

# **Supervised and Unsupervised Cell Segmentation with Deep Learning**

Linh Vo

STUDENT NUMBER: 2030898

THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE  
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE  
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES  
TILBURG UNIVERSITY

Thesis committee:

Dr. L.L. Sharon Ong  
Dr. Emmanuel Keuleers

Tilburg University  
School of Humanities and Digital Sciences  
Department of Cognitive Science & Artificial Intelligence  
Tilburg, The Netherlands  
January 2020



## Preface

The graduation assignment project I had with dr. Sharon Ong was a great chance for me to gain deeper knowledge about deep learning in biomedical image analysis. I express my deepest thanks to Dr. Sharon Ong, who in spite of being extraordinarily busy with her work, still took time to guide me on the correct path and allowed me to carry out my project successfully.

I would like to express my sincere appreciation and gratitude to Dr. Emmanuel Keuleers for giving me valuable feedback to improve my study. His feedback was really important, which provided me with further knowledge and advises to finalize my thesis paper.

I express my deepest thanks to my family and special, unique, amazing friends who have always been supportive to me.

I perceive this graduation project as a big milestone in my life. I will strive to gain skills and knowledge in the best possible way, and I will continue to work on my improvement, in order to attain the desired professional career and life expectation.



# Supervised and Unsupervised Cell Segmentation with Deep Learning

Linh Vo

*Automatic cell segmentation is a challenge in bioinformatics. With the rise of Deep Neural Networks (DNNs), it is expected that analysis in microscopy images would be more precise and automated. In this study, three different DNNs models are investigated and compared for cell nuclei segmentation. Two supervised models (the U-Net and R2U-Net) and an unsupervised model (the W-Net), which are popular in biomedical image analysis, are studied. The results show that deep learning algorithms outperform traditional image process techniques in identifying cell's boundaries. Outputs from supervised methods, U-net and R2U-net, return the highest similarity to the ground truth provided by experts. However, labelled data is necessary to train such models which is quite labour intensive for data collection. On the other hand, the unsupervised model W-net predicts satisfactory results, and it does not require any labelled data during training process. In the long run, future work includes implementing deep learning models on different types of cells acquired from different types of microscopes to obtain the comprehensive picture of deep learning to cell image segmentation.*

## 1. Introduction

People around the world are suffering from multiple diseases including cancer, heart disease, brain tumours or chronic illness. Microscopy imaging is an important work which allows faster assessment for diagnosis and prognosis. Processing microscopy images allow quantitative and qualitative analysis of cells in the human body. Cell segmentation supports researchers to routinely check on many aspects of the diseases. It is the starting point for microscopy analysis, which can help researchers to accurately observe how to react to a cell with suitable treatments. Respectively, researchers could understand the underlying biological process. As a result, the solution leads to better treatment for patients, accelerate both treatment and drug discovery. Indeed, digital image processing plays a big role in decision making in disease, enables the success of computer vision in providing useful information to clinical practice.

In the past, most of cell segmentation works in microscopy images are traditional image processing approaches such as thresholding, morphological operations or watershed algorithms ([J. Gulo, Sementille, and Tavares 2019](#)). These approaches are not only expertise-required and time-expensive but are also difficult to achieve robust, accurate and reliable results. Normally, microscopy images include background clutters, are noisy, blurry and have poor resolutions. The image quality is degraded due to the devices used for image acquisition. Low contrast and weak boundaries images could be complicated for traditional approaches to separate cells from the foreground and the background. In addition, among cells exist significant variation in shape, size and intensity and cells are often partially overlapped one another. The potential ambiguous regions could raise the risk of unsuccessful cells segmentation ([John et al. 2014](#)).

The rapid growth in population and the evolution of diseases have raised a greater demand in automatic cell segmentation. As more biological data becomes accessible, it is the time where big data and biomedical research meet. The developing of big data and computer vision is expected to put biomedical practice and research into a new level of advancement considered not only to automate the task but also to perform near human-level. Automation could save time from manual work, contribute higher production rates and make use of available data efficiently. Thus, a study for advanced and automatic microscopy image processing is crucial and in demand. The motivation of this study insists on providing faster and more efficient analysis compared to the manual techniques currently used by researchers and clinician scientists.

Artificial Neural Networks have rapidly become the chosen algorithms for computer vision. For biomedical image analysis, the methodology is being applied and developed for object detection, segmentation and tracking (Litjens et al. 2017). It is foreseen that the methodology could also add value to microscopy image processing of biological cells. Inspired by the idea, this thesis aims to study on potential neural networks to automate cell segmentation. The U-net, R2U-net and W-net are the selected supervised and unsupervised learning models as they have been shown to be quite promising on the cell segmentation tasks. Accordingly, the main research question is defined as:

*To what extent do deep learning improve cell segmentation in microscopy images?*

In detail, this thesis aim to address the following sub-questions:

1. To what extent does the U-Net, a popular supervised deep learning model for biomedical image analysis improve cell segmentation?
2. To what extent does an adjusted version of the U-Net (an R2U-Net) improve cell segmentation?

Supervised methods require labelled data which is often not readily available in biomedical image analysis. Therefore, the third sub-question is:

3. How well does the W-Net, an unsupervised version of the U-Net model perform in cell segmentation?

In this project, three different deep learning models have been applied to the same cell nuclei segmentation tasks, namely U-Net, R2U-Net and W-Net. The contributions of this paper can be summarized as follows:

1. Compared to the traditional image processing, all three proposed deep learning methods outperformed the baseline methods, Otsu thresholding and morphological watersheds, in separating cells out of image background.
2. Both supervised methods U-Net and R2U-Net achieved promising quantitative and qualitative results, improved more than 20% accuracy compared to the selected baseline methods. Especially, R2U-Net returned the best results with approximately 92% accuracy in term of Dice coefficient. In fact, the R2U-Net model performed better than the U-Net

because the model learns from its output residuals rather than the true distribution of outputs.

3. The unsupervised method W-Net could also perform the same segmentation task with no labelled data required. The results are satisfied with 77% accuracy in Dice coefficient using validation dataset, improved more than 7% to the traditional image processing techniques.
4. The proposed supervised methods require labeled data to train, which is time-consuming and labour-intensive in biomedical data acquisition process. On the other hand, the unsupervised method does not require labelled data but still able to segment the cells.

This paper is organized as follows. Section 2 introduces the related work on microscopy, deep learning in digital microscopy and previous cell segmentation solutions. Some related knowledge about deep learning to be used in the proposed methods is also introduced in this section. Section 3 describes in details the experiment design, the proposed methods and evaluation metrics for this study. While section 4 presents the results obtained, section 5 discusses about output results, some issues this study confronts and some future work in the long run. Finally, the conclusion is derived in section 6 to summarize all the findings in this study.

## 2. Related Work

This study aims to address the use of deep learning to improve cell segmentation over traditional image processing. In this section, an overview of related work is presented. Initially, some brief information about the microscope and how it works are presented. Then, major ongoing research applications in digital microscopy are introduced, which cell classification, segmentation and detection are considered as separate problems. Focusing deeper into automating cell segmentation tasks, a detailed review of available research works in the last few years is presented. This review contributes to the selection of baseline traditional methods and proposed deep learning models to this study.

### 2.1 Background on Microscopy

Optical microscopes are widely used in biology and medicine ([Lindon, Tranter, and Koppenaal 1999](#)). They are often used to capture specific features of small specimens with accuracy and details. They can also used to visually enhance features of subsurfaces at a small scale.

Fluorescent microscopes require more complex and expensive instrumentation than standard microscopes as it uses fluorescence techniques to generate images. These techniques allow the observation of samples which are stained with fluorescent dyes ([Zhang and Hoshino 2019a](#)). A fluorescent microscope normally comprises of the light source, the excitation filter, the dichroic mirror and the emission filter (see Figure 1). When a specimen is placed on the stage, the excitation light that passes through the excitation filter is reflected by the dichroic mirror. Through the emission filter, the specimen is then illuminated with light of different wavelengths yielding a multi-channel 2D image representing different objects or object structures ([Kenneth and Michael 2019](#)). This is the type of microscopy images that are included majorly in the dataset used for this study.

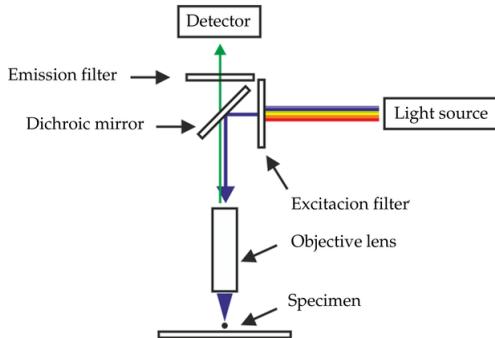


Figure 1: The scheme of the fluorescent microscope (Majtner 2015)

Confocal imaging was developed to overcome the conventional wide-field fluorescence microscopy and generate 3D images. In confocal microscopy, powerful light sources such as laser are used to focus on a pinpoint repeatedly throughout one level to another. As a result, image reconstruction software assembles the multi-level data to form a 3D image of the specimen (Thiel et al. 2019).

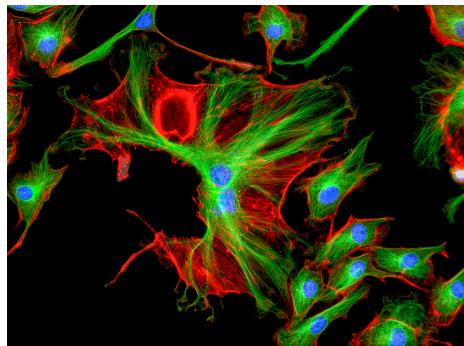


Figure 2: A sample of cells image generated by fluorescent microscopy (Discovery 2018)

Optical microscopy is useful in biology, biomedical and material sciences. For food or drug industry, microscopes could support scientists to assess the presence, structural organization and spatial distribution of components in a product (Jaime A. Rincon Cardona and Herrera 2013). They could also be used for the study of minerals such as coal and graphene oxide (Teichmuller and Wolf 2011). For particular biological imaging, fluorescent microscopy is widely used to capture structural components of cells and extract their genetic materials (DNA and RNA). This application is tremendously useful for conducting viability studies on cells, detecting particles such as neurotransmitter amines or tracing protein and nucleic acids (Zhang and Hoshino 2019b).

## 2.2 Main applications of Image Analysis in Digital Microscopy

Before going deeper into cell segmentation problem, a general overview is provided on the main applications of image analysis in digital microscopy. The field of image analysis is shifting from traditional image processing to machine learning and deep

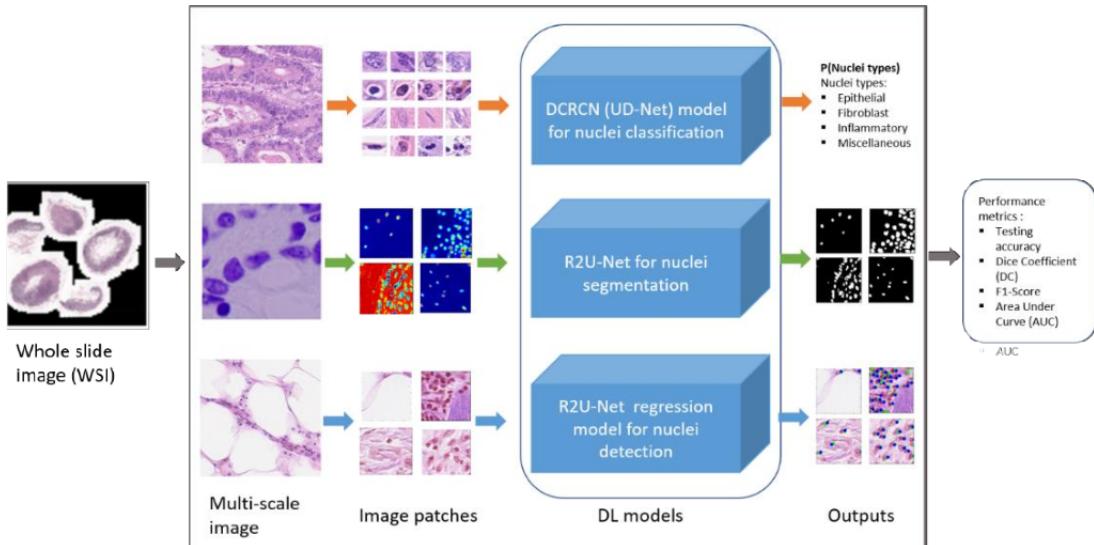


Figure 3: An example of different Deep Learning approaches for cell classification, segmentation and detection ([Alom et al. 2018b](#))

learning neural networks methods. In processing fluorescent microscopy images, digital microscopy has been researched to perform automatic tasks in three domains; cell classification, segmentation and detection.

#### *Cell classification.*

Cell classification is active research field where machine learning is applied to classify different types of cells with the given area in shape and size ([Alom et al. 2018b](#)). Before the deep learning revolution, many research works have been proposed with traditional machine learning approaches. Most of the attempts comprise of separate stages, where hand-crafted features are extracted and fed in to classify models. Most popular extracted image features are intensity histogram, wavelet coefficients, Fourier spectrum or grey-scale co-occurrence matrix ([Soda and Iannello 2009](#)). Indeed, more elaborated features have been introduced to learn the distribution of image micro-patterns and provide useful information for identification. Recently, the rise in DLNNs has led to widespread in Convolutional Neural Networks (CNN) based-methods in cell classification tasks. In some researches, these methods has been proven to outperform the normal hand-crafted algorithms ([Wang et al. 2017](#)). With the benefits of deep learning architecture in eliminating the feature extraction process, the model has advantages by learning directly from training data. One representative work of successful application could be from Oei et al. ([Oei et al. 2019](#)) when the proposed model demonstrated the use of CNN in classifying normal breast epithelial cell line versus two breast cancer cell lines based on images.

