue), and 144.000 to the non-rotated state (orange). The data had an initial pixel size of 1.06 Å and a box size of 320 by 320 pixel. It was coarsened by a factor of 8 resulting in a (theoretical) maximum obtainable resolution of 16.96 Å.

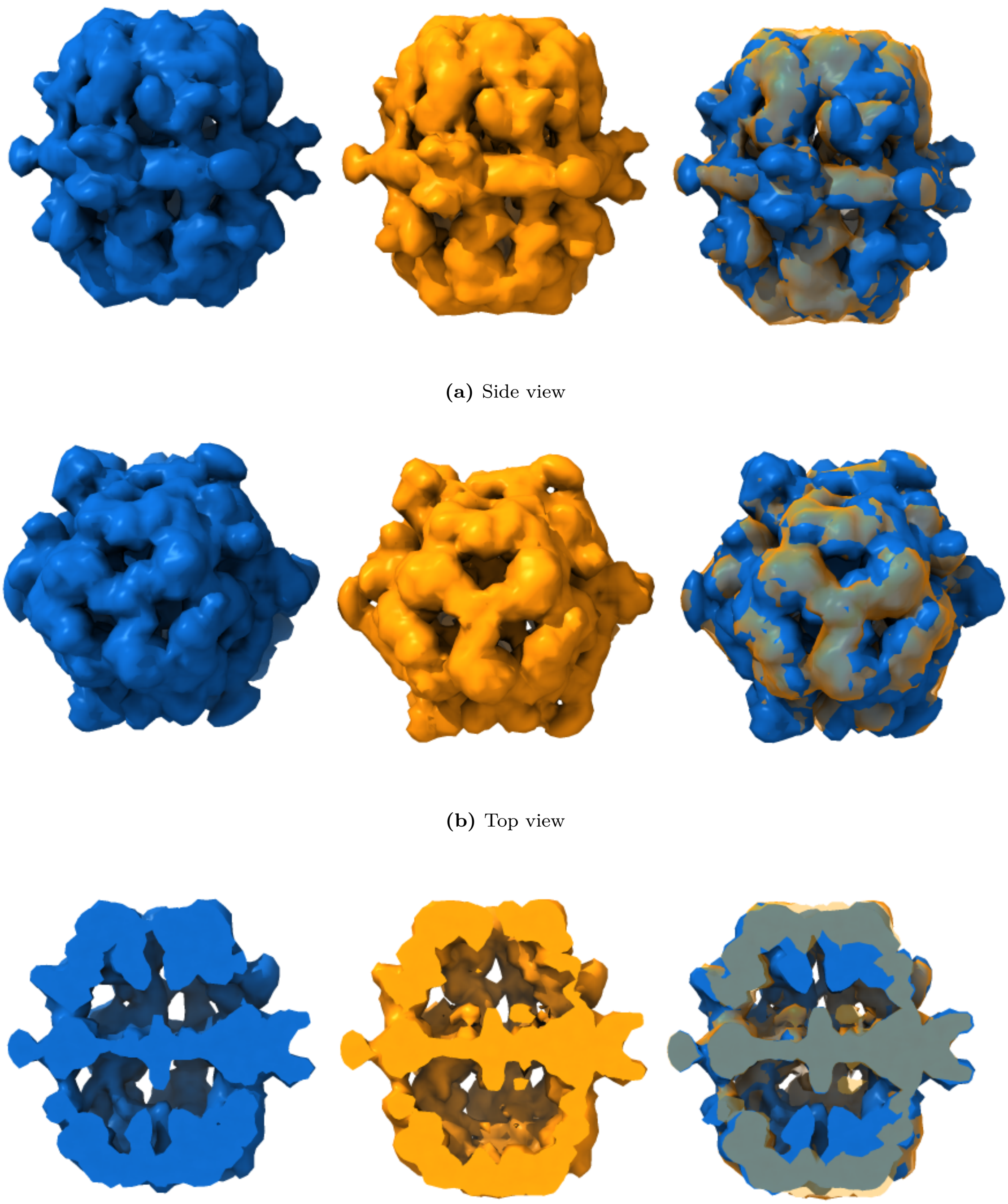
After coarsening the data was normalized with the tanh estimator (spread 0.3) and masked with a circular mask with a diameter of 95% of the box size. The data was then fed into the ANN without further modifications.

The training was performed on one of the previously described nodes with a total of 4 GPUs. The hyperparameters for the training were 21 dimensions for the latent space, a batch size of 120 images per GPU and a combined total learning rate of 10^{-5}.

3.4.3.2 Observations

Training The training on the FAS dataset converged much quicker compared to the 26S dataset. An overview of the training progress is given in Figure 3.15. Already after 1.000 iterations the ANN discovered dominant features in the images, e.g. the oval shape and the central alpha wheel, which determines the orientation of each particle in space. The top view (image 9 in Figure 3.15) took longer to be discovered. In general the training appeared to be more stable compared to the training of the 26S proteasome. The loss curves for the training of the FAS dataset are depicted in Figure 3.16. When compared to Figure 3.10 the projection loss drops quicker, while also the KL divergence increases quicker.

Naturally, the shape of the macromolecule plays an important role here. As already mentioned the FAS complex has more distinct features and additionally features a D3 symmetry. Another aspect might be that both chiiralities resemble each other more than in case of the 26S proteasome with the distinct RPN1 subunit. Also the kind of conformational dynamics might play a role, because in case of the FAS complex the



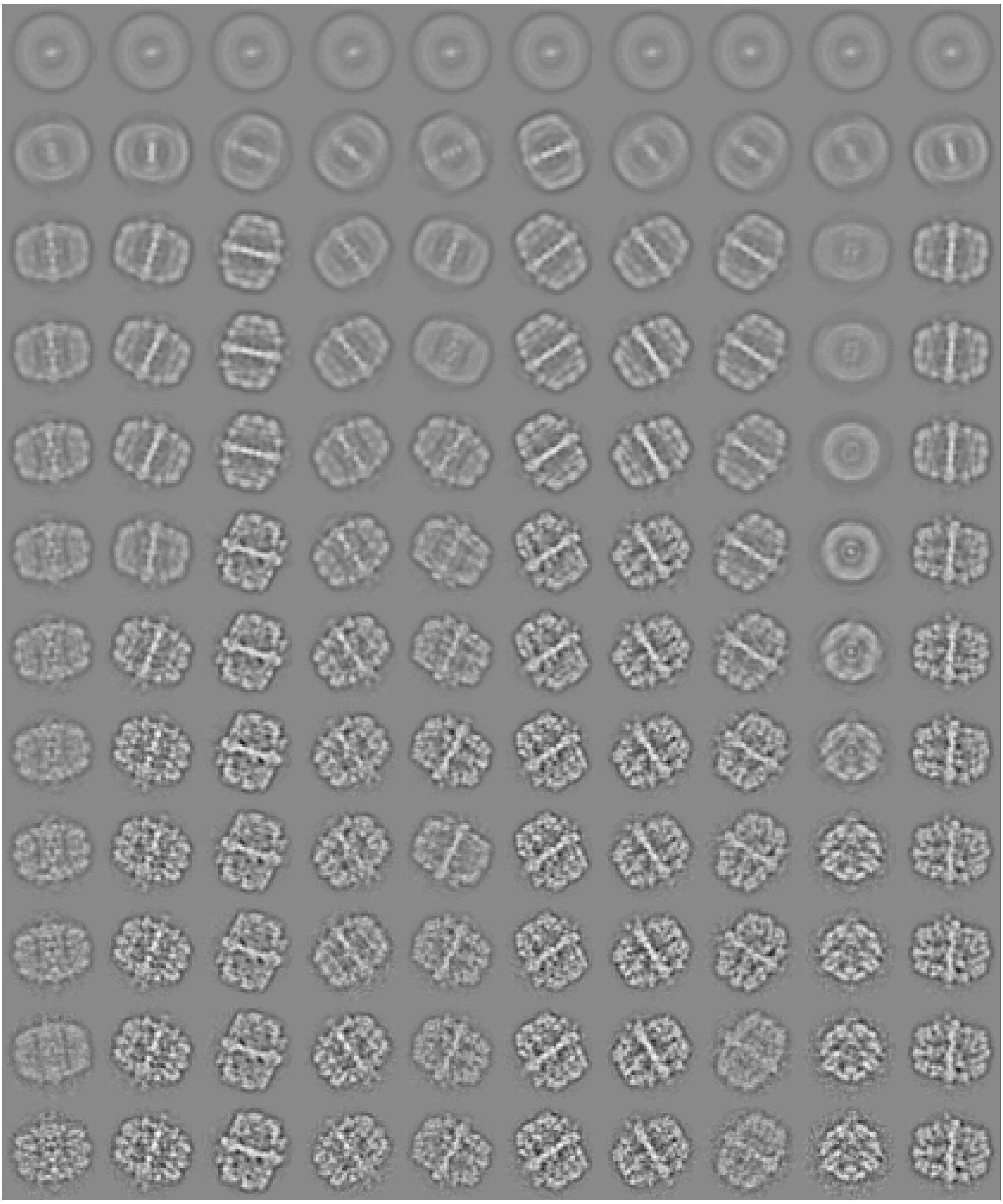
(a) Side view

(b) Top view

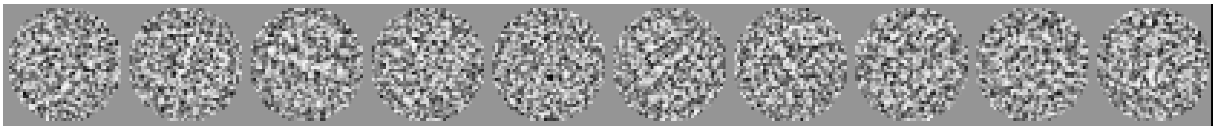
(c) Clipping through center

Fig. 3.14: Comparison of two different states of the FAS (blue and orange) and a superposition of both states. Both structures were reconstructed by Singh et al. [65]. The blue state (EMPIAR-10454) is more compact (i.e. shorter) and rotated compared to the orange state (EMPIAR-10470).

whole molecule rotates/shrinks, while for the 26S proteasome the lower half of the molecule remains fairly stable while it is mostly the lid which is moving.



