Deriving structure and heterogeneity from microscopy data

Susan Cox

Randall Centre for Cell and Molecular Biophysics
King College London

Good quality information about sample structures often requires combining information from multiple instances of the structure. In techniques such as X-ray crystallography and neutron scattering, this is achieved as part of the technique. In biology, a major recent advance was the ability to average over multiple electron microscopy images of a structure, allowing the structures of proteins to be imaged.

Larger cell structures are usually imaged using fluorescence microscopy, which allows us to specifically label different chemical components of the cell, and to image live cells (which is not possible using electron microscopy). Analysis of these images is usually performed by segmenting areas of the images thought to be significant and then fitting a mathematical model to the data. To accumulate data usually the same type of structure (e.g. a nucleus) will be imaged many times, and the data from different observations combined.

Here we take a different approach. We model the entire dataset (consisting of many instances of the same structure) as arising from a base structure with fluorophores at certain positions. The base structure can be at a different angle in each image, and a deep learning approach is taken to optimise this. We also allow the structure to vary, for example by repeating, stretching, changing size, or deforming in some other way. By taking this approach we are able to derive both the structure and its heterogeneity (how it varies across many instances of the structure).

While we have focused on biological applications, the general principle of combining data from multiple instances of a measurement is a general problem and some of our approaches may have applications elsewhere.