#### *Cell segmentation.*

Cell segmentation is the application of machine learning in learning and predicting locations of the cells and their nuclei in an image. ([Alom et al. 2018b](#)). An automatic segmentation model is a model to identify pixels and separate cells boundaries in the

image individually. Over the years, the traditional machine learning approaches mostly involved with manual engineering features that used to identifying pixels independently. Widely used segmentation algorithms include watershed, level-set, k-means, and stochastic image processing (Méndez-Vilas and Díaz 2011). With the development of deep learning, a lot of new models have been proposed using deeper networks for segmentation tasks. For instance, the U-Net model and its variations have been introduced with specific segmentation purposes for medical images (Alom et al. 2018a). As this research is focused mainly on segmenting cells, a detailed literature review on the previous machine learning research in cell segmentation is provided in section 2.3.

#### *Cell detection.*

Automatic cell detection is defined as the use of machine learning to determine and localize certain types of cells presented in microscopy images (Alom et al. 2018b). It could be divided into two different techniques: detection-based counting and estimation-based density. For the detection-based technique, the model requires prior high-efficient region detection and segmentation of the overlapping subset of those regions. The strategy of this approach is to segment and locate objects first and then associate those objects among frame (Arteta et al. 2012). In contrast, the density estimation-based model could work without prior object detection or segmentation. Thus, the number of cells is counted by integrating over density maps(Fiaschi et al. 2012).

Overall, the success of computer vision in digital microscopy can provide useful information to clinical practice. A key computer vision application is the biomedical image analysis of cells and tissues, for example, to detect cancerous cells or tumours (Thomas and John 2017). Besides, micro-data corresponding to specific structures, cell nuclei counting, detecting abnormal cell nuclei and analysing the presence of antigens within target cells could be also presented (Xue and Ray 2017). This information is crucial to pathologists to routinely check on many aspects of the diseases such as cancer sub-type, the grade or reaction of the immune system. In the end, patients could be given the most timely and optimized treatment plan (Naylor et al. 2018).

### **2.3 Related work in Cell Segmentation**

This section concentrates on providing relevant research works in cell segmentation, which belongs to the main focus for this study. Indeed, efficient and robust cell segmentation serves as a critical prerequisite for many subsequent biomedical image analysis. It remains a challenging task due to overlapping cells, background noise, and large variations in cell sizes and shapes. Historically, the work of cell segmentation has been through a revolution from handcrafted image processing techniques to machine learning and deep convolutional neural networks (DCNNs).

Initially, the solution to cell segmentation problem only dealt with clear free-lying cells. The most popular handcrafted approach is based on simple intensity thresholding methods, which is preceded by basic mean filters for noise removal (Wu, Gil, and Barba 1998). The inherent difficulty is to find the optimal threshold. The adaptive thresholding schemes are Ostu, Yen, mean and isodata. According to Vala and Baxi, Otsu thresholding is able to reduce time spent on computation and still maintain reasonable segmentation results (Vala and Baxi 2013). The result of their experiment showed that the algorithm enhanced the speed computation and returned good segmentation on salt and pepper noise images but did not perform well on Gaussian noise images. For this

study, Otsu thresholding is chosen as the first baseline segmentation to compare with the proposed deep learning models due to its simplicity and popularity.

Besides, another popular method employed in the previous works is the combination of morphological operations and watershed algorithms. Tulsani, Saxena and Yadav mentioned three advantages of using morphological watersheds in cell segmentation. Firstly, the region contours from watersheds algorithm attach well to the cell boundaries. Secondly, the results produced by watersheds operation is equal to the original image. Thirdly, the overlapping cells' boundaries could be disclosed if single mask for each cell is provided ([Tulsani, Saxena, and Yadav 2013](#)). In contrast, it is found out that morphological watershed could easily suffer from over-segmentation problems in a research work conducted on living cells images. The algorithm was recognised as quite sensitive to noise and does not work well with images with low contrast boundaries ([Hashia and Mir 2014](#)). For this study, watershed morphology is chosen as the second baseline segmentation method in order to observe its limitation in segmenting cell dataset and how deep learning perform on the same task.

From the modest beginning of machine learning, a diversity of methods that differ in terms of mathematical foundations have been researched in numerous biomedical application scenarios. The proposed methods have two directions: supervised and unsupervised methods. In a study on lymphocyte cell segmentation, Mohammed et al. focused on using SVM classifier as a supervised method to extract the accurate mask for nucleus and cell ([Mohammed et al. 2013](#)). By considering cell segmentation as a classification problem, they proved that SVM is capable of classifying the image pixels into cell or non-cell pixel. On another hand, Dhanachandra, Manglem and Chanu demonstrated K-means clustering as an unsupervised method that could be frequently suitable for biomedical object segmentation ([Dhanachandra, Manglem, and Chanu 2015](#)). Their results showed that the clusters could segment targeted cell from the background based on K-centroids. As applying K-means alone may result in the segmentation of noise, they also recommended combining K-mean and a mean filter to obtain noise-free results.

Recently, the rise of DCNNs has revolutionized nuclei research by establishing big data approaches as the main strategy of this domain. According to a study evaluating different strategies for nucleus segmentation in fluorescence images, the results demonstrated that DL improves accuracy and reduce the number of biologically relevant errors by half compared to other classical approaches ([Caicedo et al. 2019](#)). CNN is indeed considered as a well-suited model for segmenting object shapes, even when supplied with vague seed and boundary constraints ([Eschweiler et al. 2019](#)). The model requires feeding ground truth masks of images as input, which labels the pixels. While scanning the original image, a CNN looks at a small "filter" of several pixels each time until it has mapped the entire image ([Standford 2019](#)). There is also a related work which combines CNN with other classic image processing technique such as watershed segmentation, generate a two-stage segmentation that could improve cell tracking ([Jiang et al. 2019](#)). Last but not least, other well-known CNN-based models such as Deep Residual Network (ResNet) by Microsoft and VGG-16 by Oxford are also suggested to perform cell segmentation tasks.

While traditional CNNs are unable to manage different input sizes, fully convolutional networks (FCNs) have been proposed as a new approach because they use convolutional layers to process varying input sizes and work faster. Indeed, different FCNs models have been researched for particular end-to-end medical image segmentation. A study by the University of Freiberg has introduced the U-net as a supervised method for end-to-end biomedical image segmentation ([Ronneberger, P.Fischer, and](#)

Brox 2015). This encoder-decoder model has recently become popular with the application in different types of cells ranging from white blood cells, cell nuclei, inner cells to tumours. Based on the original U-net, the adjustment versions to improve better performance were proposed consequently such as RU-net, R2U-net (Alom et al. 2018a) or even an unsupervised method W-net (Xia and Kulis 2017). U-Net and R2U-net have been applied by its creators to nuclei segmentation task and gained impressive results (Alom et al. 2018). Therefore, this study could apply the same methods on a different dataset to observe their improvement over traditional techniques. On the other hand, W-net is a quite new method to cell segmentation because only one study of its adjustment, a residual W-net, has been conducted until now (Das et al. 2019). Thus, W-Net could be selected as an unsupervised model to investigate how deep learning could perform without any labelled data.

Training a comprehensive deep learning model is quite challenging due to vanishing gradient problem. Several works have suggested that implementing modern activation functions such as Rectified Linear Units ReLU (Maas, Hannun, and Ng 2013) or Exponential Linear Units ELU (Trottier, Gigure, and Chaib-draa 2017) could be a solution. Another work from Microsoft proposed that conducting identify mapping in each Residual Unit in the model architecture also helps to overcome the problem (He et al. 2016). For this study, the proposed deep learning methods should pay attention to the vanishing gradient problems and therefore, use the appropriate activation function.

This thesis aims to provide a comparative study of how deep learning models could improve cell segmentation over traditional image processing techniques. Based relevant research works, the U-Net and its adjustment, the R2U-Net, are chosen the supervised models to explore. The W-Net is the unsupervised model selected to observe how deep learning models could perform the same task without any labelled data. There is no known work which compares these three models together. All suggested deep learning methods in this study use ReLU as the activation function to avoid over-fitting. As introduce above, Otsu thresholding and morphological watershed are the chosen baseline techniques to compare results with the proposed deep learning methods.

## 2.4 Background in Deep Learning for Image Analysis

This section provides the definition and mathematical operations of all layers used in the proposed deep learning architectures for this study.

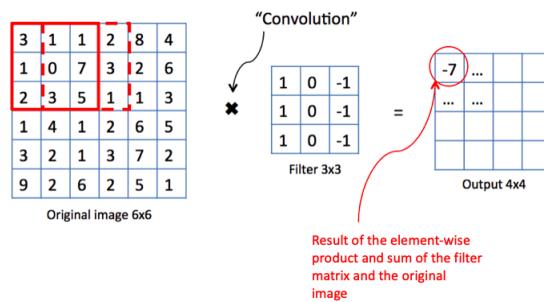


Figure 4: Convolutional operation with filer (Cavaioni 2018)

*Convolutional layer with filter.* Convolutional layer with filter is an operation to extract features from images. By observing the relationship between pixels, the model is able to learn image features using small squares of input data. In mathematics, convolution is a operation between an image matrix and a feature map. The output dot product of convolutional layer has a spatial size of:

$$n_{\text{out}} = \frac{n_{\text{in}} + 2p - k}{s} + 1 \quad (1)$$

where

$n_{\text{in}}$  : number of input features, which has the size of (Width x Height x Depth).

$n_{\text{out}}$ : number of output features.

$k$ : convolution kernel size, which is the volume of filter matrix (also called "Feature Map").

$p$ : convolution padding size, which is the number of zero padding around the border of the input volume.

$s$ : convolution stride size, which controls how the filter shifting over input matrix.

*Recurrent convolutional layer.* Recurrent convolutional layers (RCL) comprise of repetitive associations of inside each convolution layer. Each RCL is created by computing numerous convolutions and the states of RCL units evolve over time-step range from 0 to T. At T= 0, only feed-forward convolution computation takes place.

*Residual unit.* When the input is  $x$ , the residual unit  $R(x)$  or the difference to the true distribution output  $F(x)$  is denoted as

$$R(x) = \text{Output} - \text{Input} = F(x) - x \quad (2)$$

Rearranging the equation, the output or so called the skip connection is defined as

$$F(x) = R(x) + x \quad (3)$$

*Batch normalization.* Batch normalization was applied after each convolutional layer to normalize the output by subtracting the batch mean and dividing the batch standard deviation. To increase the stability for the neural networks, batch normalization allows each layer to learn by itself a little independently and reduces overfitting.

*Dropout.* Dropout is used to ignore some neurons which is chosen randomly during the training phase. As the deep learning model contains numerous of parameters, the model might learn the pattern of training data too well which leads to an over-fitting and bias model. Dropping out some random neurons might raise the ability of the model to work efficiently on unseen data.

*2x2 Max pooling.* Max pooling is a function to reduce the size of the feature map so the network has fewer parameters. A 2x2 Max pooling layer takes the largest element from each 2x2 rectified feature map.

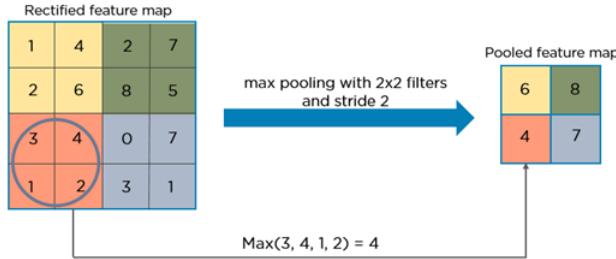


Figure 5: Max pooling 2x2 operation ([Lamba 2017](#))

*Rectified linear unit (ReLU) activation.* ReLU activation is the function to replace all the negative values by zeros. It helps to gain better gradient propagation and reduce vanishing gradient problems.

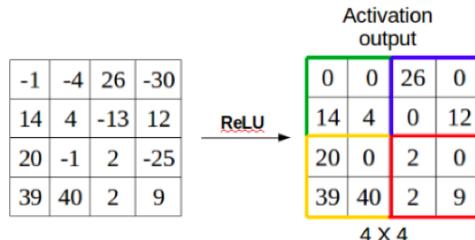


Figure 6: ReLU activation ([Gurumurthy, Kestur, and Narasipura 2019](#))

*Transposed convolution.* Transposed Convolution is a layer to combine convolution with up-scaling of an image. The technique is used to merely reconstruct the spatial transformation. In the U-net decoder path, 2x2 transposed convolution is used to perform upsampling for the feature map.

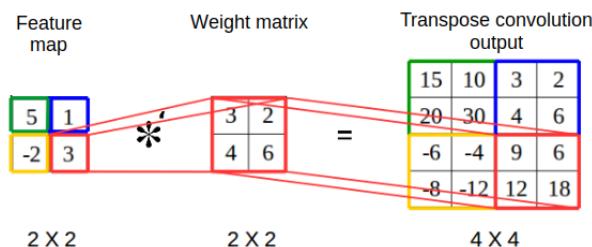


Figure 7: 2x2 Transposed convolution operation ([Gurumurthy, Kestur, and Narasipura 2019](#))

### 3. Experimental setup

This experiment is a comparative study to examine the abilities of deep learning methods in automated cell segmentation. This experiment is conducted using available fluorescent microscopy dataset and applying several methods including manual image processing techniques and deep learning methods used in biomedical image analysis. Python is chosen as the main programming language.

