Microscopy Cell Nuclei Segmentation with U-Nets

Humza Iqbal

December 19th, 2023

1 Introduction to Problem

In microscopy, cell segmentation poses a fundamental challenge in image analysis, as it involves the precise identification and delineation of individual cells within complex biological structures. Accurate cell segmentation is crucial for various biomedical applications, such as understanding cell behavior, tracking cellular processes, and diagnosing diseases.

There are many difficult problems when processing these images. For example, cells may have compartments like nuclei, which we may not want to reclassify. The position and orientation of cells within images may vary, making it challenging to design handcrafted features that are invariant to these transformations. These problems are solved with Convolution Neural Networks, which have the necessary functional capability and expressivity.

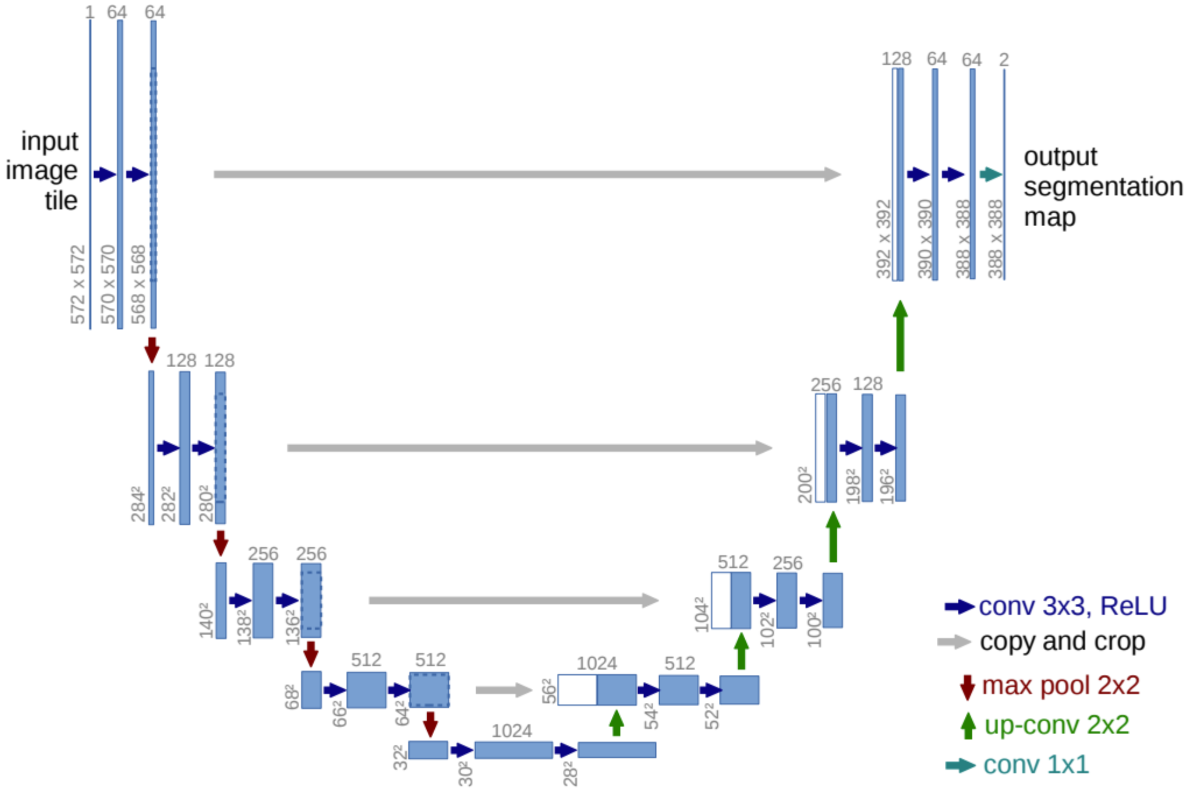


Figure 1: U-net architecture. Each blue box corresponds to a multi-channel feature map. From Ronneberger's, ”*U-Net: Convolutional Networks for Biomedical Image Segmentation*”

To the main point, The central challenge in cell segmentation often revolves around the scarcity of a substantial number of samples. However, U-nets offer a promising solution to address this limitation. As a preprocessing step to use a U-Net, we perform augmentations on the input images, effectively creating a pseudo-bootstrap approach. These augmentations may change the viewpoint, illumination, deformation, or occlusion of the image. With these various augmentations, we have added enough variance to non-trivially increase the number of data points we have. This is something like a pseudo-bootstrap. By doing so, U-nets generate an augmented dataset that significantly expands the available training data.

The augmentation process enhances the robustness of the U-net model by exposing it to a broader range of scenarios and variations in the input data. This not only mitigates the challenge of limited samples but also allows the network to learn more effectively from the augmented dataset. The augmentation-driven pseudo-bootstrap serves as a practical workaround when acquiring an extensive set of original samples is impractical.

