Development of an Artificial Neural Network for Semantic Segmentation of Nuclei and Mitochondria in Transmission Electron Microscopy and Serial Block Face Scanning Electron Microscopy Images

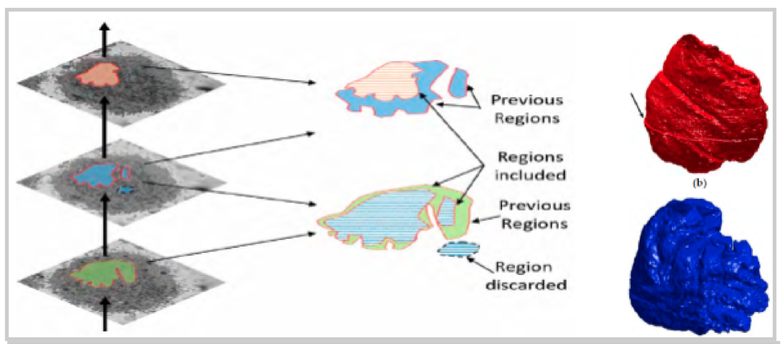
OBJECTIVES

The goal of this study is to develop an artificial neural network with Residual U-Net architecture using the Keras/Tensorflow platform, for the purpose of automating the identification of the nuclear envelope in mammalian cells.

INTRODUCTION

The field of 'Volume Electron Microscopy' (VEM), a collective term for Electron Microscopy (EM) techniques aimed at analyzing 'large volumes', is undergoing a new renaissance thanks to the prospect of applying the tools of Deep Learning to volume reconstructions of cellular and sub-cellular components. VEM can be performed using Transmission (TEM) or Scanning (SEM) electron microscopy. Traditionally, serial section TEM (ssTEM) has been the tool of choice for 3D ultra-structural investigation of biological material. Nowadays, serial section SEM (ssSEM) is increasingly used as an alternative to ssTEM. In its most popular implementation, ssSEM encompasses two complementary techniques: Serial Blockface SEM (sbfSEM) and Focused Ion Beam SEM (fibSEM). In these two methods sections are repeatedly removed and discarded from a block of specimen surface inside the electron microscope, either *via* a diamond knife (sbfSEM) or *via* a focused ion beam (fibSEM). For example, sbfSEM has been used for the 3D reconstruction of liver cell nuclei. fibSEM has been used in a number of studies, including the segmentation and reconstruction of cellular components within an entire yeast cell, and the modeling of liver cells mitochondria.

Once serial images from a block of biological material are derived from sbfSEM or fibSEM, restoring continuity to the serial image stack is essential for fully understanding the specimen. Until recently, analysis of EM images was carried out by specialists with experience in the identification of biological features in the complex grayscale world of EM. For example the reconstruction of the 3D shape of nuclei from serial sections would require an experienced microscopist to manually generate a '2D mask' of the nuclei in each imaged slice, followed by stacking all the 2D masks together.



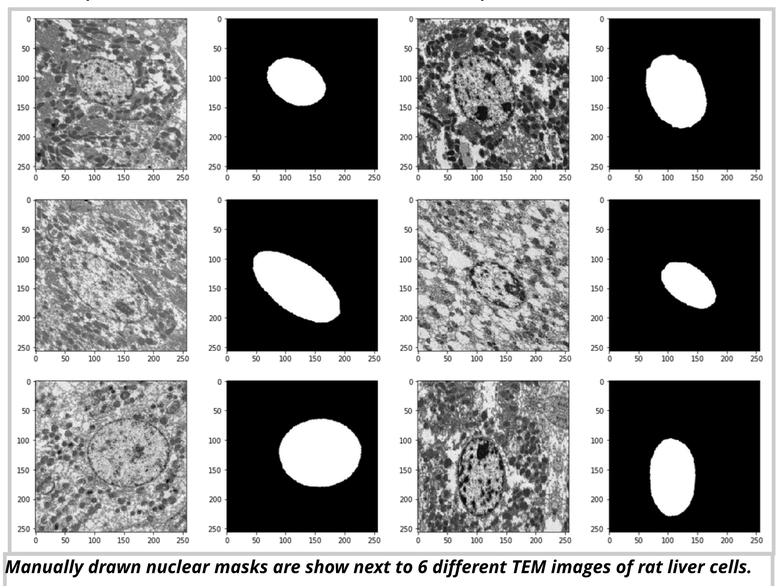
Stack of images showing one nucleus (left) and the 3D reconstruction of two nuclei (right).