(a) From top to bottom: reconstruction of test set after 500, 1k, 5k, 10k, 20k, 50k, 100k, 200k, 300k, 400k, 500k, 1M iterations



(b) Illustrated are 10 random samples from the test set, which correspond to the reconstruction in **(a)**.

Fig. 3.15: Reconstruction of FAS data. Subfigure **(b)** shows a random subset of the 2D test set. Subfigure **(a)** illustrates the training progress of the network. The illustrated images show projections of the reconstructed 3D structures that correspond to the 2D input images.

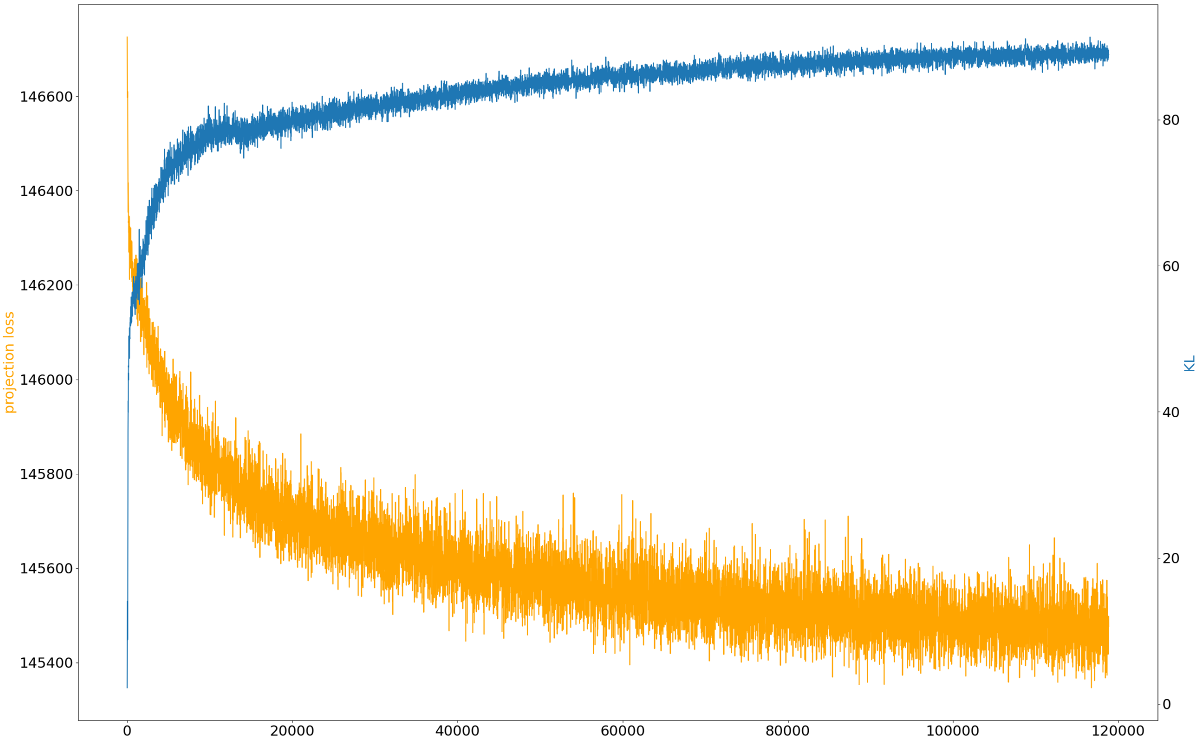


Fig. 3.16: Loss curves for the training on the FAS dataset. The loss was evaluated for every tenth iteration.

Reconstruction Quality Assessment The reconstruction quality for this experiment was again assessed visually and also in terms of the FSC curves. Two reconstructed volumes, that resembled the reference structures were chosen and afterwards fitted into their respective reference structure. The result can be seen in Figure 3.18. Visually evaluated, the fit for state 10470 looks tighter than the fit for 10454. However, this effect does not influence the FSC values much (see Figure 3.17). Based on the 0.143 criterion the resolution levels are identical for both structures (19.95 Å) compared to their references. When compared to the maximum obtainable resolution of 16.69 Å this is fairly close. In terms of the 0.5 criterion state 14454 reaches a lower resolution of 26.09 Å compared to 21.2 Å.

Zooming in on the details (see Figure 3.18e and Figure 3.18f) reveals a deviation that can potentially be systematic, which would hint towards the assumption that states exist in the reconstruction which are more tilted than the already tilted state 10454. However, the CCC based 3D alignment used for fitting the map into the reference could also be the cause of this behavior.

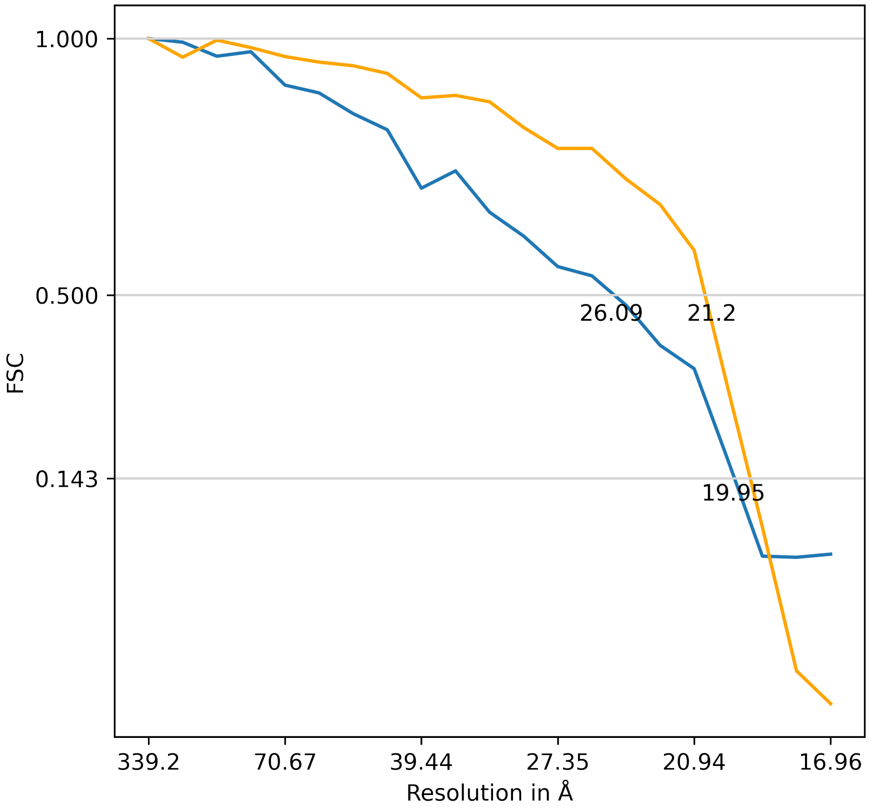
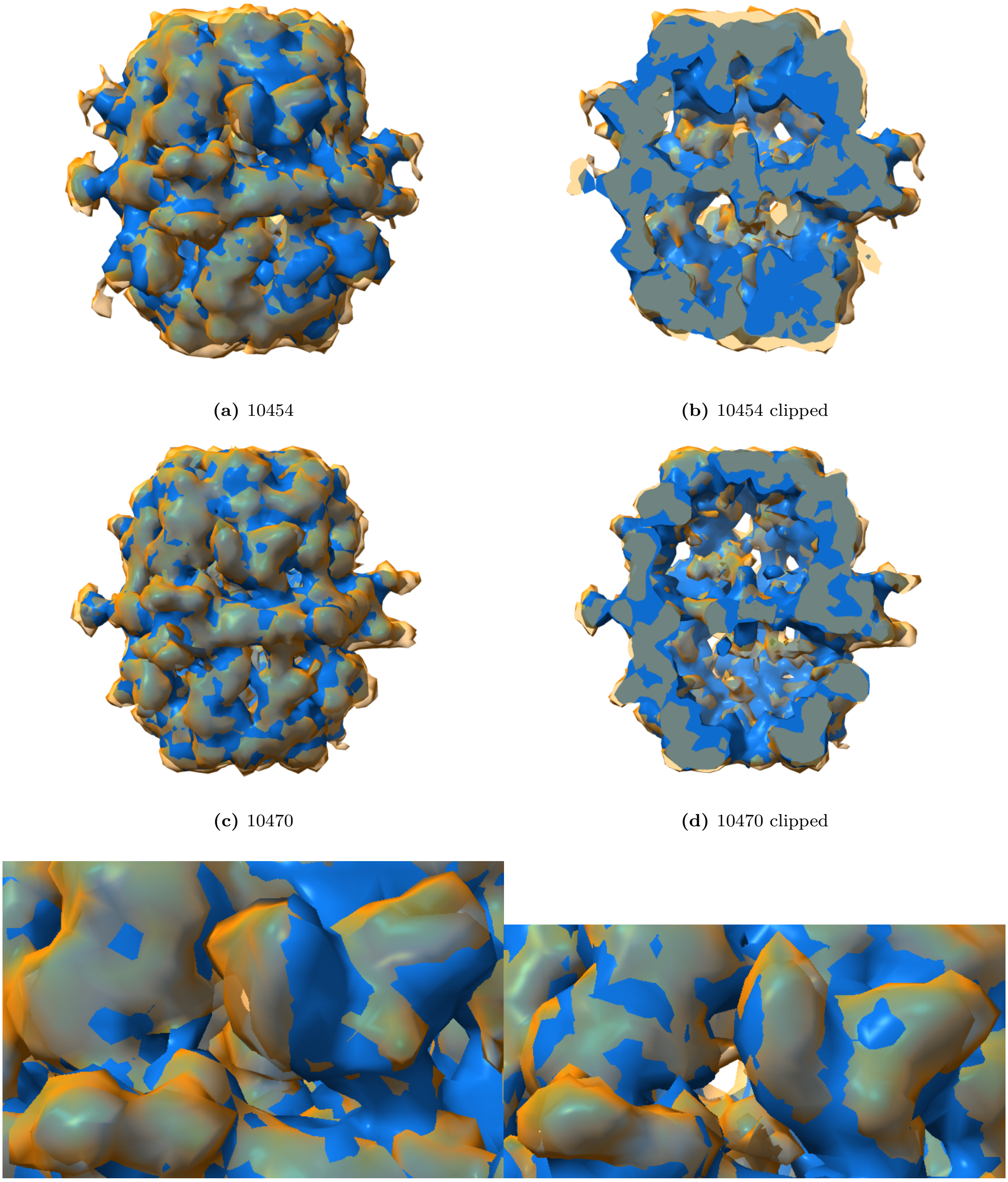


