Machine learning tool to understand living cells and identify new drug treatments for diseases

Cryo-electron tomography (cryo-ET) is an imaging tool that researchers use to see proteins close-up and in 3D. This helps them understand life in our cells and develop new drugs to fight many kinds of disease. But processing cryo-ET images is time consuming and difficult.

Researchers at STFC Scientific Computing, the Rosalind Franklin Institute and The Alan Turing Institute have developed Affinity-VAE, a powerful machine learning tool with the potential to automate cryo-ET image processing, saving time, money, and energy. The work was supported by the Ada Lovelace Centre (ALC).

The Challenge

To develop new drugs to treat disease, researchers need to understand what is going on inside the cells of all living things - from plants to animals and humans. And they need a fast, efficient, way to do this. Cryo-ET uniquely allows proteins to be seen inside their natural cell environment at incredibly high resolution which reveals how they interact with each other.

Cryo-ET works by taking pictures of proteins inside a cell from various angles. Several images of the same type of protein are then carefully combined to create a high-resolution 3D representation of the protein.

Before the individual images can be pieced together, there are two labour-intensive steps that currently create a bottleneck for image processing, slowing down research and using up resources.

First, researchers analyse the images and identify proteins of the same type. This is usually done manually, which is very time-consuming. The images contain many thousands of proteins, so selecting proteins of the same type from a grainy photo is very difficult, and it's easy to make mistakes.

Next, researchers determine the orientation of each identified protein within the cell. This allows the images to be pieced together in the right way.

I believe Cryo-ET enables the next generation of structural biology – there is great interest in looking at proteins inside the cell to understand life at the molecular level."
Dr Tom Burnley, Molecular and Cellular Microscopy Group Leader,
STFC Scientific Computing

"Traditionally, a computer algorithm uses a 'brute force' approach to test different potential orientations of the protein, which is like trying every combination on a padlock to unlock it! This process is not very efficient and uses lots of computer time and energy."
Dr Jola Mirecka, Senior Computational Scientist, STFC Scientific Computing
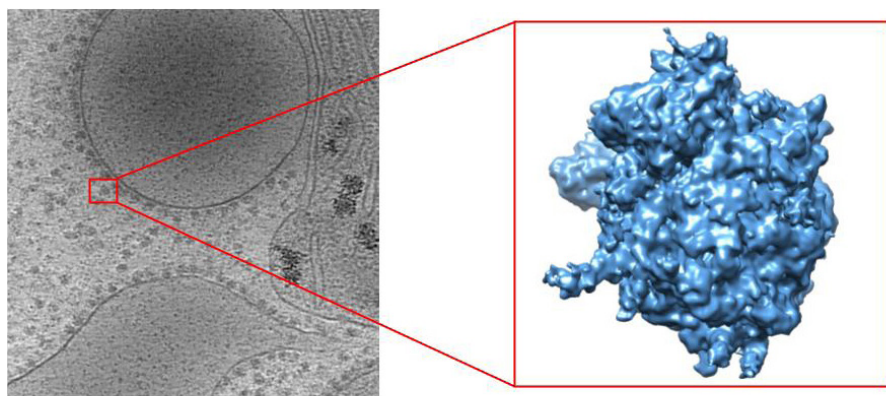
Left: Image of cell components in pancreatic cells of mice: the type of image that Affinity-VAE would analyse.

Right: 3D reconstruction of a cell component (called a ribosome) which would be possible after analysing the image with Affinity-VAE.

Image credit: Freyberg group

Our Approach

Affinity-VAE is a new machine learning model which automates both stages of identifying proteins and finding their orientation.

Affinity-VAE belongs to a group of machine learning tools called 'variational auto-encoders' (VAEs). In contrast to traditional VAEs, Affinity-VAE is very good at identifying and grouping proteins of the same shape thanks to the injection of some prior biological knowledge. This is a welcome alternative to manual identification of proteins and brute force alignment.

Affinity-VAE can also determine the rough orientation of the protein in the image. This makes it much easier to find the exact orientation of the protein: imagine trying to open the combination padlock but you already know the first numbers in the sequence, so you only need to find the final number.

The Benefits

Affinity-VAE eliminates the need for researchers to spend valuable time on tedious, repetitive, and difficult tasks. It could also lead to fewer mistakes in selecting proteins, which results in higher resolution 3D images. This means it could improve the efficiency of expensive microscope time.

Affinity-VAE also provides a more efficient approach to finding the orientation of proteins, which reduces the use of computer resources and saves energy.

"Researchers are spending a huge amount of time manually processing data from cryo–ET, and Affinity could free up this time, speed up research, and save a lot of money."
Dr Jola Mirecka, Senior Computational Scientist, STFC Scientific Computing

"What's exciting is that Affinity doesn't just find the proteins we know about – it also helps reveal ones we weren't aware of."
Dr Mark Basham, Science Director and Challenge Lead, Rosalind Franklin Institute.

The model will advance understanding of life inside our cells and even help us discover new and unknown proteins.

Overall, Affinity-VAE could save huge amounts of time, money, and energy for cryo-ET research, maximise the valuable data that can be extracted from images, and boost scientific discovery.

The Future

The developers of Affinity-VAE are currently writing a subsequent publication which includes extensive experiments on real biological data.

Affinity-VAE is a flexible model: it has been applied to cryo-ET but has the potential to be used in any field where classifying and determining orientation is needed, such as astrophysics.

Read the full paper: arxiv.org/pdf/2209.04517

Contact jola.mirecka@stfc.ac.uk for more information

Dr Mirecka's work on Affinity-VAE was funded by the Ada Lovelace Centre.

Written and Designed by Esme Mirzoeff, STFC Scientific Computing Communications and Impact team

UKRI Science and Technology Facilities Council
Ada Lovelace Centre

The Rosalind Franklin Institute

The Alan Turing Institute