However, it has been calculated that this operation would require over 2000 years for a single person to reconstruct the 3D structure of 1 cubic millimeter of mammalian tissue. This hypothetical example exemplifies the need for an automated identification of subcellular components within each imaged slice, a process called semantic segmentation.

METHODOLOGY

1. Image pre-processing

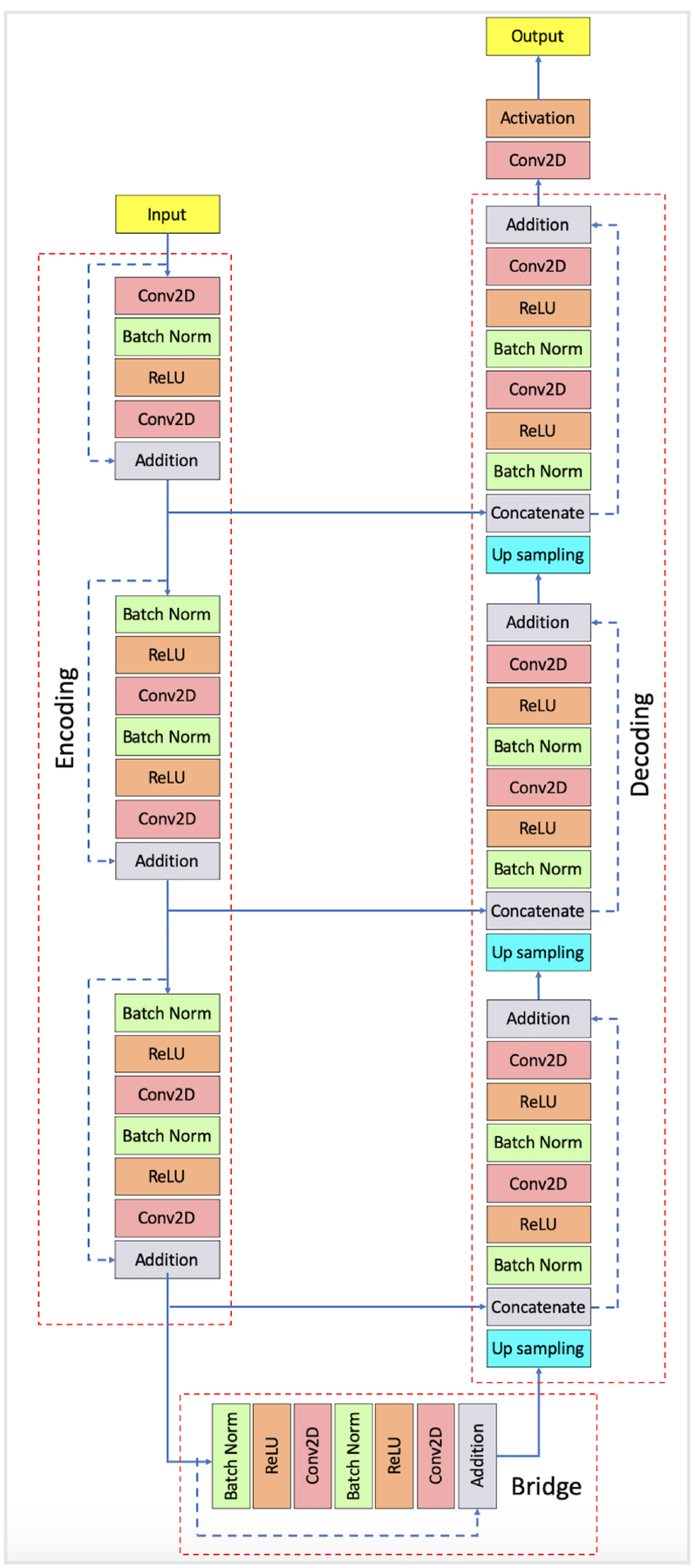
Nuclear masks in 21 TEM images of rat liver cells were drawn manually using Matlab *Image Segmenter* (below). Paired images and masks were then split into a training set of 15 pairs and a validation set of 4 pairs .