Fig. 3.17: FSC curves comparing the reconstruction with the ground truth for state 10454 (blue) and state 10470 (orange). The FSC values for the 0.143 criterion and the 0.5 criterion are marked in the plot. The theoretical maximum resolution is 16.96 Å



(a) 10454

(b) 10454 clipped

(c) 10470

(d) 10470 clipped

(e) 10454

(f) 10470

Fig. 3.18: Reconstruction of two different states within the dataset that resemble the reference volumes. Subfigure **(e)** and Subfigure **(f)** illustrate some details of the reconstructed structures. Visually evaluated, the fit for the non-rotated state (10470) seems tighter than for the rotated state (10454). However, the deviation from the reference seems to be systematic for 10454. This might be result of misalignment or artifacts in the reconstruction or could potentially hint towards the existence of FAS molecules that are even more rotated than the reference structure.

3.4.3.3 Exploring the Latent Space

In the previous experiments conformational dynamics played a less important role, therefore it was sufficient to study structures based on observed 2D particle images from the training set. However, when it comes to conformational dynamics, the ability of the ANN to generate intermediate structures that cannot be directly observed during training, gains importance. Therefore, during this experiment investigations on the anatomy of the latent space were made.

To gain an overview of the coverage of the latent space, 100,000 3D structures were generated based on observed 2D particle images. Their corresponding encoding vectors were gathered at the same time. As can be seen in Table 3.1 the dimensions resemble each other in terms of the distribution of their respective values. All of the dimensions are covered by the ANN with a mean of around zero and a standard deviation of around one.

Table 3.1: Statistics for the latent encodings (first 10 dimensions, for the complete table see Table B.1).

Latent dim	1	2	3	4	5	6	7	8	9	10
mean	0.01	-0.04	0.00	-0.09	0.00	0.16	0.00	0.00	0.01	-0.00
std	0.99	1.07	1.00	1.37	0.99	1.74	1.01	0.99	1.00	0.99
min	-4.51	-6.40	-4.43	-6.50	-4.27	-7.18	-4.50	-4.48	-4.23	-4.49
25%	-0.66	-0.51	-0.67	-1.01	-0.66	-1.08	-0.68	-0.66	-0.66	-0.67
50%	0.01	-0.06	0.00	-0.02	0.00	0.32	0.01	0.01	0.01	-0.00
75%	0.66	0.51	0.67	0.89	0.67	1.53	0.69	0.68	0.67	0.66
max	4.14	8.66	4.40	4.40	4.34	5.30	4.70	4.34	4.06	4.16

In the next step t-SNE (see Section 2.3.8) was applied onto the encoding vectors in order to see if certain cluster in the latent space existed and if they were meaningful. The result can be seen in Figure 3.19. The clustering with t-SNE did not produce any meaningful result, which means that the information is spread evenly across the latent space. The cross-validation with k-means clustering on the complete space discovered cluster in the data. However, the manual evaluation of the clusters did not show any systematics in the data, which can have multiple reasons. For example the number of classes was chosen arbitrarily. Another possible explanation would be that the features were not directly human-interpretable.

Therefore, the dimensions of the latent space were studied separately. This was done such that one dimension was sampled based on a regular grid from its minimum to its maximum, while the other dimensions were filled with their respective mean value. It

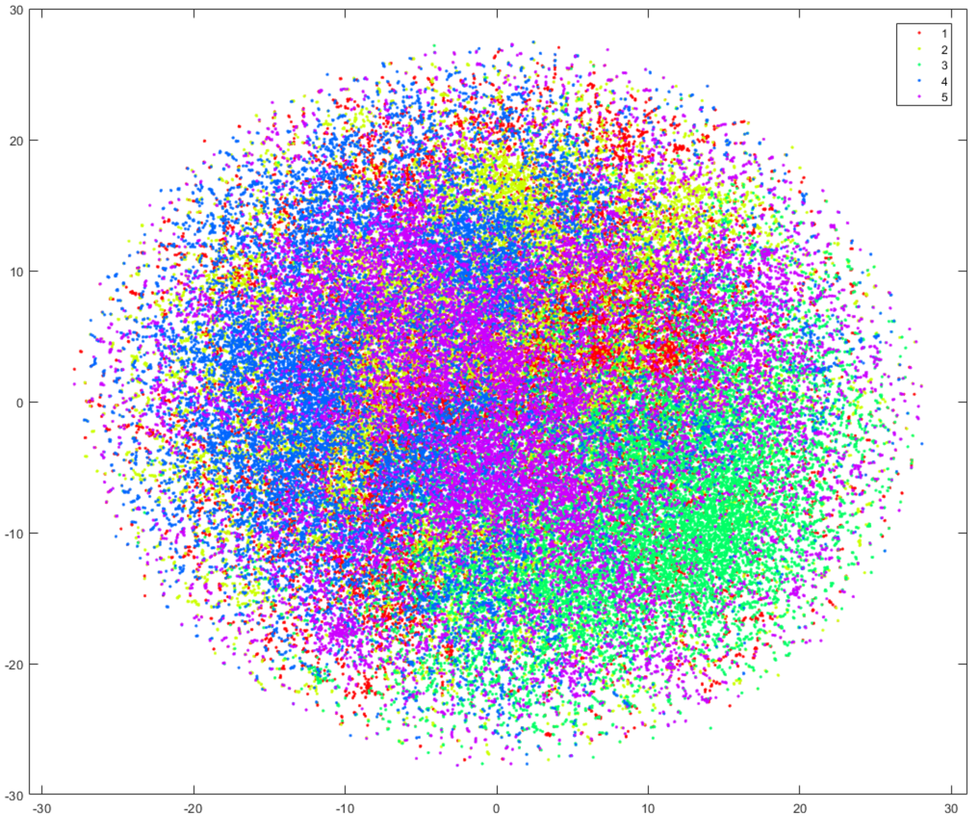


Fig. 3.19: This figure illustrates the t-SNE embedding for the latent encodings. The colored overlay was generated by performing k-means with 5 classes on the complete encodings in order to see if clusters could be discovered that are present in the projected 2D t-SNE space, but also the complete space. t-SNE was unable to discover reasonable clusters. K-mean suggests the existence of clusters, however when extracting 3D structures that belong to clusters discovered by k-means, no systematic patterns could be determined.

turned out that most of the dimensions showed different dynamics of the FAS complex. The generated 3D structures were aligned, as observed previously in other experiments, which indicated that the ANN was able to learn an orientation invariant representation of the data. However, both chiralities could be found in the latent space.

One of the key features of the ANN based approach is the ability to generate continuous trajectories from one vector in the latent space to another, such that e.g. a linear interpolation on the latent space can result in motion of the macromolecule in the voxel space. As the dynamics, discovered in the latent space, only showed minor amplitudes, which are hard to demonstrate in 2D, the latent space was used to interpolate between two existing chiralities in order to demonstrate noticeable “motion” of the macromolecule. Two volumes with different handedness’ were therefore chosen with their respective encodings. Afterwards a linear interpolation in six steps was performed from one vector to the other. The result can be seen in Figure 3.20. It becomes apparent that this kind of interpolation is possible and produces meaningful results. Unlike linear interpolations on the voxel (or PCA) space, the transport of mass preserves the overall integrity of the structure. Effects like the sudden appearance and disappearance of mass that would be result of linear interpolation on the voxel space could not be observed.

Interestingly, the interpolation discovered intermediate states which in principle could not be observed by the ANN, because they cannot exist in the actual data. This, in other words, means that the ANN is capable of generating a consistent conformational space, even without fully observing it during the training process. This would be advantageous compared to the discrete multi-body refinement approach by RELION, which only allows the interpolation between observed discrete states. However, more investigation and validation is needed here.

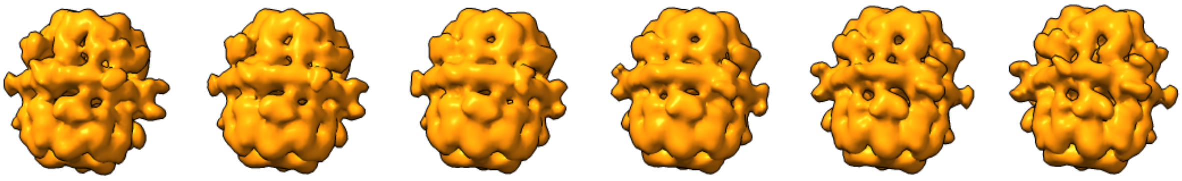


Fig. 3.20: Interpolation along the latent space. This figure illustrates how 3D structures can be generated by sampling the latent space in a regular manner. Note how the handedness of the FAS structure changes from left to right. Due to the continuous nature of the latent space, the sampling can be as coarse or fine as needed. Since the states with mixed handedness obviously were never observed in particle images, this hints towards the assumption that the latent space is ordered, such that neighboring 3D structures resemble each other. The ability to generate non- or rarely observed 3D states is of utmost importance to understand the dynamics (hence, the mechanisms) of biological macromolecules.

