Summary: Deep learning for cellular image analysis

Authors of the article talk about how advancements in computer vision and deep learning have transformed research in biological sciences. This has led to an increase in demand of image analysis techniques in biosciences, such as unsupervised image exploration, image classification, image segmentation and object tracking. Additionally, advancements in hardware and new deep learning frameworks have unlocked these techniques for software developers. However, amount of data required for deep learning and the knowledge of model performance optimization and model out interpretation is paramount for harnessing the full power of these tools and life scientists must familiarize themselves with them.

Paper further discusses about the mechanics of deep learning, importance of training data and different frameworks and strategies for creating robust models. After the training data have been gathered, frameworks such as PyTorch, Caffe and MXNet etc. can be used for training deep learning models. Only training the model is not sufficient, as one needs to experiment with model architecture, hyperparameter tuning to avoid overfitting issues and dealing with small datasets or class imbalances. Once trained, these models need to be containerized and deployed to be accessible for wide range of users. TensorFlow, MXNet frameworks have built in deployment features.

Four primary uses of deep learning in the context of biological applications are tasks such as image classification, image segmentation, object tracking and augmented microscopy.

Classification applications include identifying structures in cell, conditions and compounds that lead to cell morphology, changes in cell state and image activated cell sorting. Architectures used for biomedical image classification tasks are similar to those models used in commercial application. Due to the lack of large amount of annotated biological training datasets, transfer learning techniques are used such as tweaking the final layer of large pre-trained classifiers, typically trained on ImageNet dataset.

Another very important application is image segmentation tasks which involves identifying meaningful objects in the image, such as identifying a single cell or different parts of a cell structure. Semantic segmentation techniques are used to partitioning image pixels into different classes they belong to, whereas instance segmentation is used for identifying each instance of a class in the image. Model architectures such as U-Net and DeepCell are traditionally used for segmentation tasks, but object detection-based methods are also used to create new segmentation architectures such as R-CNN, Faster R-CNN, Retinanet and mask R-CNNs. Deep learning has also enabled the ability to segment overlapping objects using vector embedding and unsupervised clustering on the embedding space.

Object tracking is the task of following objects through a series of time-lapse images by object detection and object linkage. Object detection is done using nearest-neighbor search, state-space models, linear programming, or segmentation. Complex behaviors such as objects like splitting, disappearing can be handled using linear programming. This approach is used in software like TrackMate and CellProfiler. Deep learning has recently been applied to follow diffraction-limited particles, animals, and cells. While there is a scarcity of training data in this area, software programs for curation are helping to alleviate the problem. Most of the previous work has been on object detection, owing to a scarcity of training data; the 3D aspect of object tracking makes training data more difficult to come by.

The extraction of latent information from biological images, such as the identification of the positions of cellular nuclei in bright-field images, is known as augmented microscopy. Although technologies like as phase-contrast microscopy and differential interference contrast microscopy can yield data, extracting that data is time-consuming. This was addressed by scientists by converting it to a problem of supervised learning.

Although the application of deep learning to biological image analysis is still in its early days, there has already been remarkable progress in adapting deep learning to biological discovery.