#### 3.1 Dataset

The dataset is acquired from Kaggle competition Data Science Bowl 2018 ([Hamilton 2018](#)). It contains numerous microscopy images and their labelled masks. The images are under a diversity of conditions and vary in cell type, magnification and imaging modalities. The dataset contains two sub-sets:

- Train set: contains 670 fluorescent microscopy images and their masks provided by experts. Each image comprises of several masks as a single mask only presents a single cell. The number of masks indicates the number of cells in the image.
- Test set: contains 65 images from unseen experimental conditions. It also contains images that are not scoring. There is no ground truth provided for the test set.

#### 3.2 Data Preparation and Exploration

The dataset is downloaded using Kaggle API. Because each microscopy image contains several masks, the pre-processing work comprises of combining all the masks of each microscopy image into a comprehensive gray-scale ground truth. Matplotlib is the Python package used to visualize images in this study. All images are then resized into size 256x256x3 and their gray-level masks are resized into 256x256x1 to standardize the inputs before going to segmentation phase. The train set is split into 80% training and 20% validation.

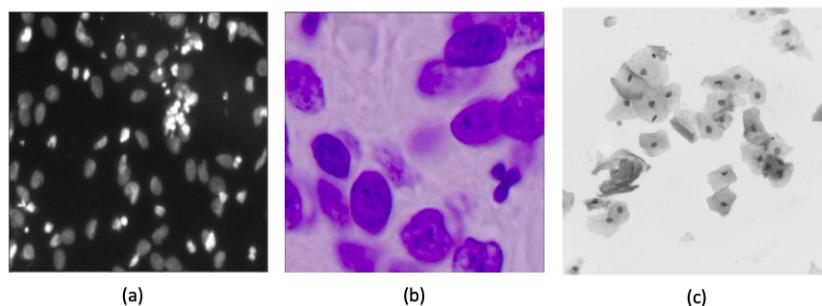


Figure 8: Three types of image modalities in the dataset. From left to right: (a) Gray-scale fluorescent image, (b) Histological slide and (c) Bright-field image ([Hamilton 2018](#))

By applying k-means clustering on the color channels, the outcome reveals three types of microscopy images in the train set. Based on the dominant colors distributions, the statistics of different image modalities in the train set is presented in Figure 9. It

could be seen that most of the images are gray-scale fluorescent type. Histological slides are also provided, in which the nuclei are violet-dominant. Finally, some bright-field images are provided with the purpose to introduce some unexpected environment to the cell segmentation problem.

Type	Number of images	Percentages
Gray-scale fluorescent images	546	81.49%
Histological slides	108	16.12%
Bright-field images	16	2.39%

Figure 9: Different image modalities of train set

### 3.3 Baseline Methods

To observe the improvement of deep learning over traditional image processing, conducting baseline traditional segmentation is required. According to the related work, Ostu method and morphological watershed with distance transform are chosen as they are two of the most popular handcrafted methods in identifying cell boundaries. The baseline methods are implemented using Skimage package.

**3.3.1 Ostu thresholding.** Ostu method is the most commonly used segmentation method and is simple to implement. It is based on the image intensity histogram, which Ostu threshold acts as the criterion to selects whether the pixel belongs to the background or the foreground (Liu, Li, and Tian 1993).

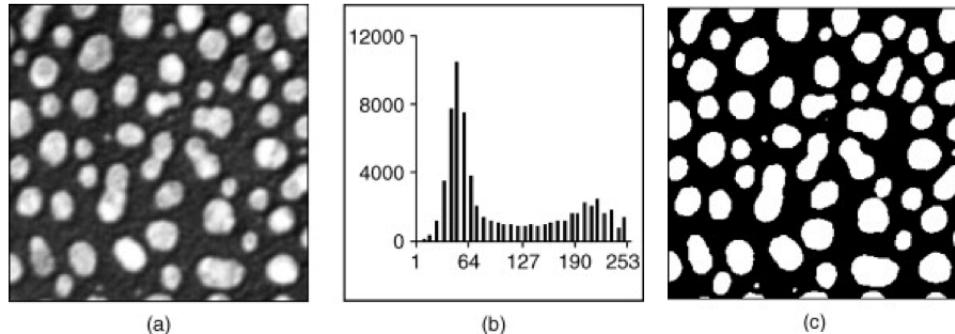


Figure 10: **An example of Otsu thresholding.** (a) Original image, (b) Histogram of image (a), (c) Result of thresholding with  $T = 127$  (Rogowska 2009)

In Otsu thresholding, if  $L$  is the gray level of an image,  $f_i$  indicates the number of the pixel points that is gray level  $i$ , the total points for the pixels are  $N = \sum_{i=1}^L f_i$ , the image pixels histogram is expressed as the probability density distribution as following

$$p_i = \frac{f_i}{N} \quad \sum_{i=1}^L p_i = 1 \quad p_i \geq 0 \quad (4)$$

The threshold divides the image into two categories; foreground and background. Image pixels with intensities greater than the threshold value is assigned as the foreground. Otherwise, these pixels are assigned to the background value ([Qidan Zhu, Liqui Jing, and Rongsheng Bi 2010](#)). In the conducted Otsu algorithm, the output results returned a series of binary images of 0 and 1, where 0 stands for black and 1 stands for white.

**3.3.2 Morphological watersheds based on distance transform.** Watershed segmentation is the second baseline method which is applied to separate cells after Otsu thresholding. The operation are useful for segmenting cells that are touching one another. The watershed transform finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low.

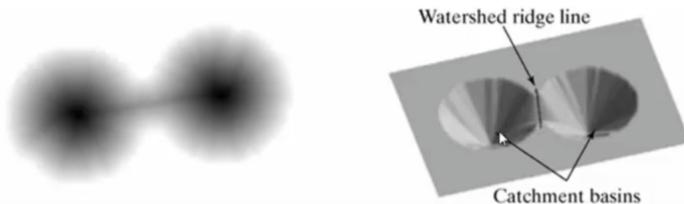


Figure 11: Visualising morphological watersheds: the image on the left is the topological representation of the image on the right ([Agarwal 2015](#))

The distance transform is an image processing technique for binary images. It receives a binary image as an input, while the output is a distance map has the same dimensions with the input image. However, each object pixel in the output is assigned to the closed obstacle pixel ([Fisher et al. 2003](#)). The important drawback of distance transform is its lack of robustness in treatment for outliers. In fact, it might lead to a shape that is less smooth than the original image ([Brunet and Sills 2017](#)). In this study, Euclidean distance is selected as a distance metric. This metric computes directly straight-line distance between two pixels ([MathWorks 2017](#)).

0	0	0		1.41	1.0	1.41
0	1	0		1.0	0.0	1.0
0	0	0		1.41	1.0	1.41

Image                          Distance Transform

Figure 12: Computation of the euclidean distance transform ([MathWorks 2017](#))

The watershed transform is a region-based segmentation approach for image segmentation. In geography, watersheds are the line that distinguishes different catchment basins drained by the river systems. In morphological operations, a gradient image has a set of pixels along gray levels that changes sharply and peaks to an edge which can be considered as the watersheds (Pratimacharja and Ghoshal 2012). Based on the computed distance map, the association of each influence zone to each of the regional minima is defined. The watershed is then defined as the boundaries of these influence zones.

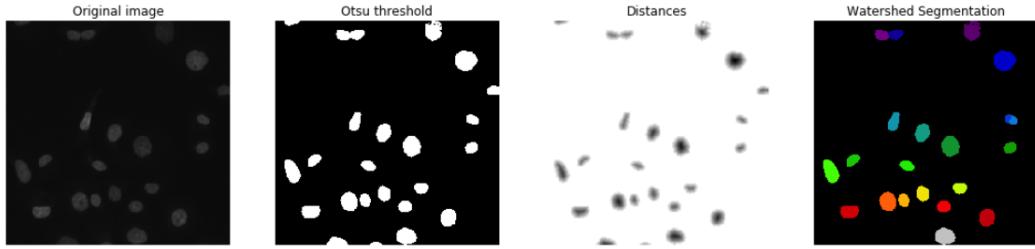


Figure 13: **An example of Distance Transform Watershed on cell images.** From left to right: original cell nuclei image from fluorescent microscope, binary mask computed after applying Ostu threshold, distance map applied to binary mask and resulting image after applying watershed algorithm (Vo 2019)

As watershed could suffer from over-segmentation problems due to large number potential minima, markers based watershed segmentation was implemented in this study. A marker is a connected component for each object, which minimizes the effect of small spatial details (Costa, Mascarenhas, and de Andrade Netto 1997). To compute the marker, several criterion was defined. In details, smooth size was set with 5 pixels, minimum radius was 4 and maximum radius was 20. After creating the marker, the watershed algorithm segmented cells based on distance map and imposed their boundaries on the original image.

### 3.4 Deep Learning Models

Deep learning is a subset of machine learning that is inspired by the structure of the brain. It uses neural networks to learn useful representations of features directly from data (Bengio, Goodfellow, and Courville 2017). In recent years, deep neural networks have proven their abilities over other machine learning methods in visual recognition tasks. Based on related works of biomedical image segmentation, U-net, R2U-Net and W-net are chosen as the promising models that would robust performance and yield high accuracy. The proposed deep learning methods are implemented with Keras and TensorFlow frameworks and trained on Google Colab GPU. The implemented code for this study could be found at: <https://github.com/linhvo2808/Cell-Segmentation>.

#### 3.4.1 U-Net model.

U-net is the first chosen CNNs model to perform cell segmentation task. U-Net is a supervised model that takes both input images and their masks into the training. This model consists of 23 convolutional layers allocating in both contacting path (encoder) and the expansive path (decoder)(Ronneberger, P.Fischer, and Brox 2015).

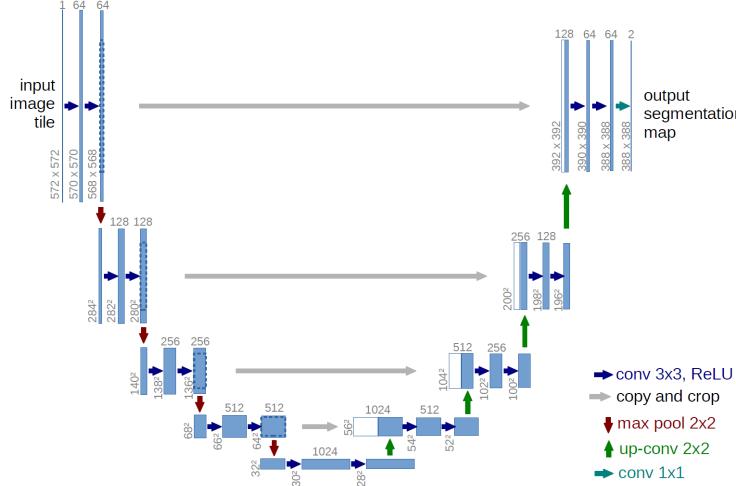


Figure 14: U-net Architecture ([Ronneberger, P.Fischer, and Brox 2015](#))

**Contracting path:** The contracting path is used to capture the context of the images. It consists of 5 convolutional units, followed by a downsampling 2x2 max pooling operation with 2 strides. Each convolutional unit is set with two 3x3 convolutional layers, batch normalization and a ReLU activation (figure 15). The feature channels are doubled for each downsampling step.

**Expensive path:** In the expansive path, the feature map is up-sampling with 2x2 transposed convolution, concatenating with the corresponding cropped image from the contracting path and then conducting the convolutional unit. In contrast with the contracting path, the number of features channels are halved for each up-sampling step. The final convolutional layer is a 1x1 convolution to map 64-component feature vector back to the desired number of classes.

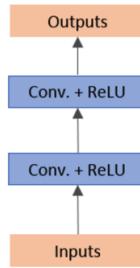


Figure 15: Forward convolutional unit in U-net ([Alom et al. 2018a](#))

In the U-net model, a kernel size of (3,3) is applied. The number of strides applied is 2, which means the model moves the filter to 2 pixels at a time. The model applies padding to input if needed so that input image gets fully covered by filter and stride specified. The drop-out rate is set as 0.25. Adam optimizer is used and ReLU is the activation function.

### 3.4.2 R2U-Net model.

R2U-Net is an adjusted U-Net version inspired by deep residual model. The R2U-Net is implemented similar to the U-net in term of general encoder-decoder architecture, number of input features, kernel size, padding and strides. The difference is that the model replaces the forward convolution unit with the residual recurrent convolution unit (RRCN). The RRCN unit structure is depicted in figure 16a.

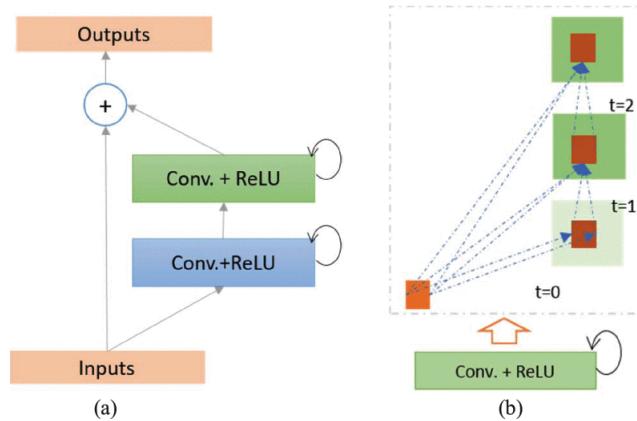


Figure 16: Elements of R2U-Net: (a) the recurrent residual convolutional unit and (b) the unfolded version of the recurrent convolutional layer. ([Alom et al. 2018a](#))

While the convolutional unit in U-Net learns the true output  $F(x)$ , the recurrent residual convolutional unit in R2U-Net learns the residual  $R(x)$  in the association between the input  $x$  and the true output  $F(x)$  ([Sahoo 2018](#)). This recurrent unit aims to solve the vanishing gradient problems with the increasing number of layers. Therefore, larger gradients are propagated to initial layers so they could learn as fast as the final layer, which enable the model to train deeper network ([Alom et al. 2017](#)). In the conducted R2U-Net, each residual unit contains two recurrent convolutional layers, which is combined into a residual recurrent convolutional unit (figure 16b).