3.4.4 Structural Heterogeneity - Fatty Acid Synthase (FAS)

During the previous experiments the ANN was evaluated regarding its ability to reconstruct steady structures and also its ability to handle conformational heterogeneity. Another potential use-case is the study of heterogeneous datasets that not just include one kind of macromolecule. As described previously multiple reasons exist for this. One reason can be that the chemical conditions during sample preparation can cause the macromolecule to disintegrate.

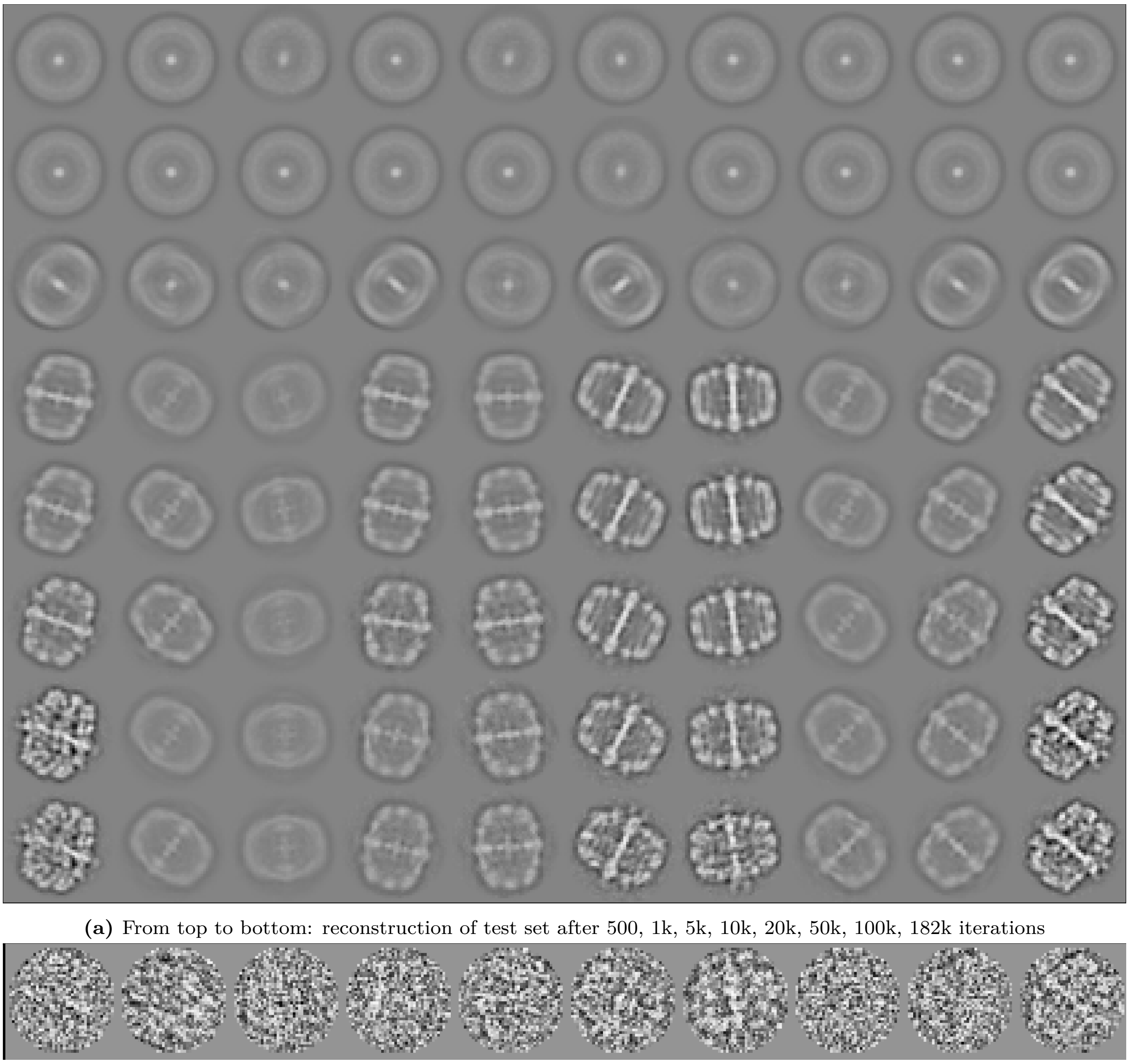
Often, structural heterogeneity causes difficulties in the reconstruction process, or prevent it from converging. This could be observed in a dataset of FAS molecules. In this experiment the ANN was trained on the dataset with broken particles, to investigate on its ability to deal with structural heterogeneity. Naturally, this is a much harder task than dealing with conformational heterogeneity, as the shape of broken particles can differ tremendously from their intact counterparts.

3.4.4.1 Design of the Experiment

The dataset used for this experiment was composed of 240,000 images. The box size was 160 by 160 pixel and was coarsened by a factor of 4, resulting in a box size of 40 by 40 pixel. As known from the previous experiments, the images were normalized with the tanh estimator (spread 0.3) and masked with a circular mask with a diameter of 95% of the box size. Afterwards they were fed into the ANN. The hyperparameters were again 21 dimensions for the latent space, the batch size was 120. Training was performed on one of the previously described nodes with 4 GPUs and a combined total learning rate of 10^{-5}.

3.4.4.2 Observations

Training An overview of the training process is given in Figure 3.21. In principle it did not differ much when compared to the training on the intact particles, except that more of the reconstructions looked blurry (see Figure 3.23). The loss curves (see Figure 3.22) also didn't reveal noticeable deviations from the known behavior. Various effects could be the cause of this phenomenon. One could be that in general there were also more images that did not contain any particle at all. Another possible explanation would be that the blurry images actually contain broken particles. At the moment there is no way to validate these assumptions, except visually. At least some of the images have shown features that resembled features of intact particles, others indeed appeared to be empty.



(a) From top to bottom: reconstruction of test set after 500, 1k, 5k, 10k, 20k, 50k, 100k, 182k iterations

(b) Illustrated are 10 random samples from the test set, which correspond to the reconstruction in **(a)**.

Fig. 3.21: Reconstruction of FAS data. Subfigure **(b)** shows a random subset of the 2D test set. Subfigure **(a)** illustrates the training progress of the network. The illustrated images show projections of the reconstructed 3D structures that correspond to the 2D input images.