Moreover, the effectiveness of U-nets in cell segmentation extends beyond their role in data augmentation. Olaf Ronneberger's comprehensive article, ” *U-Net: Convolutional Networks for Biomedical Image Segmentation*,” delves into various aspects of U-net performance. Ronneberger provides detailed insights into how U-nets leverage their unique architecture to excel in biomedical image segmentation tasks.

2 Technical Details and Formal Description

In this portion we go over some mathematical formulation of a general Convolutional Neural Network, which U-Nets are built on. CNNs are a specialized type of artificial neural network designed for tasks such as image recognition and processing. It excels at automatically learning hierarchical feature representations from input data.

Convolutional Layers utilize convolution operations, where small filters (kernels) slide across the input data, computing dot products at each step. This enables the network to learn local patterns and features. The linearity and commutativity of the convolution operator play crucial roles in this process, ensuring that the order of operations does not affect the final result. Mathematically, the convolution operator can be represented as,

$$(f * g)(t) = \int f(\tau)g(t-\tau)d\tau$$

where f represents the input data, g represents the filter, and t represents the position.

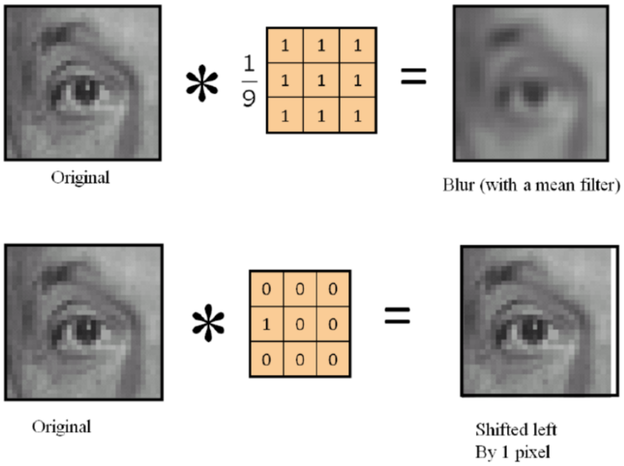


Figure 2: Two different kernels have different effects on the image

CNN's main goal is to promote properties of invariance and equivariance. CNNs exhibit translation invariance, meaning they can recognize patterns regardless of their position in the input.

$$(f * g)(t) = (f * g)(t - t_0)$$

Equivariance is an extension, suggesting that the network's output varies predictably with input transformations. Suppose f and g are related by transformation T. Then we know, that $(f * g)(t)$ and $(Tf * g)(t)$ are related. This property is crucial for tasks where the spatial relationships between features are essential.

Applied after convolution operations, activation functions (e.g., ReLU) introduce non-linearity, enabling the network to learn complex relationships. This nonlinearity is essential for capturing intricate patterns and variations in the data, especially since the convolution operator is linear. The ReLU is defined below,

$$\mathrm{ReLU}(x) = \max(0, x)$$

Pooling layers downsample spatial dimensions, reducing the number of parameters and computational load. Techniques like max-pooling retain the maximum value in local regions, contributing to translation invariance by focusing on the most significant features.

$$\mathrm{MaxPooling}(x, y) = \max_{i \in N(x,y)} f(i)$$

Where $N(x, y)$ is in the set of indices in the local region (x, y).

Typically found in the later stages of a CNN, fully connected layers connect every neuron, enabling the network to make final predictions. We have then the following objective,

$$F(x) = W_L \cdot \mathrm{ReLU}(W_{L-1} \cdot \ldots \cdot \mathrm{ReLU}(W_1 \cdot x + b_1) + b_{L-1}) + b_L$$

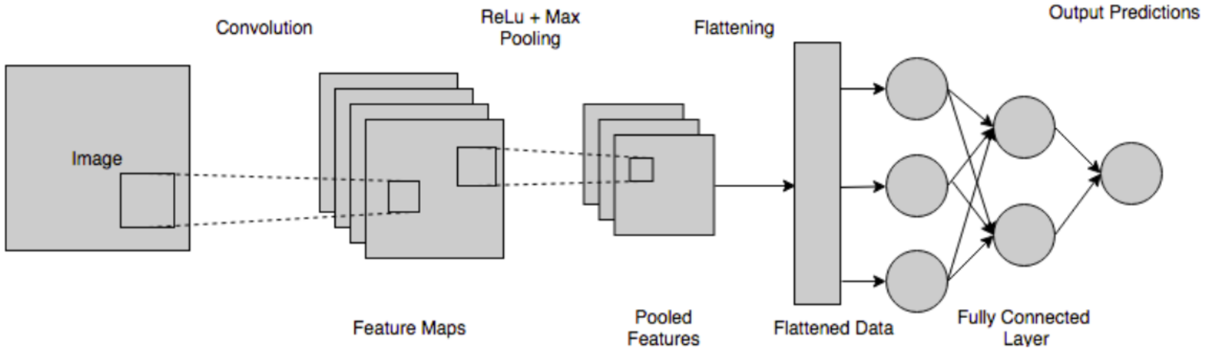


Figure 3: CNN example from *Hands-On Automated Machine Learning by Sibanjan Das, Umit Mert Cakmak*