In R2U-Net implementation,  $T = 2$  is set, which means one general convolution layer and two recurrent layers are accumulated in the recurrent residual convolutional unit. The path of max-pooling, down-sampling and up-sampling are the same as U-Net encoder-decoder architecture. The softmax layer is applied at the end of the model to calculate class probability.

### 3.4.3 W-Net Model.

W-Net is an unsupervised model inspired by U-Net. The model was designed to reconstruct input images and predict the segmentation boundaries without any labels required ([Das et al. 2019](#)). W-Net architecture consists of  $U_{Enc}$  to capture context and  $U_{Dec}$  to enable precise localization. In total, the model consists of 46 convolutional layers which could be divided into 18 small convolutional units. The conducted W-Net architecture is visualized in figure 17, each convolutional unit in W-Net is similar to the convolutional unit in U-Net (figure 15).

Skip connections are the important parts of W-Net, which are helpful to train deeper neural networks. Using concatenation in skip connections makes the network more effective in weight initialization and improves performance. However, it also leads

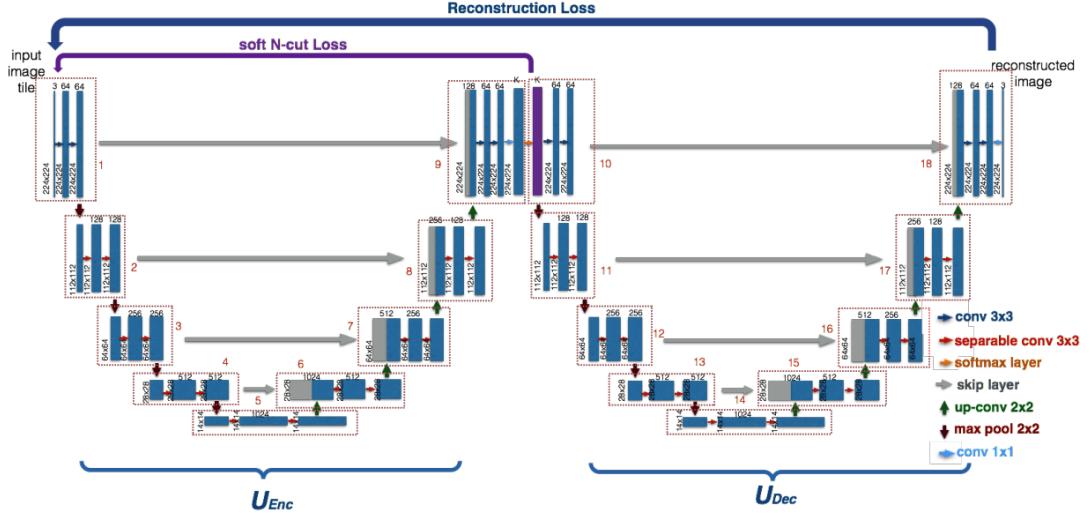


Figure 17: W-net model architecture (Xia and Kulis 2017)

to a more number of parameters, which aggravates the learning burden for W-net. In the conducted model, two U-Nets are connected using concatenation for features combination. In each U-Net, feedforward convolutional layers are applied.

### 3.5 Implementation

In deep learning implementation, all models use both ReLU and sigmoid as activation functions. Adam is set as the optimizer and binary cross-entropy is set as the loss function. The learning rate is initially set as 0.001 and changed in an adaptive way. To avoid over-fitting, drop-out rate is applied as 0.25 and early stopping is set with patience equals 10. Model size and other hyper-parameters are disclosed in figure 18. The results are then visualised using Python package Matplotlib.

Method	Parameters	Epochs	Batch size
U-net	2,164,593	100	16
R2U-net	5,995,441	100	16
W-net	12,573,730	100	16

Figure 18: Detailed hyper-parameters and model size

### 3.6 Evaluation Methods

The main purpose of this study is to observe the improvement of deep learning methods over traditional image processing. Hence, it is essential to select the right metrics that fit with image segmentation problems. After the proposed methods have

been implemented, their segmentation outputs are evaluated based on:

**3.6.1 Hausdorff distance.** Hausdorff distance measures how far the output segmentation results and the provided ground truth are from each other. In theory, the metric is suitable to measure the distance between two surfaces and also prevent possible underestimation caused by using one-sided distance (Ma et al. 2012). A lower Hausdorff distance indicates more similarity.

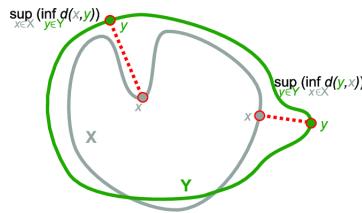


Figure 19: Hausdorff distance (Pellerin 2014)

Given the 2 set of points X and Y, the Hausdorff distance is defined as

$$H(X, Y) = \max(h(X, Y), h(Y, X)) \quad (5)$$

where

$$h(X, Y) = \sup_{x \in X} \inf_{y \in Y} \|x - y\| \quad (6)$$

which  $\|\cdot\|$  represents the space between two points set, which is calculated based on Euclidean norm.

**3.6.2 Dice coefficient.** Dice coefficient evaluates the similarity between the ground truth and output segmentation. The higher measure indicates more similarity (Sumithra and Deepa 2016). The formula used for calculating dice coefficient is given by:

$$\text{Dice}(X, Y) = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (7)$$

where X is the segmentation results by image processing techniques and Y is the ground truth provided by experts.

**3.6.3 Jaccard index.** Jaccard index measures the performance of the generated segmentation results, which is defined as the size of the intersection divided by the size of the

union with ground truth. The higher measure represents more similarity ([Sumithra and Deepa 2016](#)). Jaccard index is defined by the following formula:

$$\text{Jaccard}(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (8)$$

where X indicates the segmentation results and Y indicates the ground truth.

#### 4. Results

To demonstrate the performance of deep learning models over traditional image processing, the models have been tested on nuclei segmentation dataset from Kaggle. As has become standard, the performance results are evaluated on three different metrics: Hausdorff distance, Dice coefficient and Jaccard index. In this section, the training results and segmentation outputs of all proposed approaches are presented. Likewise, the performance of each deep learning algorithm is also evaluated. Figure 20 summarises the the average evaluation results of all outputs compared to the ground truth in validation dataset.

Method	Hausdorff distance	Dice coefficient	Jaccard index
Ostu thresholding	6.080	0.706	0.617
Watershed	7.200	0.659	0.549
U-net	4.226	0.913	0.849
R2U-net	4.210	0.918	0.857
W-net	4.868	0.770	0.797

Figure 20: Performance results on Kaggle cell nuclei dataset

##### 4.1 Baseline Segmentation

Otsu thresholding and watershed segmentation are the selected baseline methods delegating for traditional image processing. From the quantitative results in figure 20, it could be seen that morphological watersheds returned the highest Hausdorff distance and the lowest Dice coefficient and Jaccard index among the proposed methods. Therefore, the watershed algorithm is not only complex to compute but also not efficient enough for cell nuclei segmentation. Moreover, Otsu thresholding performed on average better watershed algorithm. However, this simple computation is still not able to identify cell's boundaries as accurate as proposed deep learning models.

From the quantitative results of three different types of microscopy images, it is observed that both techniques is able to segment cells in images with clear background and foreground. Figure 21 illustrates results of applying baseline image processing to fluorescent images. While Otsu thresholding is able to identify the cells based on the distribution of pixels, morphological watershed can even able to recognise some overlapping cells.

With the images that are low in quality or have complex pixels distribution, the traditional techniques fail to distinguish objects against their background. For the applied

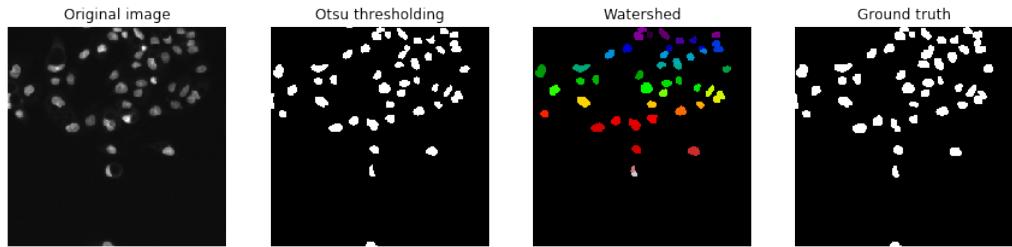


Figure 21: Results of baseline methods applied on a gray-scale fluorescent image

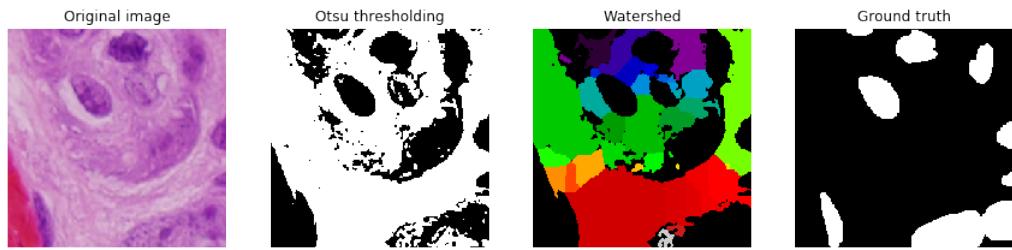


Figure 22: Results of baseline methods applied on a histological slides image

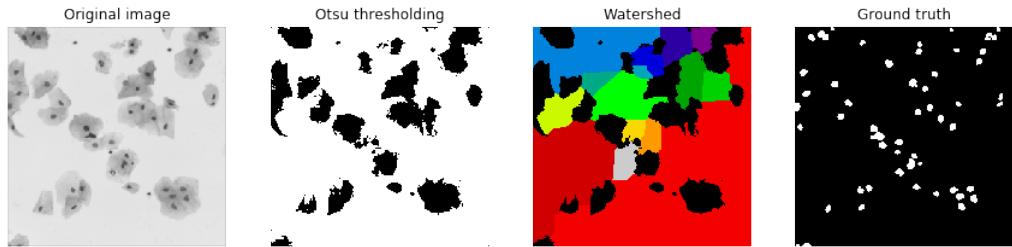


Figure 23: **Results of baseline methods applied on a bright-field image.** In watershed algorithm, each colour label represents a separate nuclei.

cell nuclei dataset, histological slides images are violet-dominant and the separation of cells and background are quite unclear. Thus, Otsu thresholding fails in separating cells and background based on the chosen metric in this type of microscopy images. In addition, the watershed algorithm is quite easy to suffer from over-segmentation problems based on distance map and the imposed boundaries. Figure 22 illustrates example results of applying baseline methods on a violet-dominant environment.

For the bright-field images provided in the dataset, the image contains both cell nucleus and cell membrane. Therefore, both baseline methods are not functional enough to identify cell boundaries out of a image with complex structure. Figure 23 illustrates the results applying baseline segmentation to bright-field microscopy images.

## 4.2 Supervised Method: U-Net

The supervised method U-net is promising while training on cell nuclei dataset. Even though the number of epochs was initially set at 100, early stopping occurred after 60 epochs. Over-fitting did not occur during the training process. The model performance on the validation set is shown in figure 20. It could be seen that the U-Net results are the second-best, slightly lower than R2U-Net results. With the mean Dice coefficient of 0.913, the model improves cell segmentation task by 20% compared to Otsu thresholding and 25% compared to morphological watersheds.

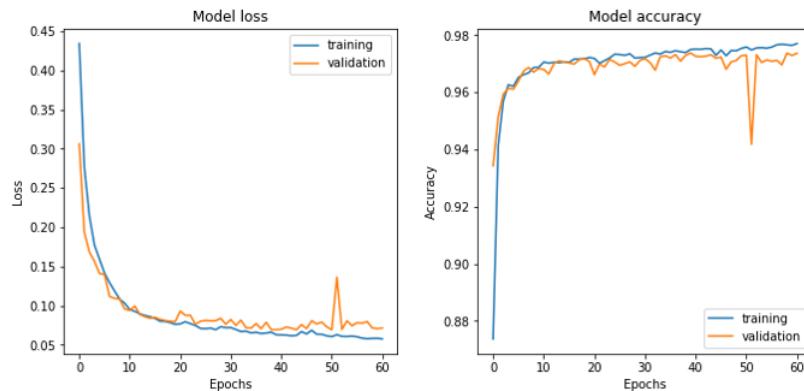
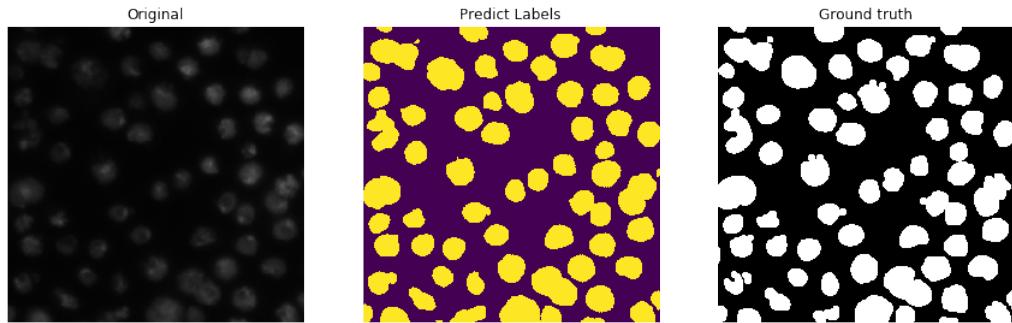


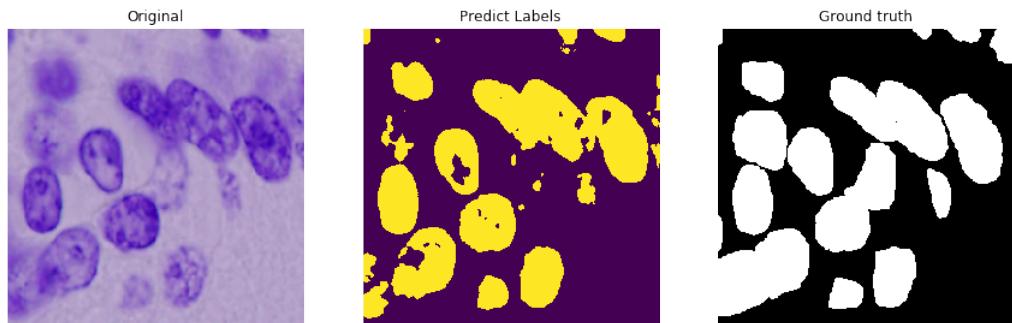
Figure 24: Learning curve over epochs of U-Net

When applying U-Net on three different types of microscopy images, the model archived the best performance on gray-scale fluorescent images. Its encoder-decoder architecture enables the model to recognise hidden representation and reconstruct the masks even though several microscopy images are low in resolutions. Figure 25 demonstrates the predicted labels of U-Net when running on a blurry fluorescent image. In this case, U-Net is able to identify the cell boundaries closed to human level.