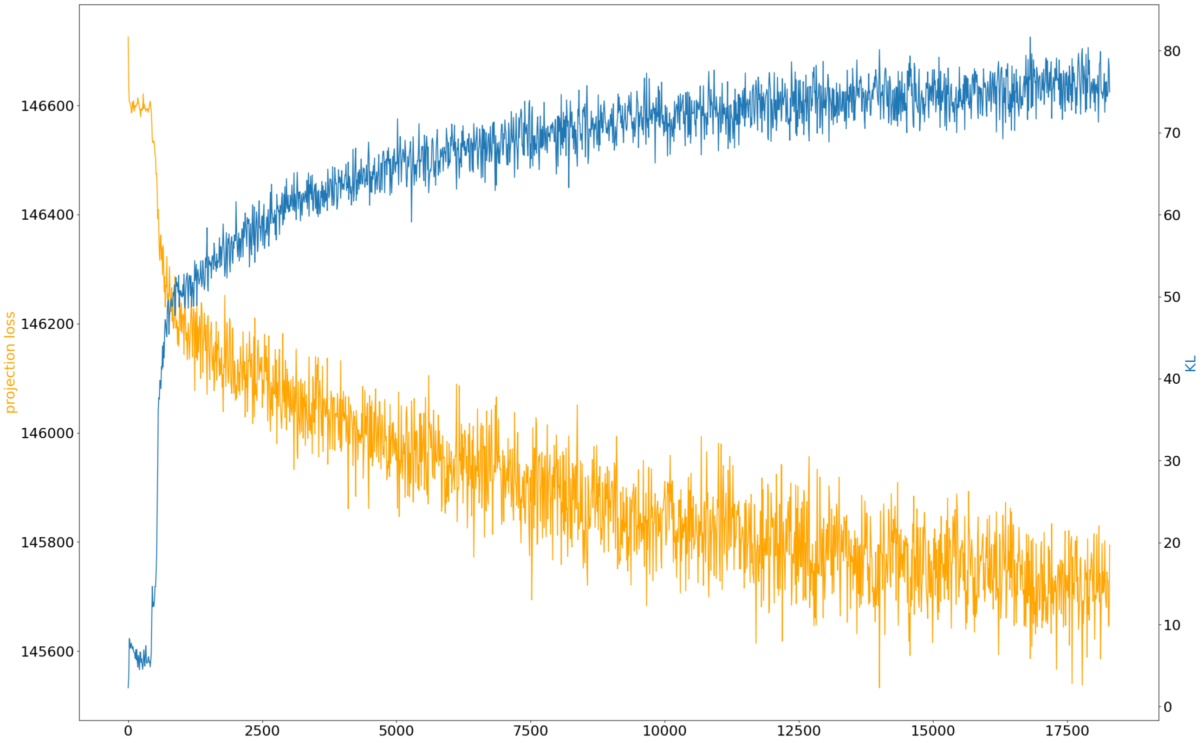
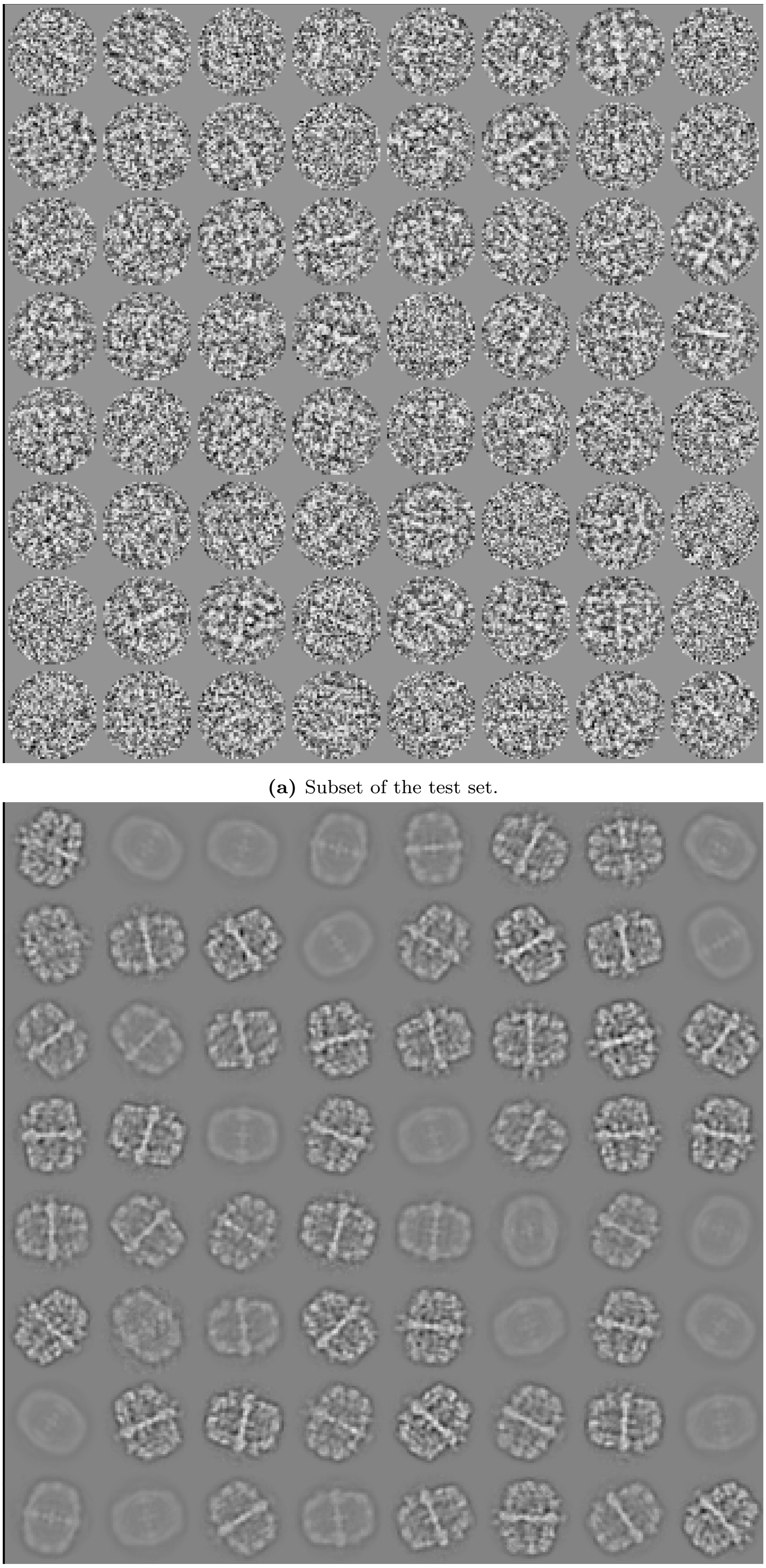


Fig. 3.22: Loss curves for the training on the FAS dataset with broken particles. The loss was evaluated for every tenth iteration.

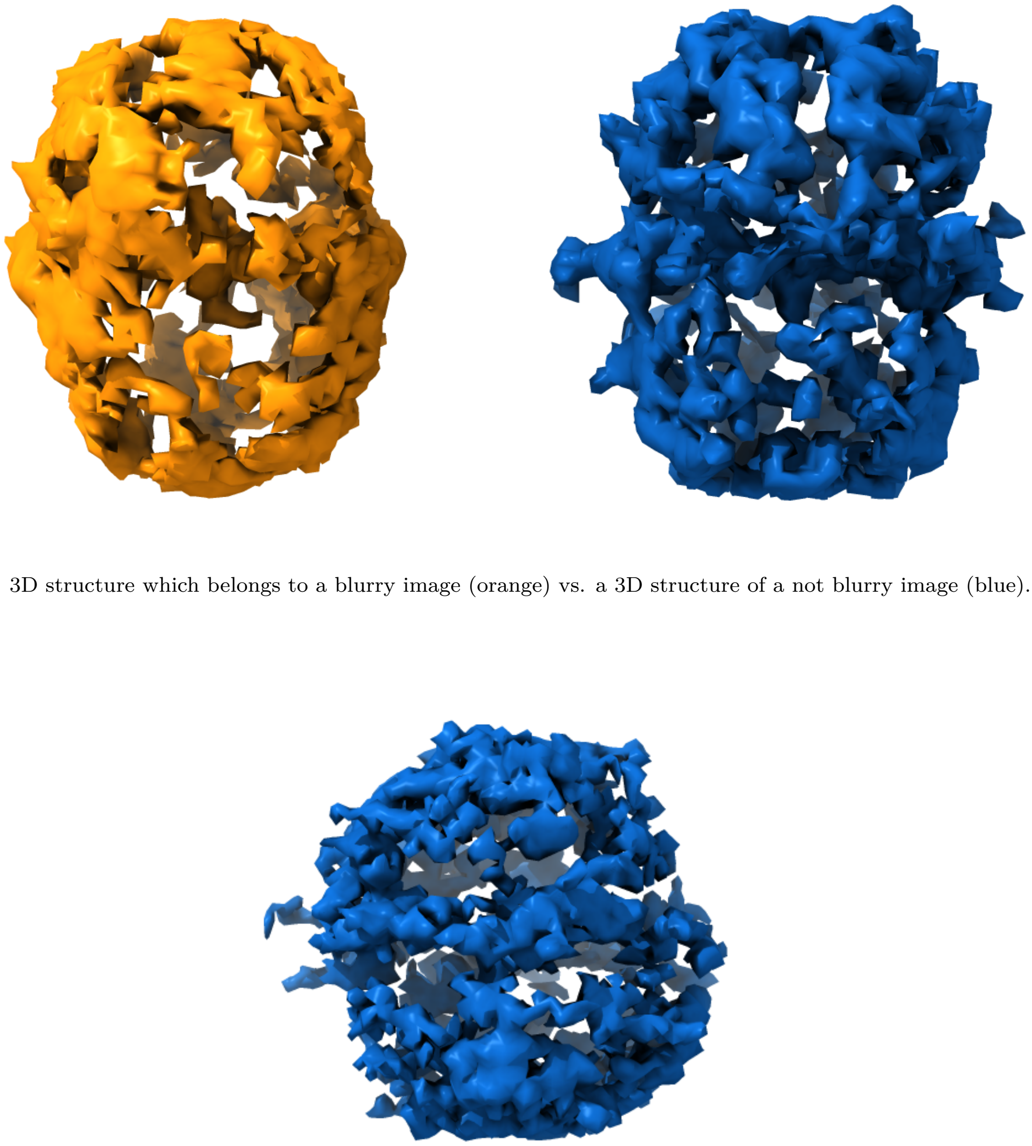


(a) Subset of the test set.

(b) Reconstruction of the ANN based on the images from the test set.

Fig. 3.23: Broken FAS complex. When compared to the reconstruction of datasets with intact particles it became apparent that significantly more reconstructions did not look like particles. Some test images, however, clearly showed features from particles (e.g. the central wheel). This might indicate that these test images contain broken particles, that were treated differently by the ANN compared to intact particles.

Reconstruction Quality Assessment The orange structure in Figure 3.24a only roughly matches the expected shape of a FAS molecule, but lacks important details. The blue structure resembles the expected shape. However, when rotated by 90 degrees (Figure 3.24b), it is obvious, that even potentially intact particles are not reconstructed correctly. They appear smeared out. Still, there is a noticeable difference between potentially broken and potentially intact, such that the ANN might be useful to distinguish them from each other, but more investigation is needed.



(a) 3D structure which belongs to a blurry image (orange) vs. a 3D structure of a not blurry image (blue).

(b) View on the blue structure in **a** rotated by 90 degrees.

Fig. 3.24: 3D structures of different FAS particles.

4 Discussion

Explanations exist: they have existed for all times,
for there is always an easy solution to every problem
– neat, plausible and wrong.
– H.L. Mencken

During the last decades, cryoEM has made enormous progress regarding every aspect of its methodical basis. The advances in biochemistry led to better sample quality, which allowed to study macromolecular complexes in high resolution in the first place, which on the other hand was only possible due to the development of instrumentation (i.e. TEMs), but also computational hardware (GPUs, storage) and advances in algorithmic image processing. This co-evolution has already enabled to study certain rigid protein complexes at atomic resolution (e.g. apoferritin). However, the vast amount of different software solutions for specific sub-tasks of the image processing pipeline often complicate reproducibility and validation. Some of them also use prior information on the data studied, which potentially induces bias, or require human intervention. It is therefore desirable to improve the process by streamlining more and more of the required steps.

One of these steps is the manual sorting process of different conformations. As cryoEM has emerged to be the method of choice when it comes to the study of dynamics in larger macromolecular complexes, this step has gained importance. The study and understanding of conformational dynamics is key to the understanding of how macromolecular machines work and which underlying mechanisms they are driven by.

Motion in the microscopic world of macromolecular complexes is – like motion in the macroscopic world – a continuous phenomenon. In the last chapter a novel machine learning based algorithm has been introduced which accounts for this fact by modeling the conformational landscape as the latent space of a VAE, a special kind of ANN. It was designed with the idea in mind that it should perform its task with as less human intervention as possible.