Manually drawn nuclear masks are show next to 6 different TEM images of rat liver cells.

2. Training

Augmented images (of dimensions 256 x 256) are then fed as *inputs* to a Res-U-net. The network designed in this study is comprised of three parts: *encoding*, *bridge*, and *decoding*. The first part encodes the input image into a compact representation. The last part recovers the representations to a pixel-wise categorization, i.e. a semantic segmentation. The middle part acts like a bridge connecting the encoding and decoding paths. All of the three parts are built with residual units. The encoding path has three residual units. In each unit, instead of using pooling operation to down-sample the feature map size, a stride of 2 is applied to the first convolution block to reduce the feature map by half. The decoding path consists also of three residual units. Before each unit, there is an up-sampling of feature maps from lower level and a concatenation with the feature maps from the corresponding encoding path. After the last level of decoding path, a 1x1 convolution and a *sigmoid* activation layer are used to project the feature maps into a semantic segmentation which is compared with the manually drawn mask of the image. During consecutive optimization cycles the Res-U-Net learns how to minimize the difference between the predicted segmentation mask and the hand-drawn segmentation mask. Training required the refinement of 75,528,113 parameters, and was carried out for 60 epochs, each epoch representing the optimization of network parameters using 30 minibatches of 5 images each.



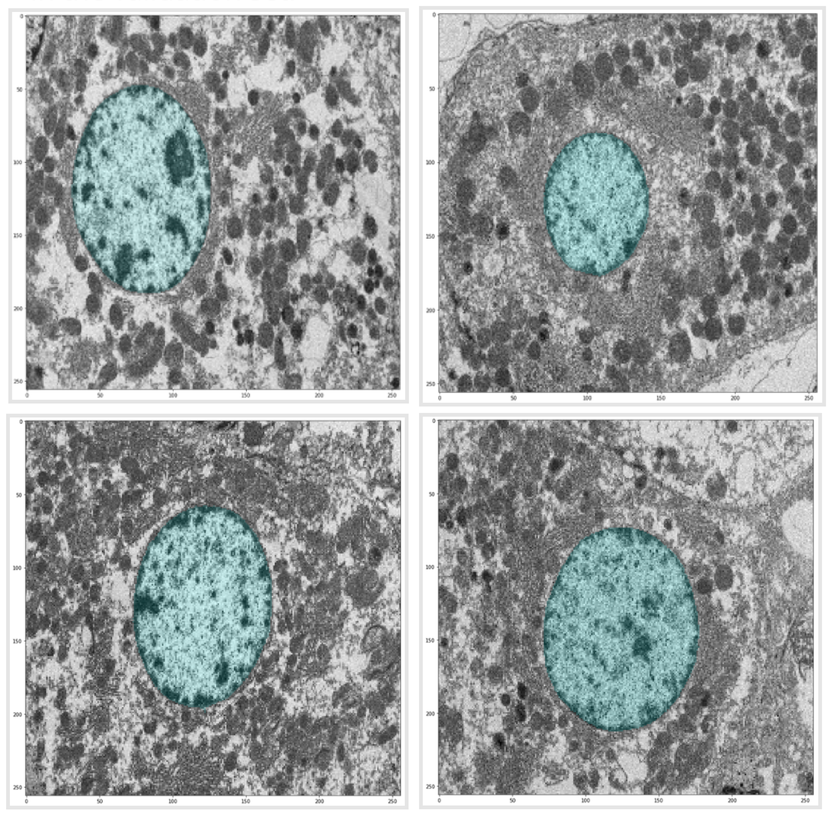
As the loss function (the difference between the predicted and hand-drawn mask), a combination of the cross entropy loss and of the Dice coefficient was used. The dice coefficient has been shown to increase performance over just the cross entropy loss. The Dice coefficient can be generalized to binary masks by the summation of the probabilities in the denominator as shown on the side:

$$D(\hat{y}, y) = \frac{2\sum_i \hat{y}_i y_i + s}{\sum_i \hat{y}_i + \sum_i y_i + s}$$

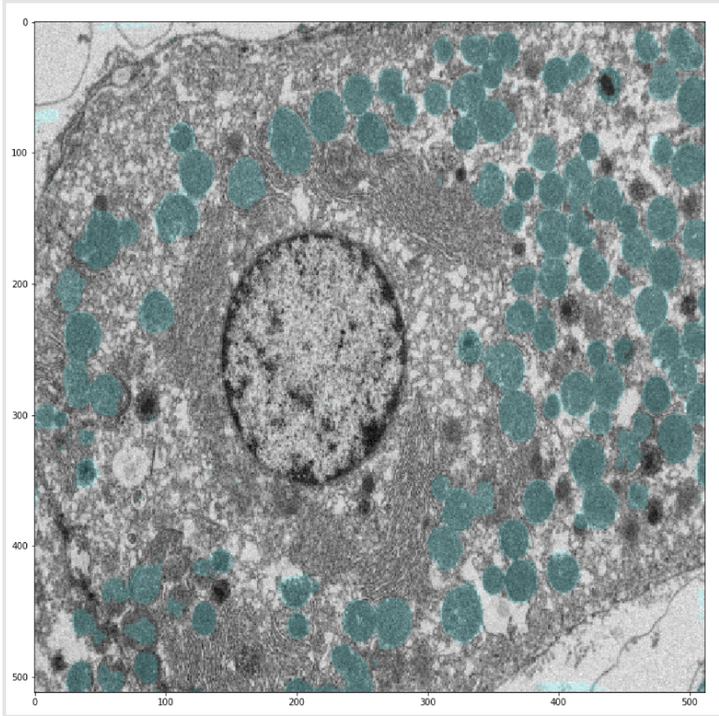
Dice coefficient. Y is the ground truth (hand drawn-mask) and Y-hat is the predicted mask (network output) refined at each iteration.

RESULTS

Once the Res-U-Net training was completed, as attested by the lack of further reduction of the loss in additional refinement cycles, the network was tested by calculating the corresponding nuclear masks for the four TEM images included in the validation set. These mask are shown below in transparent cyan color superimposed on the four images in the validation set.



Initial testing shows that the same network architecture is also effective in identifying the envelope of mitochondria. However, due to the smaller dimensions of these organelles with respect to the nucleus, it is necessary to use larger images of dimensions 512 x 512, and comparatively longer training cycles.



The results illustrated the range of capabilities of the network architecture as well as the efficiency and quickness with which the network was able to identify nuclei and mitochondria. The development of this tool can prove to be extremely valuable for the reconstruction of 3D models of cells. It is much more efficient than current methods of identification and reconstruction and just as accurate, giving pathologists and physiologists an asset that makes their work considerably easier.

CONCLUSION

The goal of this study was to design and train an Artificial Neural Network capable of identifying the nuclear envelope in TEM/SEM images of mammalian cells, for the purpose of reconstructing the 3D structure of nuclei from serial sections. I have accomplished my goal by developing a Residual-U-Network that, upon training, identified correctly the nuclear envelope in TEM images that had not been previously exposed to the network.