3 Results and Improvements

The metric used to assess the accuracy of our model is the intersection over union (IoU) metric, which takes a set of proposed pixels and a set of true pixels and calculates,

$$\mathrm{IoU}(A,B) = \frac{A \cap B}{A \cup B}$$

We achieved a final validation accuracy of 83.73% with our undersampled dataset. The following predicted masks were created by our model, with a corresponding true mask and image. From visual inspection, we see the model is correctly classifying.

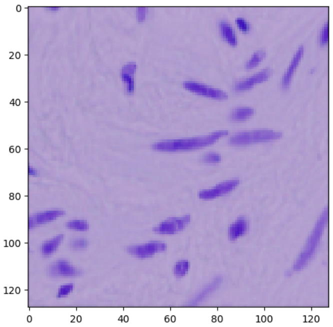
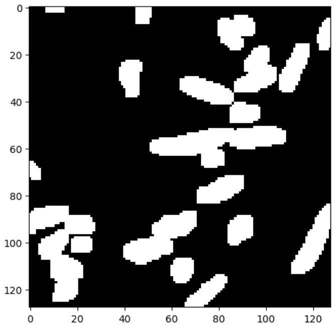
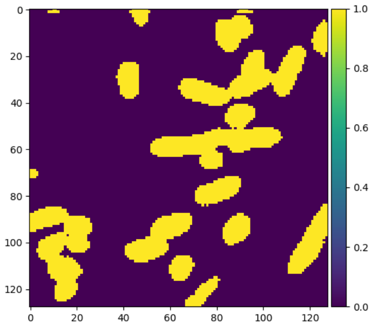
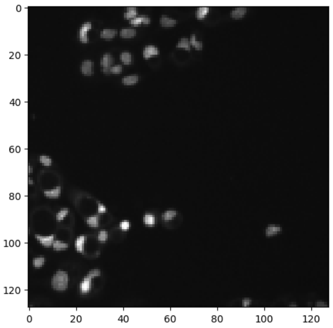
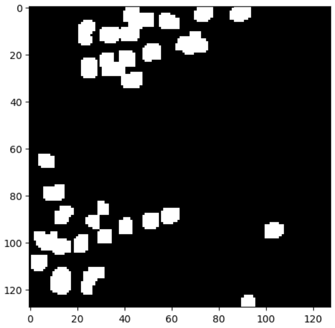
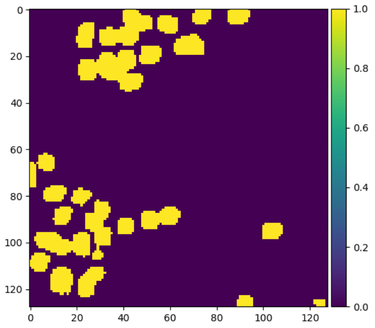
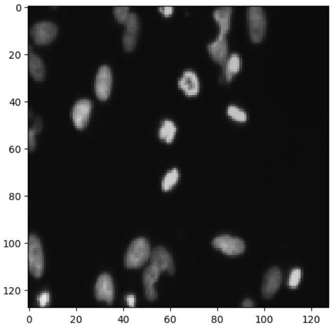
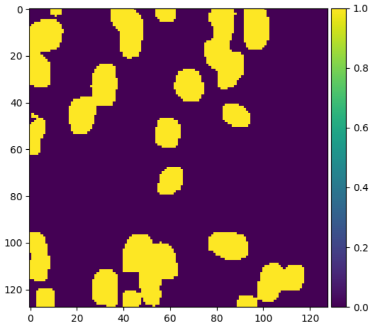
Nr.	Image	True Mask	Predicted Mask
Train			
Valid			
Test		Unknown	

Table 1: Table of Images to Masks

We can make improvements to our model by tuning the hyperparameters, using more of the given training data, and dropout methods. With this increased sophistication we may also add some regularization to our objective or use some ensemble methods.

References

Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18). *U-Net: Convolutional Networks for Biomedical Image Segmentation.* arXiv.org. https://arxiv.org/abs/1505.04597

Code and data Sources:

https://www.kaggle.com/competitions/data-science-bowl-2018

https://www.kaggle.com/code/keegil/keras-u-net-starter-lb-0-277/notebook