The following sections discuss and evaluate the outcome of the previously described experiments and emphasize on different use-cases for the algorithm.

4.1 Simulated Data

The algorithm has proven to work on simulated test data with a SNR that resembled the SNR in cryoEM density maps. The reconstructed structures matched the L-shape of the simulated input volume well. The FSC curves were calculated for an assumed pixel size of 1 Å. This in theory means a maximum obtainable resolution of 2 Å. The average FSC value for the very common 0.143 criterion (which is used for the gold standard) was 2.67 and hence almost identical with the maximum resolution. The FSC is also affected by slight misalignment of the compared volumes. As the 3D alignment was performed based on the CCC, which is prone to noise, slight misalignment is inevitable. Therefore one can conclude that in average the ANN was capable of reconstruction the volumes with the maximum obtainable resolution and also a high precision (consistency). However it is important to note that the synthetic dataset behaved differently, compared to real world datasets. The training process was quicker, the overfitting against noise present in the images took longer and the resolution was higher. Multiple reasons can be the possible cause. In general, image formation models assume the noise to be additive white gaussian noise. Therefore the synthetic test images was modeled that way. However, it is unclear if this assumption holds true in real world scenarios. Also the test set was created such that the projection angles were distributed based on a regular grid. In cryoEM datasets the distribution of projection angles, however, is usually not isotropic. Some projection angles are (based on the shape of the macromolecule) overpopulated compared to others, often top and side views. Also aberration effects of the TEM during imaging were not considered in the synthetic test set. These effects might deteriorate the reconstruction quality in real world scenarios.

Testing the algorithm on simulated data has also revealed important side effects that needed to be taken into consideration. The main effect that potentially affect the reconstruction quality turned out to be the handedness of the generated volumes. As cryoEM maps are in general invariant with respect to handedness, the ANN reconstructed volumes with similar shape, but differing handedness. Naturally, this means that the latent space, covered by the ANN, is not used up to its full potential. In state of the art algorithms this effect is avoided by using a reference structure, which implicitly biases the reconstruction process towards a certain handedness.

Like the generation of 3D structures based on noisy projection images, the generation of 3D structures by sampling the latent space turned out to work well. The internal representation of the data in the latent space showed (ignoring the handedness) implicit alignment. This on the one hand makes the latent space more powerful, as it is not affected by redundant information, but on the other hand enables for the creation of meaningful trajectories by systematically sampling the latent space (e.g. by linear interpolation on a regular grid).

4.2 Proof of Concept - 26S Proteasome

As described previously, real cryoEM datasets always behave differently, compared to test sets. Therefore a dataset from Haselbach et al. was used to quantify and validate the reconstruction quality of the ANN. The 26S proteasome was chosen because of its two different parts, the steady 20S proteasome and the highly dynamic 19S lid. It turned out that the ANN was capable of reproducing the reference map refined by REgularized LIkelihood OptiminzatioN (RELION) up to a resolution of 23.71 Å (0.143 criterion). Based on the theoretical maximum obtainable resolution of 20.32 Å this is close, but still a noticeable deviation. Multiple causes are possible for this deviation. When visually evaluating the reconstructed maps it becomes apparent that the steady 20S proteasome is much closer to the ground truth than the dynamic 19S lid. The reference was refined based on hierarchical 3D classification, such that the effects of conformational dynamics were mitigated. The ANN on the other hand has to deal with the complete spectrum of conformational dynamics in the dataset. The FSC curve for the comparison of the output of the ANN with the (steady) reference structure will therefore always show deviations, especially in the dynamic parts of the macromolecule. Structures with a resolution of 23.71 Å could without any problems be used as reference for state of the art algorithms like RELION or in the Cow Suite. It would also be an option to use multiple volumes generated by the ANN, that show different states of the conformational landscape, as a reference for the multi-body refinement of RELION in order to gain higher resolutions, if needed.

Two more effects could be observed in this experiment that were not observed using synthetic test data. One was that the ANN started to overfit against certain views (e.g. Figure 3.12b) which resulted in implausible 3D structures. This effect, however, was rarely observed and could be clearly identified by the FSC. The other effect can be observed in Figure 3.12f. For some particles the ANN reconstructed a ring-shaped mean structure in the 19S lid which might be caused by different chiralities, or by the dynamic nature of the region.

4.3 Conformational Heterogeneity - Fatty Acid Synthase (FAS)

In this experiment a dataset was studied that predominantly contained two different conformations of the FAS complex. The experiment has demonstrated that the presented algorithm works as expected. The ANN was capable of reconstructing both conformations up to a difference of 3 Å in resolution, when evaluated based on the 0.143 criterion of the FSC. Visual evaluation hints towards the assumption that some of the deviations with respect to the reference structure might be of systematic nature, e.g. result of conformational dynamics rather than error in the reconstruction.

Investigations on the latent space revealed that it is fully covered (i.e. all dimensions are used) and that it indeed encodes conformational dynamics. The 3D structures sampled from the latent space showed spatial alignment, which indicates that the representation learned by the ANN shows invariance regarding the pose of the macromolecule. However, like in the other experiments, different chiralities could be observed. The phenomenon of different chiralities most likely also affects the reconstruction accuracy or at least blocks resources in the latent space. It is to be expected that solving the issue with handedness will further increase the performance of the algorithm presented. The effect, however, was used to study the ability of the network to produce unobserved intermediate states. To achieve this, an interpolation from one handedness to the other was performed. It revealed that the ANN assumes the existence of intermediate states, e.g. a mix of both chiralities, that do not exist in real world applications. The ability of the network to generate plausible intermediate states based on observed states is extremely useful for the study of the motion of biological macromolecules. As the latent space is continuous by design it is also possible to sample as many intermediate states as needed, which sets the VAE based algorithm apart from state of the art algorithms that assume a finite amount of states.

Even though FAS and the 26S proteasome are fairly different in terms of shape and dynamics it turned out that further hyperparameter tuning was not necessary. The same set of parameters applied to both datasets. In general the role of hyperparameters should be investigated in more detail. This, however, requires an increase in computation speed, as the systematic evaluation of different sets of hyperparameters can lead to a combinatorial explosion of possible sets.

Furthermore, a more detailed investigation of the anatomy of the latent space is needed, e.g. on the distribution of different chiralities, the position of highly resolved structures compared to low resolved structures and so on. This could help to further improve the algorithm.

As the resolution for the reconstructed 3D structures does not deviate much from the ground truth it would also be interesting to see if state of the art tools like RELION or the Cow Suite can benefit from the results produced by the ANN, e.g. by using the output as starting structure for their refinements, for 3D classification and so on.

4.4 Structural Heterogeneity - Fatty Acid Synthase (FAS)

The handling of structural heterogeneity is not the main purpose of the presented algorithm. However, it revealed some interesting effects. First, it showed no noticeable difference in the training process compared to the dataset used in the previous experiment. However, the output looked much different. A lot of the reconstructed output images did not show characteristic features of the FAS molecule. Other structures showed characteristic features, but had a very strong bias towards a certain view. All that indicates that the ANN was unable to learn a plausible conformational space in the presence of structural heterogeneity. Nevertheless, the reconstructions for (presumably) particles looked much different than for (presumably) non-particles. Some chances exist here that the tool could be used to distinguish particles and non-particles. It might be interesting here to try if multiple iterations of training on the same dataset would improve on the reconstruction result, e.g. by removing the non-particles from the image stack.

5 Conclusion and Outlook

We can only see a short distance ahead,
but we can see plenty there that needs to be done.
– A. Turing

The study of the dynamics of biological macromolecules using cryoEM has gained more and more attention over the years, as is the only method known to date which can handle larger macromolecular complexes. The effort, however, that needs to be put into data processing and conformational sorting is enormous. Hence, software tools that reduce the amount of human input are required.

In this thesis a novel machine learning based software tool was presented. It has proven to be effective as a fully autonomous tool for the reconstruction of 3D structures from 2D cryoEM density maps. Furthermore it demonstrated its ability to estimate a continuous conformational space of the studied biological macromolecule, while not making any assumptions on the molecule, like symmetry or starting models.

Even though the algorithm works well in certain use cases numerous ideas exist to improve on its capabilities. For example the performance could be improved by swapping the transformation step, which requires a modification to the whole grid, for point clouds, which is expected to be more efficient. To further increase reconstruction accuracy and reliability it is crucial to overcome the challenge of different handedness during the reconstruction process.

