



|              |                                                                                            |
|--------------|--------------------------------------------------------------------------------------------|
| Title        | Volume segmentation of protein electron density maps with 3D convolutional neural networks |
| Author(s)    | Godó, Ákos                                                                                 |
| Citation     | 大阪大学, 2023, 博士論文                                                                           |
| Version Type | VoR                                                                                        |
| URL          | <a href="https://doi.org/10.18910/91983">https://doi.org/10.18910/91983</a>                |
| rights       |                                                                                            |
| Note         |                                                                                            |

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## Abstract of Thesis

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Name ( Ákos Godó )                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                       |
| Title                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Volume segmentation of protein electron density maps with 3D convolutional neural networks<br>(3D-CNN による蛋白質電子密度マップのボリューム画像セグメンテーション) |
| <p><b>Abstract of Thesis</b></p> <p>Can state-of-the-art computer vision methods be applied for protein structure determination tasks? This thesis answers the above question by defining the problem of locating and classifying amino acid (AA) residues in protein electron density (ED) maps as an image segmentation problem.</p> <p>Proteins are organic macro-molecules whose function depends on their 3D structure. By determining their structure, the information may be used in computational simulations for drug discovery, bioinformatics and computational biology.</p> <p>ED maps are three-dimensional, volumetric data acquired via an imaging method called X-ray Diffraction (XRD). ED maps can be considered a 3D image of the protein molecule's shape.</p> <p>Two limitations of protein structure determination are the dependency on sample resolution and processing time. Automatic model building toolkits have difficulties below 3Å resolution which necessitates the use of high-resolution ED maps to obtain reliable results. Unfortunately, obtaining high-resolution ED maps is difficult, contributing to a low, approx. 5% success rate of model building efforts.</p> <p>Model building toolkits create structural models by mapping the AA sequence into the ED map by matching and merging chain fragments. This is computationally intensive: the process may take hours to days. At lower resolutions AAs are less discernible, leading to incomplete models or the outright inability to handle low resolution ED maps.</p> <p>The thesis proposes two 3D convolutional neural network (CNN) architectures to overcome the above limitations. CNNs excel at parallel computations and effectively utilize the parallel computing capabilities of modern GPUs, which by itself leads to improved computational times. The proposed methods assign AA labels directly to ED maps by crop-stitching fixed size windows, only observing the densities without relying on the AA sequence. Thus, the computational time scales linearly with the volume of the ED map instead of the number of residues in the AA chain, resulting in further reduced computational time.</p> <p>A custom data set with annotations for supervised semantic and instance segmentation is created to train and evaluate network performance. It contains high-, mid- and low-resolution ED maps generated from the same approximately 7000 proteins, allowing the CNNs to specifically target each resolution. With low-resolution data in the challenging sub-3Å range, the CNNs can be trained to be functional even where state-of-the-art toolkits fail.</p> <p>A real-life data set from experimental XRD observations of over 500 protein samples is also established. These samples are used to fine-tune and measure the practical, real-life performance of the proposed CNNs.</p> <p>The first proposed CNN architecture, 3D FC-DenseNet matches the performance of state-of-the-art toolkits at high resolutions, and beats them at medium and especially at low resolutions, which are a failure case for toolkits. Its upgrade, MT-StackNet is capable of simultaneous semantic and instance segmentation of ED maps and, due its improved performance, it outperforms model building toolkits even at high resolutions. It can locate and classify all AA instances with a single forward pass using a novel offset vector regression technique.</p> <p>Both methods process ED maps with better performance and much faster than model building toolkits, showing that computer vision methods and CNNs are applicable for protein structure determination tasks.</p> |                                                                                                                                       |

## 論文審査の結果の要旨及び担当者

| 氏 名 ( Ákos Godó ) |     |     |       |
|-------------------|-----|-----|-------|
|                   | (職) | 氏 名 |       |
| 論文審査担当者           | 主 査 | 教授  | 八木 康史 |
|                   | 副 査 | 教授  | 伊野 文彦 |
|                   | 副 査 | 教授  | 長原 一  |
|                   | 副 査 | 教授  | 中川 敦史 |

**論文審査の結果の要旨**

本論文では、3次元ボリウムデータとして表現されるタンパク質の電子密度マップから、アミノ酸残基の位置と種類を推定することを目的とした、タンパク質立体構造決定のための深層学習手法を提案している。

第一章では、「生命の構成要素」であるタンパク質の構造を解き明かすことの重要性について述べている。タンパク質構造決定において一般的に用いられるイメージング技術として、X線結晶構造解析が挙げられる。X線結晶構造解析によって得られる電子密度マップは3次元ボリウムデータとして表現されるが、試料（結晶）作製の困難さから、高分解能のデータを得ることは一般に難しい。低分解能データに対して既存手法の信頼性は低く、あるいは機能しない場合もある。以上のような背景の下、本研究の意義について論じている。

第二章では、関連研究として、タンパク質構造決定のための手法やツールキット、3次元ボリウムデータの分割のためのニューラルネットワーク手法、および構造決定のためのニューラルネットワーク手法について述べている。

第三章では、電子密度マップからアミノ酸残基の位置と種類を特定するという課題の定式化について述べている。本課題は①電子密度マップの各ボクセルにアミノ酸を表すラベルを割り当てる、②各ボクセルが属すアミノ酸の中心座標を指すベクトルを推定する、③2つの結果を統合して、同一のアミノ酸に属すボクセルの集合に電子密度マップを分割する、という3つの目的に分割できる。

第四章では、データセットの構築方法について述べている。Protein Data Bank には実験によって決定された構造データが蓄積されており、その公開データを利用してアミノ酸残基の位置と種類が結び付けられた電子密度マップを生成した。

第五章では、上記①の目的のために DenseNet 構造を改良した 3D FC-DenseNet と目的関数、および実験について述べている。3D FC-DenseNet の構成要素は Dense Block であり、ブロック内の各層の出力を最終層の出力に連結することで勾配消失問題を避け、モデルの訓練が可能になる。アミノ酸残基の一部のみを含む部分データは異なるアミノ酸同士の混同を招くため、重み付きの損失関数を提案した。提案手法は比較手法の性能を上回り、特に低解像度データに対して大幅に向上した。

第六章では、上記②の目的のためにさらなる改良を加えた MT-StackNet と目的関数、および実験について述べている。MT-StackNet は①と②の2つのタスクを同時に扱う構造となっており、また Residual Block への変更などを行った。また MT-StackNet の訓練を可能にする、複数の損失を統合した損失関数を定義した。提案手法は高い精度でアミノ酸残基の位置と種類の推定が可能であることを示した。また、より複雑な課題に対応する必要があるにも関わらず、①の目的において 3D FC-DenseNet を上回る性能を達成した。

第七章では、本研究を総括し、今後の展望について述べている。

以上により、本論文は博士（情報科学）の学位論文として価値のあるものと認める。