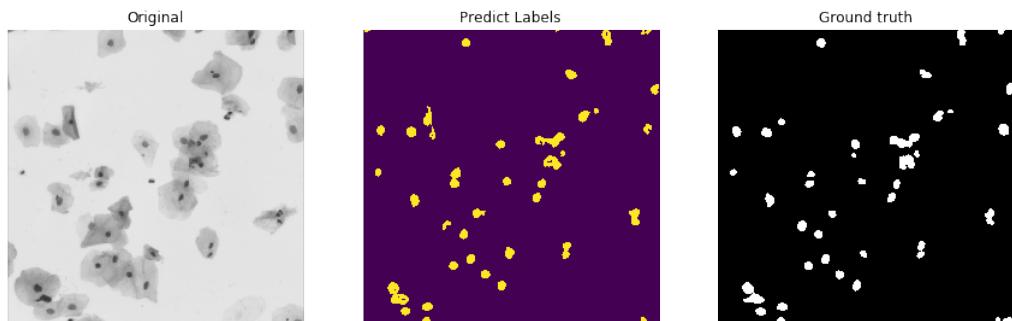
On the other hand, U-Net did not perform well on violet-dominant and special environment. Figure 26 illustrates the results of running U-Net model on a histological slides image. It could be seen that the predicted masks appear several empty spaces in the middle of the cells. For bright-field images, U-Net is able to identify the location of cells out of cell membranes and images' background. However, the predicted boundaries are not as accurate as expected. Figure 27 presents a case of U-Net model predicting on a bright-field microscopy image, while the Dice coefficient is satisfied but Jaccard similarity is quite low.



**Figure 25: Predicted labels of U-Net model on a gray-scale fluorescent image.** Compared to the ground truth, Jaccard similarity is 0.865, Hausdorff distance is 6.083 and Dice coefficient is 0.927.



**Figure 26: Predicted labels of U-Net model on a histological slides image.** Compared to the ground truth, Jaccard similarity is 0.712, Hausdorff distance is 8.367 and Dice coefficient is 0.832.



**Figure 27: Predicted labels of U-Net model on a bright-field image.** Compared to the ground truth, Jaccard similarity is 0.697, Hausdorff distance is 3.873 and Dice coefficient is 0.821.

### 4.3 Supervised method: R2U-Net

R2U-Net is the best fit in cell segmentation task among all proposed methods. In the implementation, R2U-Net model occurred early stopping after 62 epochs. Based on the fact that the model is supervised and it contains skip connections which takes residuals into consideration, R2U-Net demonstrates its robustness and automation of DNNs in cell segmentation. From figure 20, R2U-Net achieved the most impressive outcomes, with the highest Dice coefficient, Jaccard index and lowest Hausdorff distance. The method predicted on average 90% similar to the ground truths provided by experts (calculated by Dice coefficient). Comparing to the baseline segmentation which returns only around 70% similarity, R2U-Net outperform traditional image processing approaches by more than 20%.

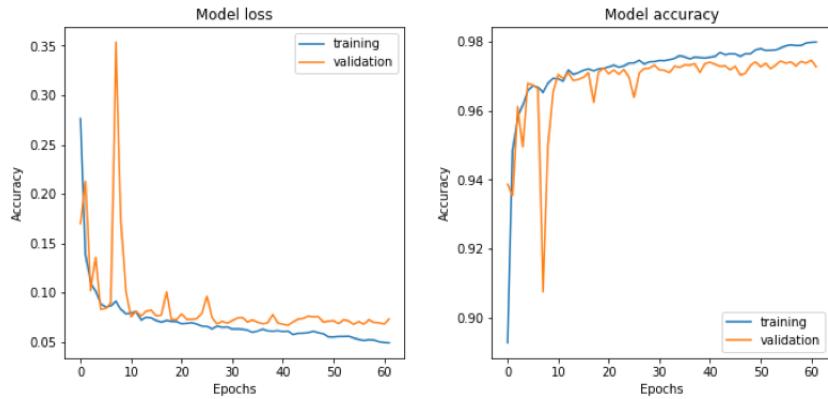
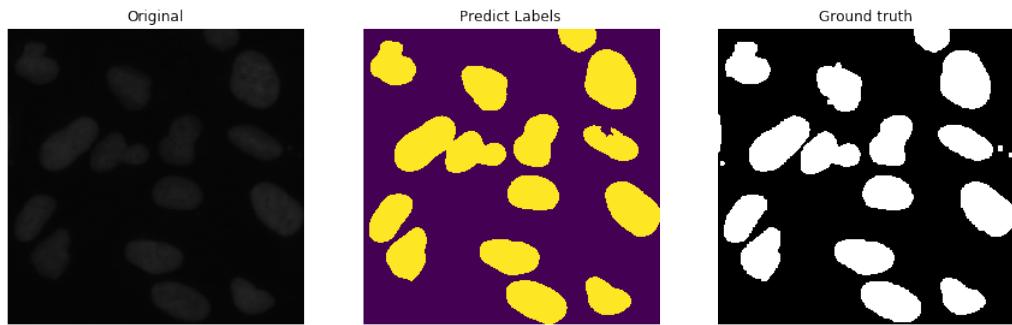


Figure 28: Learning curve over epochs of R2U-Net

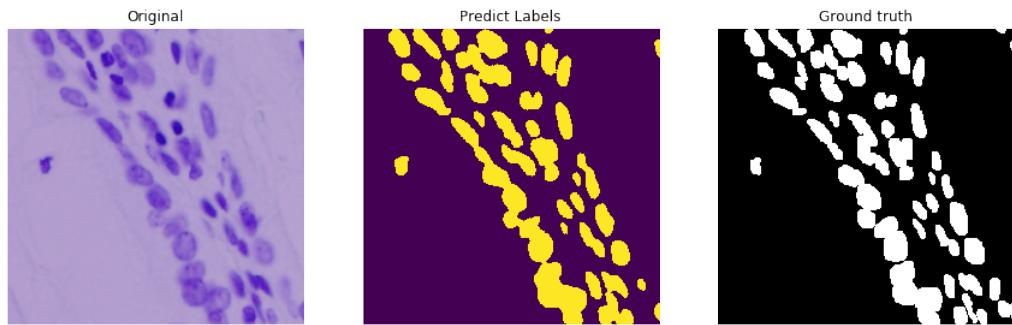
R2U-Net is able to conduct cell segmentation task for all three types of microscopy images. For gray-scale fluorescent images, the model can predict the cells' masks that are more than 95% similar to the expectation. R2U-Net overcomes the challenge of traditional image processing. The model works on images at any conditions including images that are blurry, low in quality and have unclear foreground and background. Figure 29 demonstrates the ability of R2U-Net while running on an unclear fluorescent image.

For histological slides images, the model manages to identify cells even though the images are low in quality and dominant in one colour (violet in this case). Compare with the U-Net where some empty spaces exist in the predicted masks, R2U-Net can capture the cell's structure and predict more comprehensive labels in unseen environment. Figure 30 visualises the result of running R2U-Net models on a histological slides image, where the Dice coefficient is almost 90% similar to human work.

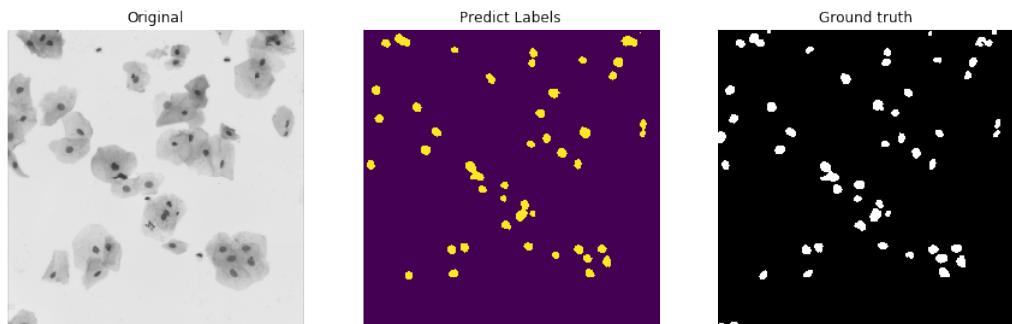
For bright-fields images, R2U-Net can identify the central nucleus out of cell membrane. These results are impressive because the model can learn the nuclei's structure in only 12 bright-field images provided in the training dataset. Figure 31 visualises the results of running R2U-Net models on a bright-field image, while the model succeed in identifying the cell location. However, segmenting accurate cell boundaries is still a challenge for the model to improve.



**Figure 29: Predicted labels of R2U-Net model on a gray-scale fluorescent image.** Compared to ground truth, Jaccard similarity is 0.946, Hausdorff distance is 4.359 and Dice coefficient is 0.972.



**Figure 30: Predicted labels of R2U-Net model on a histological slides image.** Compared to the ground truth, Jaccard similarity is 0.814, Hausdorff distance is 4.796 and Dice coefficient is 0.897.



**Figure 31: Predicted labels of R2U-Net model on a bright-field image.** Compared to the ground truth, Jaccard similarity is 0.708, Hausdorff distance is 3.873 and Dice coefficient is 0.829.

#### 4.4 Unsupervised Method: W-Net

Since W-Net is designed for unsupervised image segmentation, ground truth labels are not provided in the training phase but only used to evaluate the quality of the segmentation. With the epochs initially set at 100, W-net model early stops at epoch 31. Figure 32 shows the accuracy and loss concerning the number of epochs during training. The unsupervised W-Net generated less accurate outputs than the supervised models. However, an average of 77% similarity to the ground truth in term of Dice coefficient is an impressive result. Indeed, W-net manages to learn about cells' patterns by itself and still performs better than traditional image processing techniques. The model performance is 7% better than Otsu thresholding and 10% better than watershed algorithm.

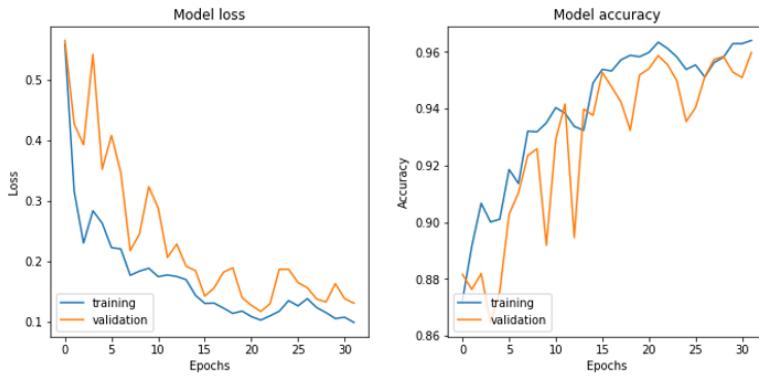
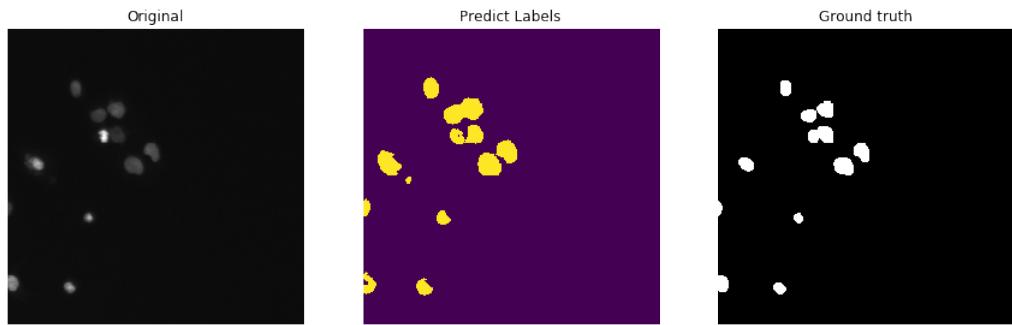


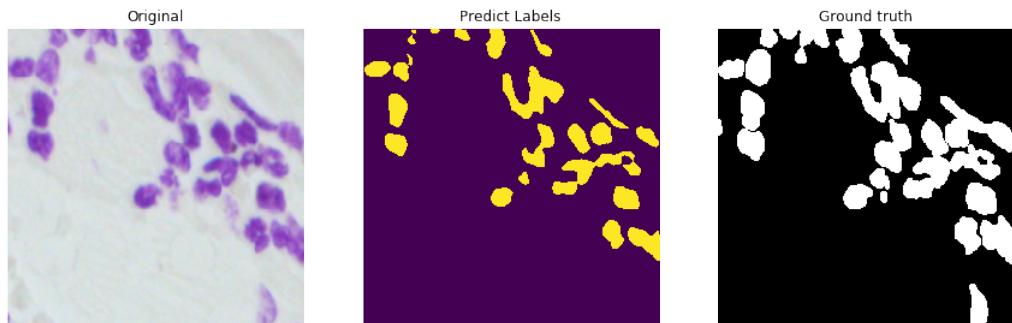
Figure 32: Learning curve over epochs of W-Net

Among three types of microscopy images in the validation dataset, the W-Net performs the best on gray-scale fluorescent images. In Figure 33 illustrates the predicted labels of the W-Net on a blurry fluorescent image. Here, the model can segment the cell boundaries but the mask is not completely fulfilled as empty holes inside the cells area still exist.