□

A Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
ANN	Artificial Neural Network
API	Application Programming Interface
CCC	cross correlation coefficient
CCD	charge coupled device
CPU	Central Processing Unit
cryoEM	3D transmission electron cryo-microscopy
CST	central slice theorem
CTF	Contrast Transfer Function
DDD	direct electron detection device
EM	electron microscope
FAS	Fatty Acid Synthase
FEG	Field Emission Gun
FRC	Fourier Ring Correlation
FSC	Fourier shell correlation
FFT	Fast Fourier transformation
GPU	graphics processing unit
HPC	High Performance Computing
iFFT	inverse Fast Fourier transformation
i.i.d.	independent and identically distributed
KL	Kullback–Leibler
MPI	Message Passing Interface
MSE	mean squared error
NMR	Nuclear Magnetic Resonance spectroscopy
PCA	Principle Component Analysis
PDH	Pyruvate dehydrogenase
PSF	Point Spread Function
RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
RELION	REgularized LIkelihood OptiminzatioN
ReLU	Rectified Linear Unit
RPN1	Ribophorin I
SEM	Scanning Electron Microscope
SIFT	Scale Invariant Feature Transform
SIRT	Simultaneous Iterative Reconstruction Technique
SURF	Speeded Up Robust Features
SNR	signal-to-noise ratio
SPA	Single Particle Analysis
SSE	sum of squared errors
STEM	Scanning Transmission Electron Microscope
TEM	transmission electron microscope
t-SNE	t-distributed stochastic neighbor embedding

VAE	Variational Autoencoder
XRC	X-ray Crystallography

B Supplementary information

B.1 26S Proteasome

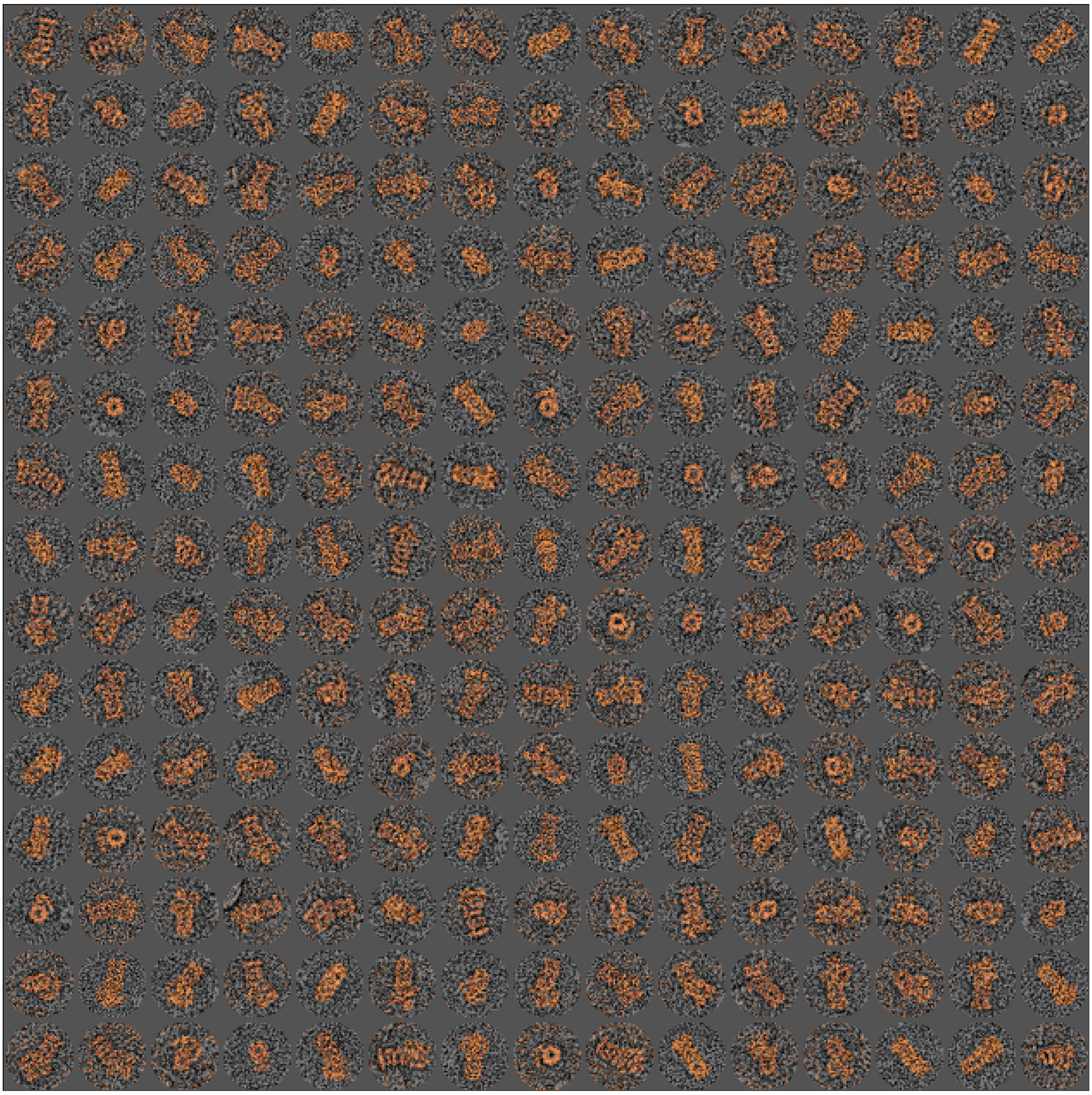


Fig. B.1: Reconstruction of the ANN as an orange color overlay over samples from the original dataset.

B.2 Conformational Heterogeneity - Fatty Acid Synthase (FAS)

Table B.1: Statistics for the latent encodings.

Latent dim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
mean	0.01	-0.04	0.00	-0.09	0.00	0.16	0.00	0.00	0.01	-0.00	-0.14	0.20	-0.01	0.01	0.03	0.01	-0.00	0.00	-0.01	-0.01	0.32
std	0.99	1.07	1.00	1.37	0.99	1.74	1.01	0.99	1.00	0.99	1.50	1.84	0.99	1.01	1.04	0.96	0.99	1.00	1.00	1.02	0.98
min	-4.51	-6.40	-4.43	-6.50	-4.27	-7.18	-4.50	-4.48	-4.23	-4.49	-11.48	-7.68	-4.37	-4.11	-7.16	-4.36	-4.56	-4.68	-4.27	-4.46	-4.23
25%	-0.66	-0.51	-0.67	-1.01	-0.66	-1.08	-0.68	-0.66	-0.66	-0.67	-1.11	-1.05	-0.67	-0.68	-0.42	-0.64	-0.68	-0.67	-0.68	-0.69	-0.28
50%	0.01	-0.06	0.00	-0.02	0.00	0.32	0.01	0.01	0.01	-0.00	-0.01	-0.32	-0.01	0.01	-0.01	0.01	-0.01	0.00	-0.01	-0.01	0.42
75%	0.66	0.51	0.67	0.89	0.67	1.53	0.69	0.68	0.67	0.66	0.91	1.80	0.67	0.69	0.45	0.65	0.67	0.67	0.67	0.68	1.01
max	4.14	8.66	4.40	4.40	4.34	5.30	4.70	4.34	4.06	4.16	5.00	6.45	4.00	4.67	7.83	3.96	4.32	4.36	4.56	4.20	5.36

C References

[1] Verse by verse ai. `https://sites.research.google/versebyverse/`, 2021. [Online; accessed 04-August-2021].

[2] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[3] Ernst Abbe. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–468, 1873.

[4] A Andersen. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, January 1984. doi: 10.1016/0161-7346(84)90008-7. URL `https://doi.org/10.1016/0161-7346(84)90008-7`.

[5] Jean Baptiste Joseph baron Fourier. *Théorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822.

[6] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. ISSN 01628828. doi: 10.1109/TPAMI.2013.50.

[8] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 16. 2007. ISBN 9780387310732. doi: 10.1117/1.2819119. URL `http://www.library.wisc.edu/selectedtocs/bg0137.pdf`.

[9] B. Böttcher, S. A. Wynne, and R. A. Crowther. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy, 1997. ISSN 00280836.

[10] R. N. Bracewell. Numerical transforms. *Science*, 248(4956):697–704, 1990. ISSN 00368075. doi: 10.1126/science.248.4956.697.

[11] S Brenner and RW Horne. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta*, 34:103–110, 1959. ISSN 00063002. doi: 10.1016/0006-3002(59)90237-9. URL `https://linkinghub.elsevier.com/retrieve/pii/0006300259902379`.

[12] E. O. Brigham and R. E. Morrow. The fast fourier transform. *IEEE Spectrum*, 4(12):63–70, December 1967. doi: 10.1109/mspec.1967.5217220. URL `https://doi.org/10.1109/mspec.1967.5217220`.

[13] Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte, Shuchismita Dutta, Zukang Feng, Sai Ganesan, David S Goodsell, Sutapa Ghosh, Rachel Kramer Green, Vladimir Guranović, Dmytro Guzenko, Brian P Hudson, Catherine L Lawson, Yuhe Liang, Robert Lowe, Harry Namkoong, Ezra Peisach, Irina Persikova, Chris Randle, Alexander Rose, Yana Rose, Andrej Sali, Joan Segura, Monica Sekharan, Chenghua Shao, Yi-Ping Tao, Maria Voigt, John D Westbrook, Jasmine Y Young, Christine Zardecki, and Marina Zhuravleva. RCSB protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, 49 (D1):D437–D451, nov 2020. doi: 10.1093/nar/gkaa1038.

[14] Tom Burnley, Colin M. Palmer, and Martyn Winn. Recent developments in the CCP-EM software suite. *Acta Crystallographica Section D: Structural Biology*, 73(6):469–477, 2017. ISSN 20597983. doi: 10.1107/S2059798317007859.

[15] Boris Busche. New Algorithms for Automated Processing of Electronmicroscopic Images. page 158, 2013.

[16] Xiang Chen, Zachary Dorris, Dan Shi, Rick K Huang, Htet Khant, Tara Fox, Natalia de Val, Dewight Williams, Ping Zhang, and Kylie J Walters. Cryo-em reveals unanchored m1-ubiquitin chain binding at hrpn11 of the 26s proteasome. *Structure (London, England : 1993)*, 28(11):1206,Äì1217.e4, November 2020. ISSN 0969-2126. doi: 10.1016/j.str.2020.07.011. URL https://doi.org/10.1016/j.str.2020.07.011.

[17] Wikimedia Commons. Joseph fourier (circa 1820), 2010. URL https://commons.wikimedia.org/wiki/File:Joseph_Fourier_(circa_1820).jpg. File:Joseph Fourier (circa 1820).jpg.

[18] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. URL http://www.blender.org.

[19] R.A. Crowther, R. Henderson, and J.M. Smith. Mrc image processing programs. *Journal of Structural Biology*, 116(1):9–16, 1996. ISSN 1047-8477. doi: https://doi.org/10.1006/jsbi.1996.0003. URL https://www.sciencedirect.com/science/article/pii/S1047847796900039.

[20] CROWTHER RA, DEROSIER DJ, and KLUG A. Reconstruction of a Three-Dimensional Structure From Projection and Its Application To Electron Microscopy. 317(1530):319–340, 1970. ISSN 0080-4630. doi: 10.1098/rspa.1970.0119.