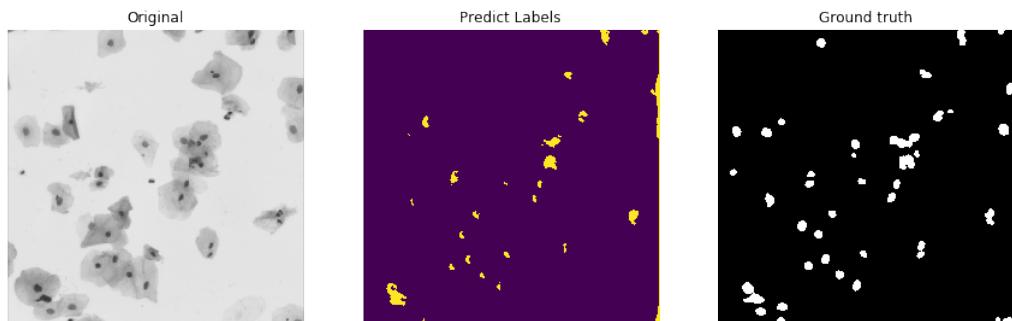
On other conditions rather than gray-scale, W-Net is not able to conduct the segmentation task up to high expectation. Figure 34 illustrates the failure while applying W-Net in a histological slides image, where W-Net is able to identify the cells' location but unable to segment them accordingly. On bright-field images, W-Net can locate the cell but once again the model can not determine a complete cells' shape. Figure 35 presents the failure of W-Net on a bright-field image in fulfilling the predicted mask with many off-limit boundaries.



**Figure 33: Predicted labels of W-Net model on a gray-scale fluorescent image.** Compared to ground truth, Jaccard similarity is 0.564, Hausdorff distance is 4.243 and Dice coefficient is 0.721.



**Figure 34: Predicted labels of W-Net model on a histological slides image.** Compared to ground truth, Jaccard similarity is 0.518, Hausdorff distance is 6.557 and Dice coefficient is 0.682.



**Figure 35: Predicted labels of W-Net model on a bright-field image.** For this result, Jaccard similarity is 0.242, Hausdorff distance 5.292 and Dice coefficient is 0.389.

## 5. Discussion

This study is conducted to demonstrate the abilities of deep learning methods over traditional image processing. Based on related work in biomedical image analysis, two supervised models, namely U-Net and R2U-Net, and an unsupervised model, namely W-Net, have been selected for the experiment. Otsu thresholding and morphological watersheds are the baseline methods delegating traditional image processing due to its popularity in image segmentation. All proposed methods are evaluated using Hausdorff distance, Dice coefficient and Jaccard Index.

After implementing all the proposed methods, the output results returned positively for the study purposes. All deep learning methods performed better than traditional image processing. Supervised methods achieved the best performance with Dice coefficient of more than 0.9. Especially, R2U-Net generated slightly more accurate outputs than U-Net due to its abilities to learn from the residuals rather than just the true distribution outputs. On the other hand, W-Net as an unsupervised model performed the cell segmentation task with 77% similar to the desired outputs.

### 5.1 Answering the research questions

The main research question for this study is:

*To what extent do deep learning improve cell segmentation in fluorescence microscopy images?*

From the experimental results, it can be concluded that deep learning outperforms classical image processing in cell segmentation task. Compared to the Otsu thresholding algorithm, R2U-Net, U-Net and W-Net improved 21.1%, 20.7% and 6.4% respectively. Compared to morphological watersheds, deep learning methods performed even better because morphological watersheds are easy to face over-segmenting problems. Indeed, traditional processing methods are able to segment cells in gray-level and good conditions images. However, they fail to segment images that are blurry, noisy and contain cells overlapping and in multi-colours conditions. In contrast, deep learning models could automate the cell segmentation task and achieve near human-level accuracy in gray-scale fluorescent images. In other image conditions, the models still achieve the outputs with satisfied results compared to the expert-provided data in three-channel images but still quite sensitive to violet-dominant histological slides images.

Contributing to the main question, the sub-questions are answered as follows:

*SQ1: To what extent does the U-Net, a popular supervised deep learning model for medical image analysis improve cell segmentation?*

U-Net returned impressive results with the mean Dice coefficient of 0.913. Compared to traditional Otsu and morphological watersheds, the results of this model improved 20.7% and 25.4% respectively. From the qualitative results, it could be seen that U-Net achieved the best performance on gray-scale fluorescent images. In this type of image modality, the model could predict labels that are similar to the ground truth provided by experts. In the other image modalities, U-Net could return satisfied results for bright-field images but not for histological ones. Indeed, the model confronts difficulties to distinguish correctly between cells and backgrounds in the

violet-dominant environment. Because only a small percentage of histological slides are included in the train set, the conducted U-Net was not well introduced about this type of microscopy images.

*SQ2: To what extent does an adjusted version of the U-Net (an R2U-Net) improve cell segmentation?*

The adjusted version of U-Net, a R2U-Net, returned the best performance among three proposed deep learning models. With the mean Dice coefficient of 0.918, the model showed an average improvement of 21.2% compared to Otsu thresholding and 25.9% compared to morphological watersheds. There are two main reasons that lead to this results. Firstly, R2U-Net is a supervised model, which labelled data is provided during the training process. Secondly, the model architecture comprises of the residual recurrent convolutional layers instead of the feedforward convolutional layers, which enables it to learn from the output residuals rather than the distribution of outputs itself. Therefore, the model has advantages when dealing with noisy inputs coming from the unexpected environment. From the qualitative outputs in appendix ??, it could be seen that R2U-Net performed better than U-Net and W-Net in all three types of microscopy images.

*SQ3: How well does the W-Net, an unsupervised version of the U-Net model perform in cell segmentation?*

Based on the evaluation results on validation dataset, W-Net performed with a mean Dice coefficient of 0.77. Compared to the baseline segmentation, the W-Net model is 6.4% better than Otsu thresholding and 11% better than morphological watersheds. However, W-Net results are not as good as the supervised U-Net and R2U-net results because no labelled data is provided to train the model. Looking at the qualitative results, W-Net could be able to predict particularly on gray-scale fluorescent images. For other type of image modalities, W-Net confronts difficulties in figuring the pattern of cell boundaries by itself. The model is able to identify the cells' location but unable to segment them accordingly. W-Net with this characteristics might be fit with the cell density counting task rather than the cell segmentation task. To sum up, unsupervised W-Net is better than traditional image processing when conducting cell segmentation tasks but the model cannot be compared to the supervised ones.

## 5.2 Shortcomings

From the finding results, it could be seen that supervised models outperformed the unsupervised model in cell segmentation tasks. However, both original microscopy images and their ground truths are compulsory to train any supervised model. For that reason, more time and expertise labour are required in the data preparation phase to create labelled datasets. On the other hand, while the unsupervised model W-net performs with less accuracy compared to the supervised ones, no ground truth is required to train the model. As the model was originally introduced with unsupervised biomedical segmentation purpose, this approach enables the model to learn the cells' pattern by itself.

The dataset used in this study might not be large enough for deep learning models to generalize well from training data to unseen data. The train set contains of only 670 images, which is quite limited for the proposed models to learn as early stop-

ping did occur. Besides, the dataset is biased on gray-level fluorescent images, which make some models fail to predict on other image modalities such as violet-dominant histologiscal slides or bright-field microscopy images. In the future, some noise could be introduced to dataset to enhance model accuracy. The input images could be pre-processed by adding Gaussian filter, salt and pepper to degrade the images. Another solution to this problem could be applying image augmentation. Image augmentation is an automatic way to boost the number of input images. By using Python libraries such as OpenCV, Pillow or scikit-image, new images could be generated by three possible transformations: random rotation, random noise and horizontal flip. Therefore, the hundred-images dataset will be simply enhanced into a thousand-images dataset.

### 5.3 Future Work

In this experiment, the proposed methods are only applied to the cell nuclei dataset provided by Kaggle. Therefore, this comparative study has only been applied on cell nuclei segmentation task. In the future, the same experiment should be implemented in other deep learning models on variant types of cells in order to investigate the most optimized solution in different scenarios. Besides, all proposed models are trained only on 2D data. Future work includes studying how these models perform on confocal microscopy data which acquires 3D images. Last but not least, this study is on focused on cell segmentation problem, which is one of the three main applications of digital microscopy in biomedical research. Future works would also include the same study for the other two applications which are cell classification and cell detection.

## 6. Conclusion

In this study, different deep learning approaches have been applied to the same cell nuclei segmentation tasks. From the experimental results, it is proven that deep learning outperforms classical image processing techniques in separating cells versus their background. Compared to the traditional Otsu thresholding algorithm, R2U-Net, U-net and W-net improved 21.1%, 20.7% and 6.4% respectively. Compared to morphological watersheds, the improvement percentages are 25.9%, 25.4% and 8.8% following the same order. It is observed that traditional processing methods are able to implement cell segmentation in gray-level and good conditions images. However, they fail to segment images that are blurry, noisy, cells overlapping and in multi-colours conditions. In contrast, the proposed deep learning models could automate the cell segmentation task and achieve near human-level accuracy in microscopy images. In other image conditions, the models still achieve the results with satisfied similarity to the expert-provided data.

Among the proposed approaches, supervised methods achieved the best performance with Dice coefficient at 90% similar to the ground truth provided by experts. The R2U-Net model generated more accurate outputs than the U-net due to its ability to learn from the residuals rather than just the true distribution outputs. Indeed, R2U-Net could able to perform well on three different types of image modalities. U-Net could perform quite well on gray-scale fluorescent images but only returned satisfying results on histological slides and bright-field images. All supervised methods require labelled data to train. Therefore, the process of data acquisition is more labour-intensive and time-consuming to create labelled data. On the other hand, the W-Net, an unsupervised model, achieved a result of 77% similar to the desired outputs. This encoder-decoder framework does not require any labelled data and still manage to identify cells in gray-scale fluorescent images. For other image modalities, there are still several challenges for the W-Net. Because less data is needed, this unsupervised model could be considered as a future completely automatic solution that saves time and requires less expertise.

## References

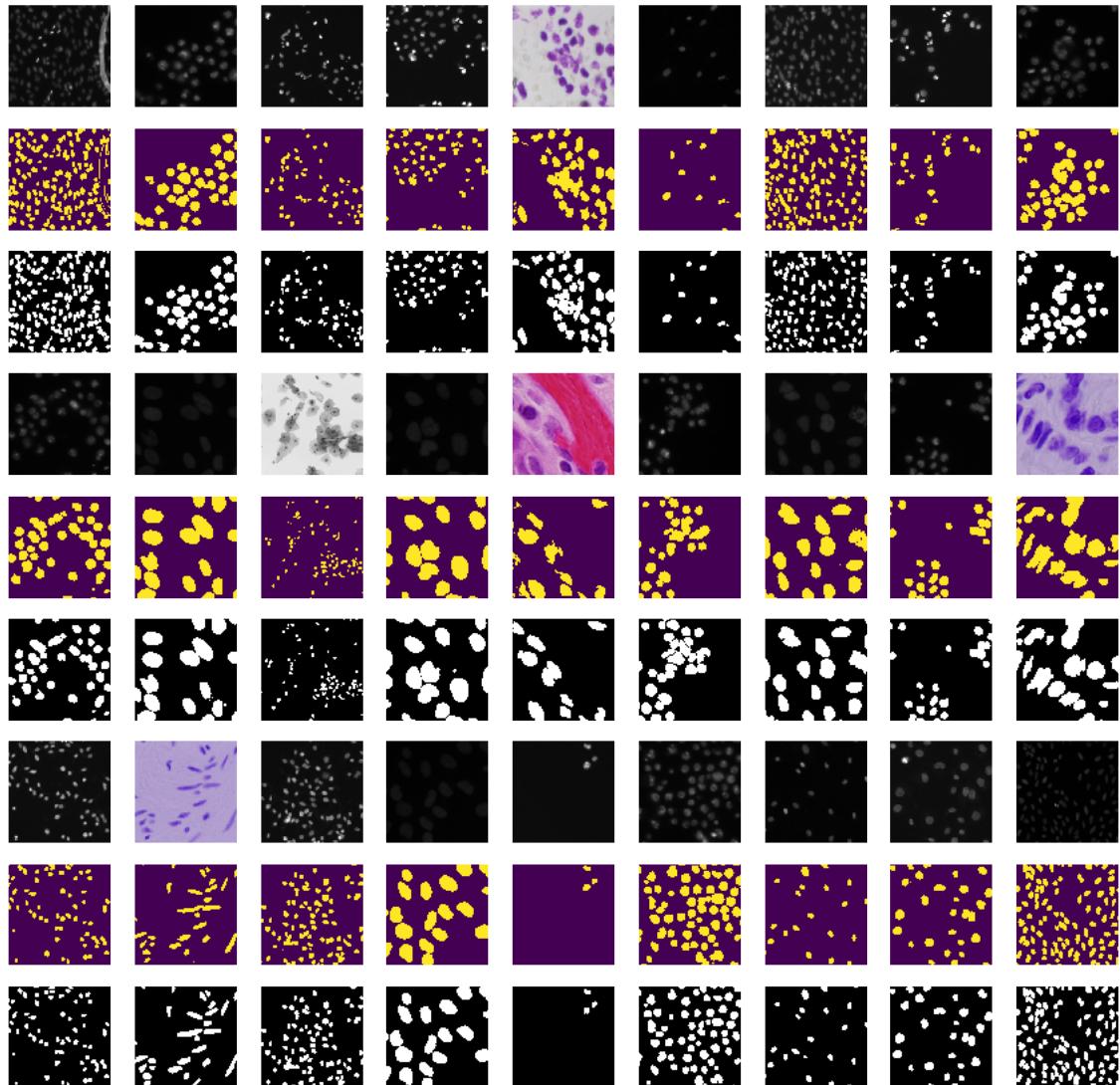
- Agarwal, Rashi. 2015. Segmentation using watershed algorithm in matlab.
- Alom, M. Z., C. Yakopcic, T. M. Taha, and V. K. Asari. 2018. Nuclei segmentation with recurrent residual convolutional neural networks based u-net (r2u-net). In *NAECON 2018 - IEEE National Aerospace and Electronics Conference*.
- Alom, Md. Zahangir, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. 2017. Improved inception-residual convolutional neural network for object recognition. *CoRR*, abs/1712.09888.
- Alom, Md. Zahangir, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. 2018a. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *CoRR*, abs/1802.06955.
- Alom, Md Zahangir, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. 2018b. Microscopic nuclei classification, segmentation and detection with improved deep convolutional neural network (dcnn) approaches.
- Arteta, Carlos, Victor Lempitsky, J. Alison Noble1, and Andrew Zisserman. 2012. Learning to detect cells using non-overlapping extremal regions.
- Bengio, Yoshua, Ian Goodfellow, and Aaron Courville. 2017. *Deep learning*, volume 1. Citeseer.
- Brunet, Dominique and David Sills. 2017. A generalized distance transform: Theory and applications to weather analysis and forecasting. *IEEE Transactions on Geoscience and Remote Sensing*, 55(3):1752–1764.
- Caicedo, Juan C, Jonathan Roth, Allen Goodman, Tim Becker, Kyle W Karhohs, Matthieu Broisin, Molnar Csaba, Claire McQuin, Shantanu Singh, Fabian Theis, et al. 2019. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *BioRxiv*, page 335216.
- Cavaioni, Michele. 2018. Deeplearning series: Convolutional neural networks.
- Costa, Jose Alfredo Ferreira, Nelson DA Mascarenhas, and Marcio Luiz de Andrade Netto. 1997. Cell nuclei segmentation in noisy images using morphological watersheds. In *Applications of Digital Image Processing XX*, volume 3164, pages 314–324, International Society for Optics and Photonics.
- Das, Sushmita, Ankur Deka, Yuji Iwahori, MK Bhuyan, Takashi Iwamoto, and Jun Ueda. 2019. Contour-aware residual w-net for nuclei segmentation. *Procedia Computer Science*, 159:1479–1488.
- Dhanachandra, Nameirkpam, Khumanthem Manglem, and Yambem Jina Chanu. 2015. Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764–771.
- Discovery. 2018. Fluorescence microscope image of bovine pulmonary artery endothelial cells bpae stained for mitochondria, phalloidin, and nuclei undergoing mitosis.
- Eschweiler, D., T. V. Spina, R. C. Choudhury, E. Meyerowitz, A. Cunha, and J. Stegmaier. 2019. Cnn-based preprocessing to optimize watershed-based cell segmentation in 3d confocal microscopy images. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 223–227.
- Fiaschi, Luca, Rahul Nair, Ullrich Koethe, and Fred A. Hamprecht. 2012. Learning to count with regression forest and structured labels.
- Fisher, R., S. Perkins, A. Walker, and E. Wolfart. 2003. Distance transform.
- Gurumurthy, Vikas Agaradahalli, Ramesh Kestur, and Omkar Narasipura. 2019. Mango tree net – a fully convolutional network for semantic segmentation and individual crown detection of mango trees.
- Hamilton, Booz Allen. 2018. Find the nuclei in divergent images to advance medical discovery | Kaggle.
- Hashia, B. and A. H. Mir. 2014. Segmentation of living cells: A comparative study. In *2014 International Conference on Communication and Signal Processing*, pages 1524–1528.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027.
- J. Gulo, Carlos, Antonio Sementille, and Joao Tavares. 2019. Techniques of medical image processing and analysis accelerated by high-performance computing: A systematic literature review. *Journal of Real-Time Image Processing*, 16:1891–1908.
- Jaime A. Rincon Cardona, Cristian Huck Iriart and Maria Lidia Herrera. 2013. *Confocal Laser Microscopy - Principles and Applications in Medicine, Biology, and the Food Sciences*. IntechOpen.
- Jiang, Wenbo, Lehui Wu, Shihui Liu, and Min Liu. 2019. Cnn-based two-stage cell segmentation improves plant cell tracking. *Pattern Recognition Letters*, 128:311–317.

- John, J., M. S. Nair, Wilscy M, and Anil Kumar P.R. 2014. Cell image segmentation and color coding based on nucleus to cell ratio. In *2014 International Conference on Electronics and Communication Systems (ICECS)*, pages 1–5.
- Kenneth, R. Spring and W. Davidson Michael. 2019. Introduction to fluorescence microscopy. Accessed: 2019-16-09.
- Lamba, Harshall. 2017. Understanding semantic segmentation with unet. Accessed: 2019-15-11.
- Lindon, John, George E. Tranter, and David Koppenaal. 1999. *Fluorescence Microscopy, Applications*. Encyclopedia of Spectroscopy and Spectrometry. Elsevier, Ashfield, NSW, Australia.
- Litjens, G, T Kooi, B Bejnordi, A Setio, and J van der Laak B van Ginneken C Sánchez F Ciompi, M Ghafoorian. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88.
- Liu, J, W Li, and Y Tian. 1993. The automatic thresholding of gray-level picture via 2d otsu algorithm. *Acta Autom. Sin.*, 19(1):101–105.
- Ma, X., J. Zheng, Y. Shui, H. Zhou, and L. Shen. 2012. A novel method of mesh simplification using hausdorff distance. In *2012 Third World Congress on Software Engineering*, pages 136–139.
- Maas, Andrew L, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3.
- Majtner, Tomas. 2015. *Texture-Based Image Description in Fluorescence Microscopy*. Ph.D. thesis.
- MathWorks. 2017. Select a web site.
- Méndez-Vilas, Antonio and José Díaz. 2011. Feature extraction , selection and classifier design in automated time-lapse fluorescence microscope image analysis.
- Mohammed, E. A., B. H. Far, M. M. A. Mohamed, and C. Naugler. 2013. Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation. In *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, pages 622–626.
- Naylor, Peter, Marick Laé, Fabien Reyal, and Thomas Walter. 2018. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38:1–1.
- Oei, Ronald Wihal, Guanqun Hou, Fuhai Liu, Jin Zhong, Jiewen Zhang, Zhaoyi An, Luping Xu, and Yujiu Yang. 2019. Convolutional neural network for cell classification using microscope images of intracellular actin networks. *PLOS ONE*.
- Pellerin, Jeanne. 2014. *Accounting for the geometrical complexity of geological structural models in Voronoi-based meshing methods*. Ph.D. thesis.
- Pratimacharjya, Pinaki and Dibyendu Ghoshal. 2012. Watershed segmentation based on distance transform and edge detection techniques. *International Journal of Computer Applications*, 52(13):6–10.
- Qidan Zhu, Lijiu Jing, and Rongsheng Bi. 2010. Exploration and improvement of ostu threshold segmentation algorithm. In *2010 8th World Congress on Intelligent Control and Automation*, pages 6183–6188.
- Rogowska, Jadwiga. 2009. Overview and fundamentals of medical image segmentation.
- Ronneberger, O., P.Fischer, and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241, Springer. (available on arXiv:1505.04597 [cs.CV]).
- Sahoo, Sabyasachi. 2018. Residual blocks-building blocks of resnet.
- Soda, Paolo and Giulio Iannello. 2009. Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis. *Trans. Info. Tech. Biomed.*, 13(3):322–329.
- Standford. 2019. Convolutional neural network.
- Sumithra, M. G. and B. Deepa. 2016. Performance analysis of various segmentation techniques for detection of brain abnormality. In *2016 IEEE Region 10 Conference (TENCON)*, pages 2056–2061.
- Teichmuller, M. and M. Wolf. 2011. Application of fluorescence microscopy in coal petrology and oil exploration. *Journal of Microscopy*, 109:49 – 73.
- Thiel, Cora Sandra, Svantje Tauber, Christian Seebacher, Martin Schropp, Rainer Uhl, Beatrice Lauber, Jennifer Polzer, Srujana Neelam, Ye Zhang 4, and Oliver Ullrich. 2019. Real-time 3d high-resolution microscopy of human cells on the international space station. *International Journal of Molecular Sciences*, page 2033.
- Thomas, R. M. and J. John. 2017. A review on cell detection and segmentation in microscopic images. In *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*,

- pages 1–5.
- Trottier, Ludovic, Philippe Gigure, and Brahim Chaib-draa. 2017. Parametric exponential linear unit for deep convolutional neural networks. pages 207–214.
- Tulsani, Hemant, Saransh Saxena, and Naveen Yadav. 2013. Segmentation using morphological watershed transformation for counting blood cells. *IJCAIT*, 2(3):28–36.
- Vala, Hetal J and Astha Baxi. 2013. A review on otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(2):387–389.
- Wang, Hongkai, Zongwei Zhou, Yingci Li, Zhonghua Chen, Peiou Lu, Wenzhi Wang, Wanyu Liu, and Lijuan Yu. 2017. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18f-fdg pet/ct images. *EJNMMI Research*, 7(1).
- Wu, H. ., J. Gil, and J. Barba. 1998. Optimal segmentation of cell images. *IEE Proceedings - Vision, Image and Signal Processing*, 145(1):50–56.
- Xia, Xide and Brian Kulis. 2017. W-net: A deep model for fully unsupervised image segmentation.
- Xue, Yao and Nilanjan Ray. 2017. Cell detection with deep convolutional neural network and compressed sensing. *CoRR*, abs/1708.03307.
- Zhang, J and K Hoshino. 2019a. *Molecular sensors and nanodevices : principles, designs and applications in biomedical engineering*. Academic Press.
- Zhang, John X.J. and Kazunori Hoshino. 2019b. Fluorescence microscopy- an overview.

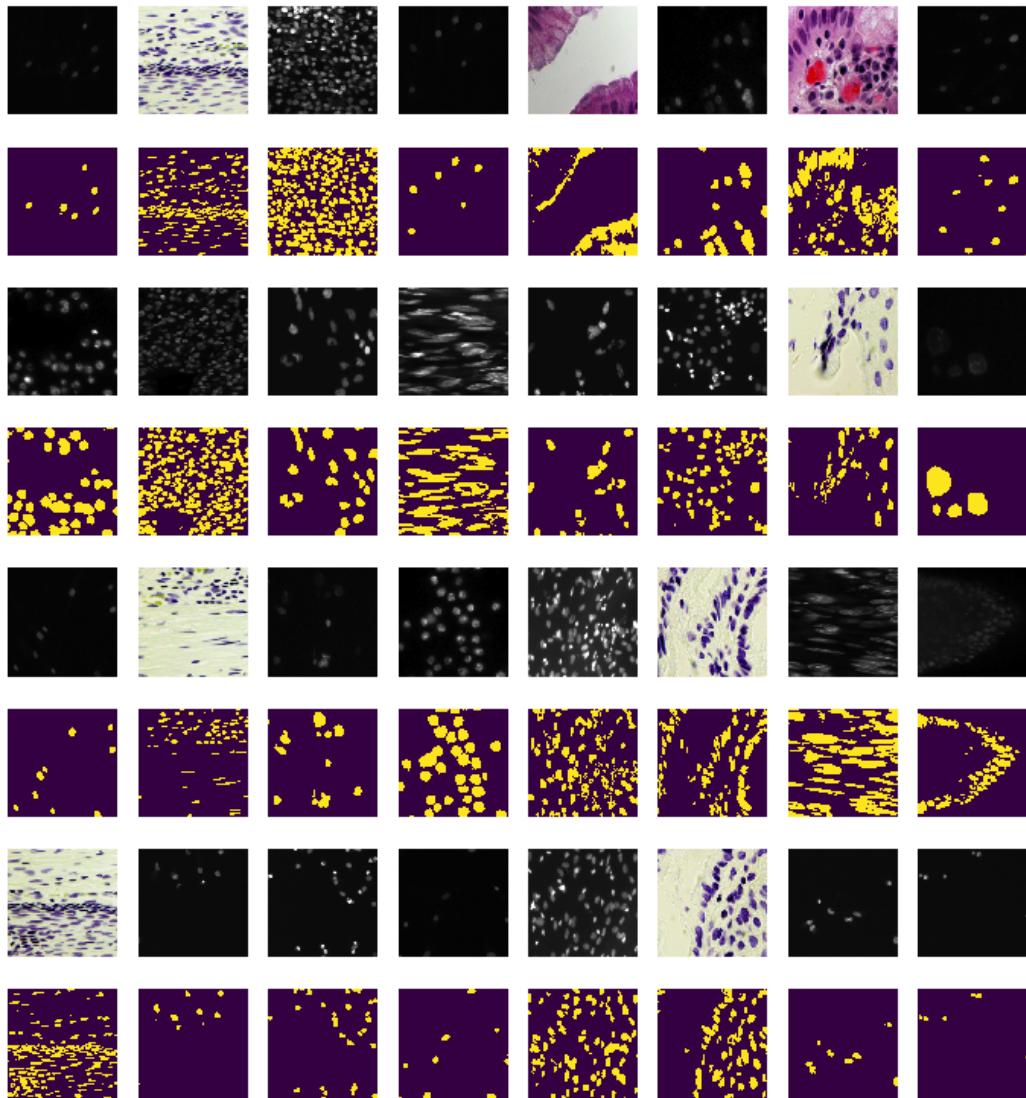
**Appendix G: Predicted results of U-Net on validation dataset**

The results of running U-Net model on validation dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by U-Net and third row show the ground truth provided by experts. The sequence repeats continuously.