[21] Tim R. Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M. Tomczak. Hyperspherical Variational Auto-Encoders. 2018. URL http://arxiv.org/abs/1804.00891.

[22] Louis de Broglie. Recherches sur la théorie des Quanta Louis De Broglie. page 109, 1924. URL https://tel.archives-ouvertes.fr/tel-00006807.

[23] J. Dubochet and A.W. McDowall. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *Journal of Microscopy*, 124(3):3–4, dec 1981. ISSN 00222720. doi: 10.1111/j.1365-2818.1981.tb02483.x.

[24] J. Dubochet, J. Lepault, R. Freeman, J. A. Berriman, and J.-C. Homo. Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy*, 128(3):219–237, dec 1982. ISSN 00222720. doi: 10.1111/j.1365-2818.1982.tb04625.x.

[25] J. J. Fernández, D. Luque, J. R. Castón, and J. L. Carrascosa. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of Structural Biology*, 164 (1):170–175, 2008. ISSN 10478477. doi: 10.1016/j.jsb.2008.05.010.

[26] Joachim Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 2nd edition, 2006. ISBN 9780195182187.

[27] Carl Friedrich Gauß. Theoria interpolationis methodo nova tractata, 1866.

[28] P. F. Gilbert. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. II. Direct methods. *Proceedings of the Royal Society of London. Series B. Biological sciences*, 182(66):89–102, 1972. ISSN 09628452. doi: 10.1098/rspb.1972.0068.

[29] Krzysztof M. Gorski, Benjamin D. Wandelt, Frode K. Hansen, Eric Hivon, and Anthony J. Banday. The HEALPix Primer. CM(2), 1999. URL `http://arxiv.org/abs/astro-ph/9905275`.

[30] David Haselbach, Jil Schrader, Felix Lambrecht, Fabian Henneberg, Ashwin Chari, and Holger Stark. Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature Communications*, 8(May):1–8, 2017. ISSN 20411723. doi: 10.1038/ncomms15578. URL `http://dx.doi.org/10.1038/ncomms15578`.

[31] M Van Heel, W Keegstra, W Schutter, and EJF Van Bruggen. Arthropod hemocyanin structures studied by image analysis, 1982.

[32] Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11): 780–782, 1994.

[33] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):18037–18041, 2013. ISSN 00278424. doi: 10.1073/pnas.1314449110.

[34] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β -VAE : L EARNING B ASIC V ISUAL C ONCEPTS WITH A C ONSTRAINED V ARIATIONAL F RAMEWORK. pages 1–22, 2017.

[35] Geoffrey Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 2003. ISSN 10495258.

[36] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 08936080. doi: 10.1016/0893-6080(89)90020-8.

[37] Lewis E. Kay. NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance*, 213(2):477–491, 2011. ISSN 10960856. doi: 10.1016/j.jmr.2011.09.009.

[38] James Keeler. Understanding nmr spectroscopy. 01 2013.

[39] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. 2013. ISSN 1312.6114v10. doi: 10.1051/0004-6361/201527329. URL `http://arxiv.org/abs/1312.6114`.

[40] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019. ISSN 19358245. doi: 10.1561/2200000056.

[41] Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/bc7316929fe1545bf0b98d114ee3ecb8-Paper.pdf`.

[42] Jan-Martin Kirves. New Algorithms for Single Particle Cryo Electron Microscopic Image Processing. 2014.

[43] Max Knoll and Ernst Ruska. Das elektronenmikroskop. *Zeitschrift für physik*, 78(5):318–339, 1932.

[44] Mark A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991. ISSN 15475905. doi: 10.1002/aic.690370209.

[45] Gregory M. Kurtzer, Vanessa Sochat, and Michael W. Bauer. Singularity: Scientific containers for mobility of compute. *PLoS ONE*, 12(5):1–20, 2017. ISSN 19326203. doi: 10.1371/journal.pone.0177459.

[46] Felix Lambrecht. Computational methods for the structure determination of highly dynamic molecular machines by cryo-EM submitted by. 2018.

[47] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[48] Mario Lüttich. *Analytische Methoden zur hochauflösenden Strukturbestimmung in der Kryo-Elektronen-Mikroskopie.* 2007. 86 S., Ill., graph. Darst.

[49] Satya P. Mallick, Bridget Carragher, Clinton S. Potter, and David J. Kriegman. ACE: Automated CTF estimation. *Ultramicroscopy*, 104(1):8–29, 2005. ISSN 03043991. doi: 10.1016/j.ultramic.2005.02.004.

[50] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[51] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve Restricted Boltzmann machines. In *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 2010.

[52] Nguyen Phuoc Nguyen, Ilker Ersoy, Jacob Gotberg, Filiz Bunyak, and Tommi A. White. DRPnet: automated particle picking in cryo-electron micrographs using deep regression. *BMC Bioinformatics*, 22(1):1–28, 2021. ISSN 14712105. doi: 10.1186/s12859-020-03948-x.

[53] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–12, oct 2004. ISSN 0192-8651. doi: 10.1002/jcc.20084.

[54] Elad Plaut. From Principal Subspaces to Principal Components with Linear Autoencoders. pages 1–6, 2018. URL `http://arxiv.org/abs/1804.10253`.

[55] Johann Radon. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, dec 1986. ISSN 0278-0062. doi: 10.1109/TMI.1986.4307775.

[56] A. Rose. Television pickup tubes and the problem of vision. volume 1 of *Advances in Electronics and Electron Physics*, pages 131–166. Academic Press, 1948. URL `https://www.sciencedirect.com/science/article/pii/S0065253908611026`.

[57] Peter B. Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003. ISSN 00222836. doi: 10.1016/j.jmb.2003.07.013.

[58] Arthur L Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal*, pages 210–229, 1959.

[59] Sjors H W Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012. ISSN 10478477. doi: 10.1016/j.jsb.2012.09.006.

[60] Sjors H.W. Scheres and Shaoxia Chen. Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, 9(9):853–854, 2012. ISSN 15487091. doi: 10.1038/nmeth.2115.

[61] O Scherzer. Sphärische und chromatische korrektur von elektronenlinsen. *Optik*, 2:114–132, 1947.

[62] Lukas Schulte. New Computational Tools for Sample Purification and Early-Stage Data Processing in High-Resolution Cryo-Electron Microscopy. page 164, 2018.

[63] Alexander Sergeev and Mike Del Balso. Horovod: fast and easy distributed deep learning in TensorFlow. (September), 2018.

[64] Claude E. Shannon. Commnunication theory in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. ISSN 0018-9219.

[65] Kashish Singh, Benjamin Graf, Andreas Linden, Viktor Sautner, Henning Urlaub, Kai Tittmann, Holger Stark, and Ashwin Chari. Discovery of a Regulatory Subunit of the Yeast Fatty Acid Synthase. *Cell*, 180(6):1130–1143.e20, 2020. ISSN 10974172. doi: 10.1016/j.cell.2020.02.034.

[66] Søren Kaae Sønderby, Casper Kaae Sønderby, Lars Maaløe, and Ole Winther. Recurrent Spatial Transformer Networks. pages 1–15, 2015. ISSN 1087-0156. doi: 10.1038/nbt.3343. URL `http://arxiv.org/abs/1509.05329`.

[67] C. O.S. Sorzano, L. G. De La Fraga, R. Clackdoyle, and J. M. Carazo. Normalizing projection images: A study of image normalizing procedures for single particle three-dimensional electron microscopy. *Ultramicroscopy*, 101(2-4):129–138, 2004. ISSN 03043991. doi: 10.1016/j.ultramic.2004.04.004.

[68] K. N. Trueblood, H. B. Bürgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli, and S. C. Abrahams. Atomic displacement parameter nomenclature report of a subcommittee on atomic displacement parameter nomenclature. *Acta Crystallographica Section A: Foundations of Crystallography*, 52(5):770–781, 1996. ISSN 01087673. doi: 10.1107/S0108767396005697.

[69] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[70] M van Heel and G Harauz. Exact filters for general geometry three dimensional reconstruction, 1986.

[71] Marin Van Heel. Similarity measures between images. *Ultramicroscopy*, 21(1):95–100, 1987. ISSN 03043991. doi: 10.1016/0304-3991(87)90010-6.

[72] Marin Van Heel. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987. ISSN 03043991. doi: 10.1016/0304-3991(87)90078-7.

[73] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005. ISSN 10478477. doi: 10.1016/j.jsb.2005.05.009.

[74] Guido van Rossum. Python tutorial, May 1995. *CWI Report CS-R9526*, (CS-R9526):1–65, 1995.

[75] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, Dennis Quentin, Daniel Roderer, Sebastian Tacke, Birte Siebolds, Evelyn Schubert, Tanvir R. Shaikh, Pascal Lill, Christos Gatsogiannis, and Stefan Raunser. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1):1–13, 2019. ISSN 23993642. doi: 10.1038/s42003-019-0437-z.

[76] Feng Wang, Huichao Gong, Gaochao Liu, Meijing Li, Chuangye Yan, Tian Xia, Xueming Li, and Jianyang Zeng. DeepPicker: a Deep Learning Approach for Fully Automated Particle Picking in Cryo-EM. *Journal of Structural Biology*, 2016. ISSN 10478477. doi: 10.1016/j.jsb.2016.07.006.

[77] David B. Williams and C. Barry Carter. The Transmission Electron Microscope. In *Transmission Electron Microscopy*, pages 3–22. Springer US, Boston, MA, 2009. ISBN 9780387765006. doi: 10.1007/978-0-387-76501-3_1.

[78] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! 2020. doi: 10.1101/2020.05.21.106740.

[79] Kai Zhang. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1):1–12, 2016. ISSN 10958657. doi: 10.1016/j.jsb.2015.11.003.