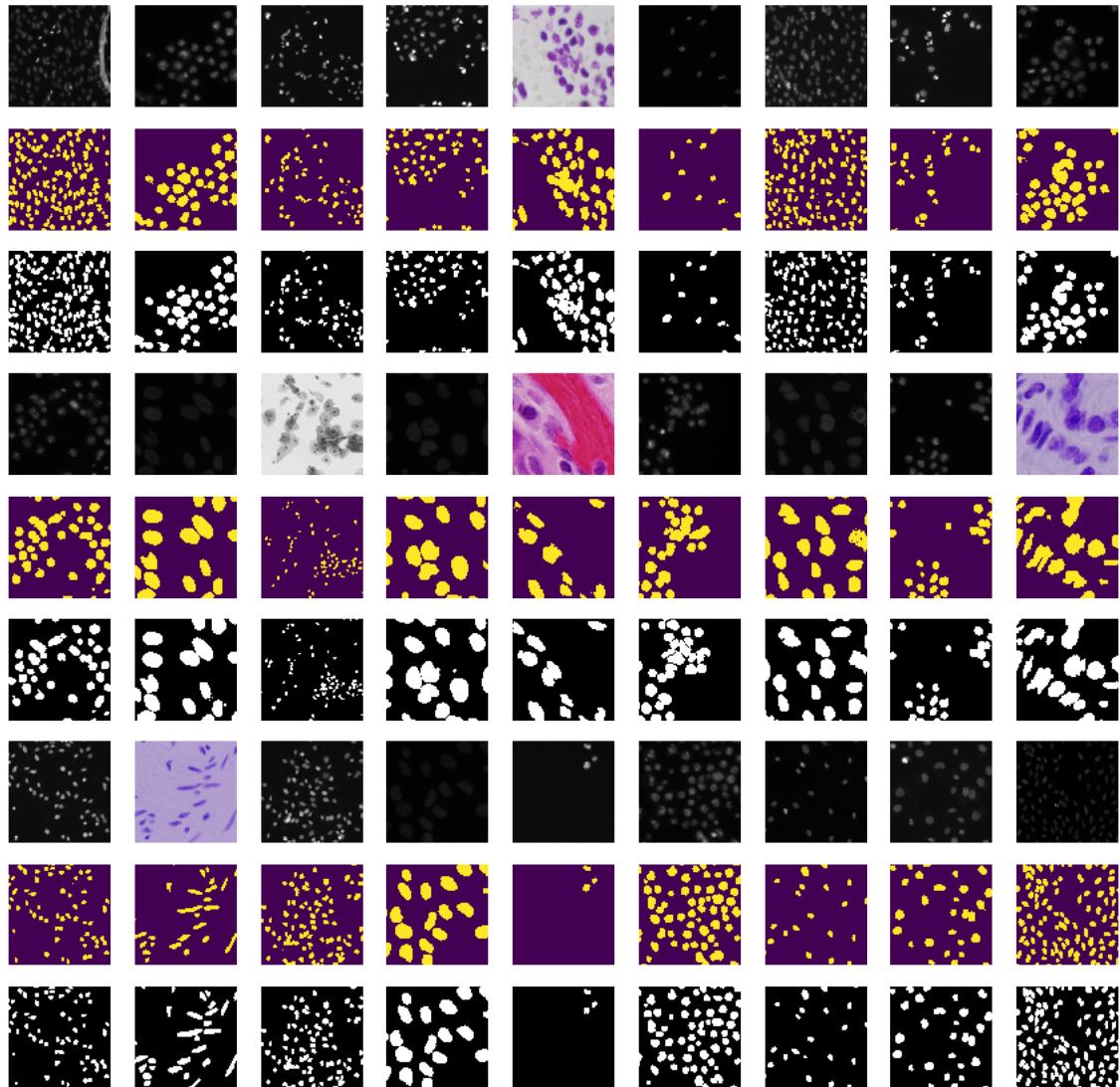
**Appendix H: Predicted results of U-Net on testing dataset**

The results of running U-Net model on testing dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by U-Net and the logic continues to apply. There is no ground truth provided for testing set to evaluate performance.



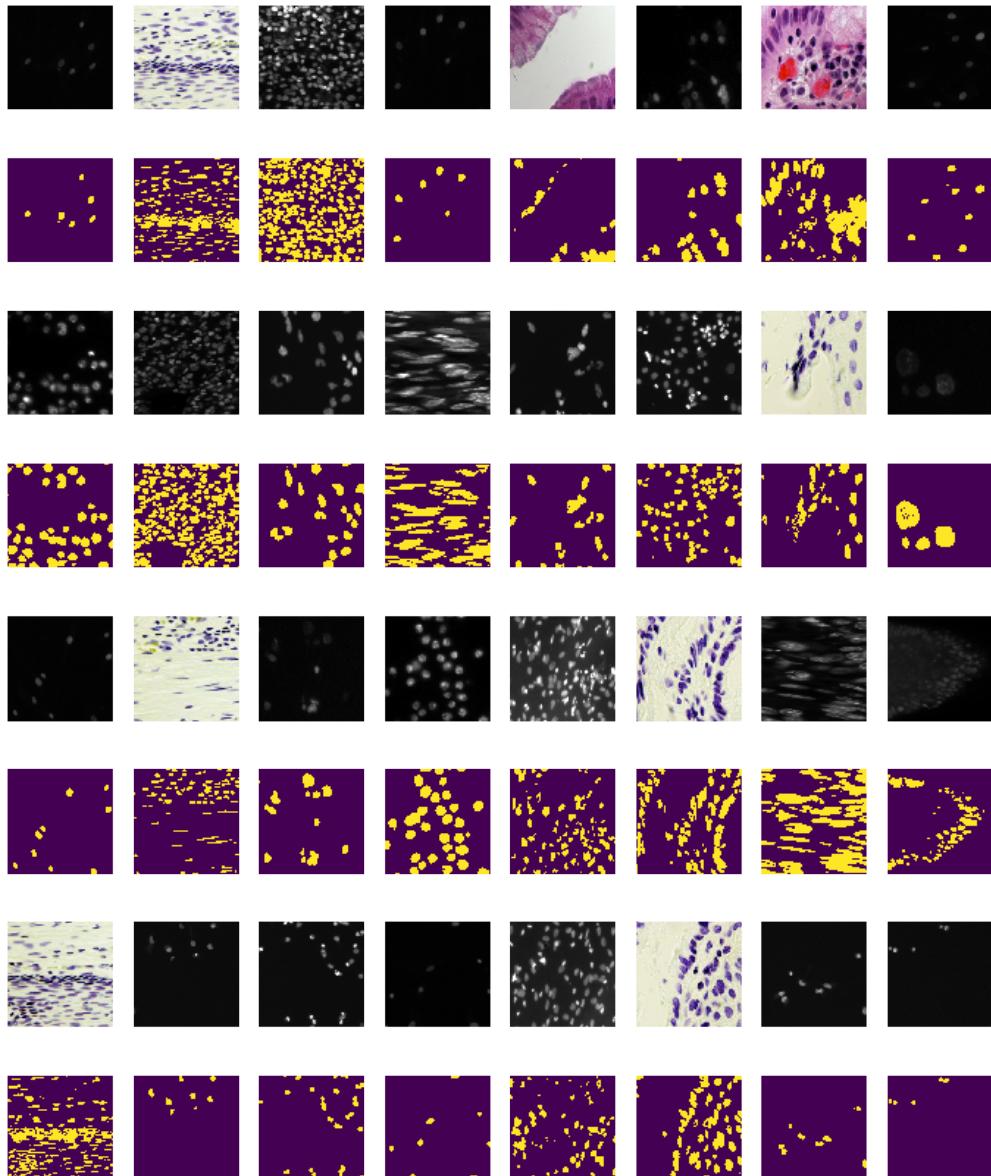
**Appendix I: Predicted results of R2U-Net on validation dataset**

The results of running R2U-Net model on validation dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by R2U-Net and third row show the ground truth provided by experts. The sequence repeats continuously.



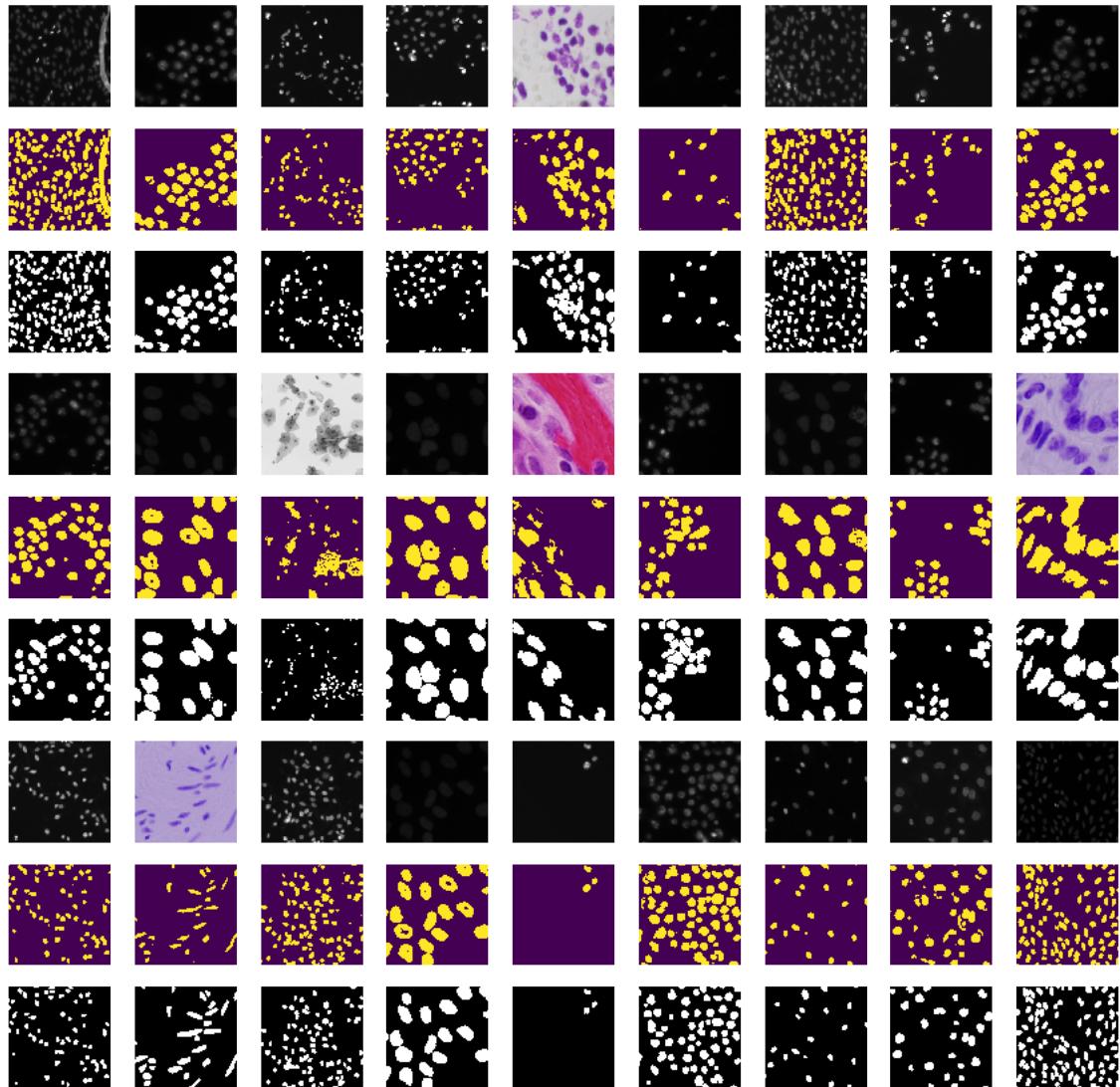
**Appendix J: Predicted results of R2U-Net on testing dataset**

The results of running U-Net model on testing dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by R2U-Net and the logic continues to apply. There is no ground truth provided for testing set to evaluate performance.



**Appendix K: Predicted results of W-Net on validation dataset**

The results of running W-Net model on validation dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by W-Net and third row show the ground truth provided by experts. The sequence repeats continuously.



**Appendix L: Predicted results of W-Net on testing dataset**

The results of running W-Net model on testing dataset are presented as below. The first row shows the original inputs; the second row shows the predicted labels produced by W-Net and the logic continues to apply. There is no ground truth provided for testing set to evaluate performance.

